E 20340 F

# Journal of Institutional and Theoretical Economics

# JITE

Vol. 173, No. 2 June 2017



ARTIBUS INGENUIS J·C·B·M· 1·8·0·1

Mohr Siebeck

# Centralized versus Decentralized Institutions for Expert Testimony

by

Chulyoung Kim*

Received August 16, 2014; in revised form March 5, 2016;
accepted March 14, 2016

The legal community has been debating the question of who should select and provide expert witnesses at trial: the litigant or the judge? Using a persuasion-game framework, I show that there is a trade-off. On one hand, the litigant may consult an expert even when the judge is reluctant to do so due to high costs. On the other hand, given the same amount of expert advice, the judge can make a more accurate decision when using her own expert's advice. I show that the cost of expert advice is an important factor in this trade-off. (JEL: C72, D82, K41)

## 1 *Introduction*

In the current American legal system, which I call the *decentralized institution*, expert witnesses are selected and retained by litigants. Thus, self-interested litigants invest in strong statements for their causes by searching for and retaining favorable expert witnesses. Proponents of such an institution argue that the competitive nature of the system provides litigants with strong incentives to collect and reveal evidence to defend their causes, in which process the truth is found.[1]

Opponents of the present system, however, argue that the "battles of the experts" observed in many civil litigations are obstacles to finding the truth. As expert witnesses are selected by and affiliated with the litigants, there exists inevitable evidence distortion: only those experts whose opinions align with the litigants' interests will be heard at trial. Such opportunistic behavior by the litigants with the help of their hired guns may work to the detriment of the accuracy of the final verdict, and thereby place the legitimacy of the legal procedure itself in question. Con-

* School of Economics, Yonsei University, Seoul, Republic of Korea. I am grateful to Biung-Ghi Ju and the referees for their valuable comments that substantially improved this article. I also thank participants at various seminars and conferences for their valuable comments. All remaining errors are mine. This work was supported by the Yonsei University Future-leading Research Initiative of 2015 (2015-22-0076).

[1] Posner (1988, 1999) presented strong arguments for such decentralized institutions.

*Journal of Institutional and Theoretical Economics* 173, 209–238 – ISSN 0932-4569
DOI: 10.1628/093245616X14664894246848 – 

cerned about the drawbacks, many scholars have long argued for a more centralized system for expert witnesses, which I call the *centralized institution*, allowing judges to appoint neutral experts. In particular, there have been numerous reform proposals suggesting that the court appoint its own experts, thereby enhancing the inquisitorial component in the American legal system. For example, see Runkle (2001), who discusses the structure of the Court Appointed Scientific Experts program created by the American Association for the Advancement of Science in order to help judges obtain independent experts. Also see Hillman (2002), Adrogué and Ratliff (2003), and Kaplan (2006), among others. Based on his experience as Judge Richard Posner's court-appointed economic expert, Sidak (2013) argues for court-appointed, neutral economic experts. Many reformers, most famously including Hand (1901), argued that the appropriate remedy for adversarial bias (combined with inexpert juries) was increased reliance on court-appointed, nonpartisan experts. The main task of this paper is to evaluate such reform proposals, focusing especially on the accuracy of the legal system.[2]

The main results show that there is a trade-off between the two institutional arrangements. On the one hand, the litigants are willing to consult an expert even when the court is reluctant to appoint its own experts due to high costs. More precisely, there exists an interval of cost parameters such that no expert is utilized in the centralized institution, whereas an expert is utilized in the decentralized institution, when the cost of using expert advice lies in the interval. This result obtains because the court, as an impartial decision-maker, must weigh the possibility that "bad news" will lead to an incorrect decision because expert advice provides imperfect information about the truth. Proposition 3 shows the ways in which such consideration by the court reduces its incentive to utilize expert advice, relative to the litigants' incentives. On the other hand, given the same amount of expert advice in both institutions, the trier of fact can make a more accurate decision when using a court-appointed expert's advice at trial. As litigants attempt to distort evidence, there exists an information loss under the decentralized institution. This behavior by litigants increases the uncertainty faced by the trier of fact, leading to a less-accurate decision than in the centralized institution. Propositions 4 and 5 provide more precise statements.

The main model in this paper is a persuasion game with endogenous information acquisition, which is adapted from Kim (2014a). In that paper, I study two commonly used forms of legal processes, the adversarial and inquisitorial systems,[3]

[2] Although the main body of this paper is presented in a civil-litigation context, the result is not limited to it. See section 6.4 for an interpretation of the model in a criminal-litigation context.

[3] For an important debate on the relative merits of the adversarial and inquisitorial systems, see Posner (1988, 1999) and Tullock (1975, 1980, 1988). The distinction between the decentralized institution and the adversarial system (the centralized institution and the inquisitorial system) is subtle. The adversarial system is a legal system in which the case under dispute is organized and developed by the initiatives of the interested parties, rather than by an impartial third party. In theory, the adversarial system can coexist with the centralized institution, relegating to the court only the role of providing the judge with

within a persuasion-game environment, and show the conditions in which one system dominates the other in terms of accuracy. An important assumption is that both litigants have access to the same source of information, and therefore they obtain the same piece of evidence if they were successful in collecting information before a trial occurs. This assumption is crucial to the finding that only one litigant searches for information in equilibrium. In contrast, the current paper assumes that litigants have access to different information sources because each litigant seeks advice from an expert who may possess pieces of evidence different from others. The main results demonstrate that both litigants may consult an expert in equilibrium, depending on the cost of expert advice. Thus, the competition between the litigants in the pursuit of more favorable evidence for their own causes is better modeled in the current paper.

In general, economic analysis has been in favor of decentralized systems of evidence collection. The main intuition obtained from various economic models, as demonstrated in an early contribution by Milgrom and Roberts (1986), is that information possessed by litigants is eventually revealed to the fact-finder because of competition among them: as a piece of evidence detrimental to one party is beneficial to the other, any evidence is eventually revealed by one of the competing parties. This intuition has been confirmed to be robust (albeit not free from debate) in a more general environment, and has provided strong support for the current form of the American legal system. Milgrom and Roberts (1986) employ a persuasion-game framework for their analysis. See, among others, Froeb and Kobayashi (1996), Shin (1998), Demougin and Fluet (2008), and Kim (2014a) for the same line of research. Also see Froeb and Kobayashi (2001), Parisi (2002), and Emons and Fluet (2009a,b) for related research. While these papers assume that the litigants always supply biased information to the fact-finder, Kim (2016) studies a situation in which a litigant is willing to provide unbiased information. Kim (2015) studies a situation in which the fact-finder does not observe the *quality* of information proffered by the litigants. Although the existing literature focuses on communication problems between informed players and an uninformed decision-maker, the current paper adds one more dimension to the literature by introducing players' information acquisition behavior.

Using a principal–agent model, Dewatripont and Tirole (1999), Palumbo (2001, 2006), Iossa and Palumbo (2007), Deffains and Demougin (2008), and Kim (2014b) study whether information can be provided to the fact-finder at a lower cost in decentralized systems. These models also provide strong support for decentralized systems, showing that incentive constraints are easily satisfied by exploiting competition among agents. Thus, pointing out another merit of employing decentralized systems, this line of research complements the persuasion-game approach adopted in the current paper.

---

expert witnesses, which is the current development of the debate regarding the reform of expert law in the United States. The focus of the current paper is only on the rule governing expert witnesses, rather than on a broader discussion on the relative merits of the adversarial system and the inquisitorial system.

The remainder of the current paper is organized as follows. Section 2 presents the basic model used for subsequent analysis. Section 3 analyzes the decentralized institution, section 4 investigates the centralized institution, and section 5 compares the two institutions with respect to accuracy. Section 6 discusses the extensions and implications of the main results. Finally, section 7 concludes. Proofs of the propositions appear in the appendix.

## 2 *Model*

Consider a lawsuit in which a plaintiff (henceforth P) contends with a defendant (henceforth D). Each litigant pleads for his cause, and a judge (henceforth J) must decide whose cause should prevail at trial. J wants to make a correct decision accurately reflecting the true state $t \in \{h,l\}$. When $t = h$, J obtains a payoff of 1 if she rules in favor of D, and a payoff of 0 otherwise. Similarly, when $t = l$, J obtains a payoff of 1 if she rules in favor of P, and a payoff of 0 otherwise. In contrast, each litigant wants to win at trial regardless of $t \in \{h,l\}$: a litigant obtains a payoff of 1 if he wins at trial, and a payoff of 0 otherwise. The prior probability that $t = h$ is denoted by $\mu = P(t = h)$.

To assist J in finding the truth, experts may be called to testify at trial.[4] An expert is someone better equipped than laypersons through "knowledge, skill, experience, training, or education" (Federal Rule of Evidence 702) to perceive the truth in his specialized domains. He can tell whether the plaintiff's illness is due to exposure to specific toxic chemicals from the workplace, whether the plaintiff underwent erroneous medical treatment in the hospital, and so forth. Such testimony provided by expert witnesses is valuable, sometimes crucial, in the fact-finding process, particularly when the dispute involves scientific and technical issues. Thus, experts play an important role in civil litigation. Formally, each expert has access to a conditionally i.i.d. random variable $x$ with probability[5] $e \in (0,1)$, where $x$ takes the value either $H$ or $L$ with the conditional probability $P(H \mid h) = P(L \mid l) = p > 1/2$.

Note that $x = H$ can be said to be *favorable* evidence for D and *unfavorable* evidence for P, because, as clarified in the main analysis, if J observes $x = H$, she believes that $t = h$ is more likely to be the true state and therefore rules in favor of D. Similarly, $x = L$ can be said to be favorable evidence for P and unfavorable evidence for D. Also note that $e$ can be thought of as the expert's quality. If $e$ is close to 1, the expert can be relied upon to provide valuable evidence for the issue, whereas if $e$ is close to 0, the expert's ability is questionable and is unlikely to be able to provide the trier of fact with useful guidance. I assume that all available experts have the same quality, i.e., they have the same chance of receiving infor-

[4] Gross (1991) notes that experts testified in 86 % of civil trials in a sample of California cases between 1985 and 1986.

[5] Thus, an expert observes the realization of $x$ with probability $e$ and fails to observe it with probability $1-e$.

mation upon investigation.[6] Another measure of an expert's quality in the model is $p$, because as $p$ increases, the evidence collected by an expert becomes more accurate. Note that an expert's quality in this sense cannot be lower than the prior probability $\mu$, because if $p$ is smaller than $\mu$, the evidence $x$ is not precise enough to persuade J to change her decision depending on the realization of $x$. Thus, I assume $\mu \in (1-p, p)$, which guarantees that J's decision is responsive to the evidence and helps us avoid uninteresting cases.

In the current American legal system, expert witnesses are selected and retained by litigants; I call this the decentralized institution (henceforth DI). Opponents of the present system argue for a more centralized system for expertise, which I call the centralized institution (henceforth CI), allowing judges to appoint neutral experts. The main task of this paper is to study the strength and weakness of each institution, focusing especially on accuracy.

Formally, DI is modeled as an incomplete-information dynamic game with two stages, the *pretrial stage* and the *trial stage*. In the pretrial stage, by paying a cost $c > 0$,[7] a litigant $i \in \{P, D\}$ can secretly[8] consult (at most) one expert to obtain evidence to present at trial. If his expert observes the hidden evidence, the litigant obtains $x_i \in \{H, L\}$. A litigant cannot obtain any evidence if either he does not consult an expert or his expert cannot observe the hidden evidence.

In the trial stage, litigants present their evidence to J, and I denote a litigant $i$'s presentation by $r_i$. I assume that the evidence is verifiable, so litigants can choose to hide but cannot falsify the evidence presented to J. Thus, when a litigant has obtained $x_i$ from his expert, he either truthfully reveals it ($r_i = x_i$) or hides it as an attorney's work product and remains silent ($r_i = \phi$). If a litigant has no evidence, he remains silent ($r_i = \phi$). Thus, when a litigant remains silent, J cannot ascertain whether the litigant is hiding evidence or simply uninformed. In such a situation, J forms a Bayesian posterior incorporating her belief about the litigants' strategies. Finally, J makes a decision regarding which party wins at trial, payoffs are realized, and the game ends.

In contrast, CI is modeled as a decision-making problem in which J makes a decision after directly consulting experts for evidence and paying a cost $c > 0$. To make the two institutions, CI and DI, comparable, I assume that J can consult at most two experts in CI, so that the maximum number of experts consulted in each

---

6 An alternative approach is to assume a pool of heterogeneous experts with a mean quality level $e$, where an expert is randomly contacted at the request of the litigants or the court. This approach is similar in spirit to the proposal by Robertson (2010). The result is the same under both approaches.

7 This cost may include the cost of searching for experts, preparing a dossier for them, reviewing their technical reports, separating relevant pieces of evidence from irrelevant ones, and so forth.

8 That is, a litigant's action is not observable to J and the other litigant. This assumption seems realistic in light of the fact that a litigant's search activity is generally not discoverable. This assumption also simplifies the analysis. It is not clear whether the equilibrium structure remains intact when J can directly observe a litigant's action, because it is possible that a litigant may adopt a mixed strategy in equilibrium.

institution is 2. I also assume that the cost of consulting an expert is the same in both institutions.

In the following analysis, I first analyze DI and find the perfect Bayesian equilibrium, which is simply referred to as the equilibrium. I then proceed to the analysis of CI and compare the results from the two institutions.

## 3 Decentralized Institution

### 3.1 Trial Stage

I first analyze the players' behavior in the trial stage. It is straightforward to see that the litigants only reveal favorable evidence (i.e., P never reveals $x_{\mathrm{P}} = H$, whereas D never reveals $x_{\mathrm{D}} = L$), because revealing unfavorable evidence only reduces their chances of winning. Thus, evidence distortion naturally arises in the trial stage, and J must take account of such incentives of the litigants when observing the litigants' presentations.[9]

In the presence of evidence distortion by the litigants, there are four possible situations:

1. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (L, \phi)$: P wins;
2. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (\phi, H)$: D wins;
3. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (L, H)$: J's decision depends on $\mu$;
4. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (\phi, \phi)$: J's decision depends on her belief about the litigants' behavior.

To be more precise, consider the first situation, in which J observes $L$ from P, and D remains silent. The "low" signal from P alone reduces J's posterior belief below $1/2$.[10] As D's silence cannot increase J's posterior belief,[11] it is easy to establish that J rules in favor of P. The reasoning under the second situation is analogous. In the third situation, both litigants reveal evidence supporting their own claims. As the signals are conditionally i.i.d., these two pieces of evidence nullify each other, inducing J to hold a posterior belief equal to the prior belief. Thus, D wins if $\mu \geq 1/2$, and P wins otherwise. This situation shows why DI is vulnerable to criticisms such as "war of attrition" or "money contest."[12] By consulting experts

[9] This feature is not new to the literature, and many papers examine various models in which evidence distortion is introduced in one way or another. See Sobel (2013) for a survey on this topic.

[10] To be more precise, suppose only P consulted an expert. Then $\phi$ has no information content, and the only piece of information is $L$. Thus, J's posterior belief becomes $P(t = h \mid x = L) = [\mu(1-p)]/[\mu(1-p) + (1-\mu)p] < 1/2$, where the inequality holds because $\mu \in (1-p, p)$.

[11] D is silent when he is uninformed or hiding $x_{\mathrm{D}} = L$. In the former case, there should be no change in J's posterior belief. In the latter case, J's posterior belief must fall. As J's posterior belief is a convex combination of those two beliefs, the posterior cannot increase following D's silence.

[12] In his papers, Tullock criticizes such decentralized legal systems for leading to excessive expenditures through unnecessary duplication and costly overproduction of misleading information. See Tullock (1975, 1980, 1988).

and selectively presenting evidence that is favorable to their causes, the litigants can provide the trier of fact with the impression that the issue at hand is subject to contestation, which leaves her equipoised without any change in her assessment regarding the dispute.

In the fourth situation, J receives no *direct* evidence, because both litigants remain silent. However, she could obtain *indirect* evidence from the litigants' behavior:

*(a)* First, suppose that J believes that neither litigant consulted an expert in the pretrial stage. Then, J believes that both litigants are silent because they are simply uninformed, and therefore J's posterior belief is equal to the prior belief. Thus, D wins if $\mu \geq 1/2$, and P wins otherwise.

*(b)* Second, suppose J believes that only one litigant consulted an expert in the pretrial stage. It turns out that J forms a posterior belief against that litigant. For example, if J believes that only P consulted an expert, her posterior belief in the no-evidence event $(r_{\text{P}}, r_{\text{D}}) = (\phi, \phi)$, denoted as $\mu(\phi, \phi)$, is given by

$$\begin{aligned} \mu(\phi,\phi) &= \frac{\mu q_h}{\mu q_h + (1-\mu) q_l} \\ &= \frac{\mu(ep+1-e)}{\mu(ep+1-e)+(1-\mu)(e(1-p)+1-e)} \\ &> \mu, \end{aligned} \tag{1}$$

where $q_t$ is the probability that P remains silent given $t \in \{h, l\}$; e.g., given that the true state is high, P remains silent either because he obtained unfavorable evidence ($x_{\text{P}} = H$) from his expert (with probability $ep$) or his expert could not observe the hidden evidence (with probability $1-e$), which gives us $q_h$. If P's silence is due to his manipulation, J's posterior belief must be higher than $\mu$, and if P's silence is due to no information, J's posterior belief must be equal to $\mu$. Thus, J's posterior belief, which is a convex combination of the beliefs under the two possibilities, becomes higher (i.e., against P) if she believes that only P consulted an expert. Based on J's posterior belief, D wins if $\mu(\phi, \phi) \geq 1/2$, and P wins otherwise.

*(c)* Third, if J believes that both litigants consulted an expert, her posterior belief is equal to the prior belief, because the indirect evidence from each litigant's silence nullifies that from the other.[13] Thus, D wins if $\mu \geq 1/2$, and P wins otherwise.

At this point, J's belief about which litigant has consulted an expert can be arbitrary. In equilibrium, however, her belief must be consistent with the litigants'

[13] This is because I assume that the experts have the same chance of observing the evidence. If I assume that the litigants randomly contact an expert from a pool of heterogeneous experts, I obtain the same result. See Sharif and Swank (2012) for an analysis of heterogeneity among litigants.

strategies, which will be clarified in section 3.3. When no direct evidence is revealed in the trial stage, D wins if $\mu(\phi,\phi) \geq 1/2$, and P wins otherwise. I say the burden of proof (henceforth BOP) is on P if $\mu(\phi,\phi) \geq 1/2$, and on D otherwise.

DEFINITION *The BOP is said to be on P if* $\mu(\phi,\phi) \geq 1/2$*, and on D otherwise.*

Note that if a litigant bears the BOP, he knows that he can win only when he presents favorable evidence in the trial stage. For example, suppose P bears the BOP. If P cannot reveal $x_P = L$ (which implies that P will remain silent), J will eventually observe $(r_P, r_D) = (\phi, H)$ or $(r_P, r_D) = (\phi, \phi)$ in the trial stage, and both cases lead to D's winning.

### *3.2 Pretrial Stage*

Using backward induction, I now analyze the litigants' behavior regarding their decisions to consult an expert in the pretrial stage. Throughout the analysis, I assume that the BOP falls on P. The opposite case in which the BOP falls on D easily follows because the result is symmetric, and therefore its analysis is omitted to save space. The analysis of this section is separated into two parts, depending on the prior probability: $\mu \geq 1/2$ and $\mu < 1/2$.

#### 3.2.1 Prior in Favor of D

In this subsection, I assume $\mu \geq 1/2$. P's expected payoff is (remember that the BOP is on P)

0 if he does not consult an expert, or

$\underbrace{\mu e(1-p)(1-ep\cdot s_D)+(1-\mu)ep(1-e(1-p)\cdot s_D)}_{(*)\text{ prob. of P's winning}} - c$ if he consults an expert,

where $s_D = 1$ if D contacts an expert and $s_D = 0$ otherwise.

If P does not consult an expert (leading to $r_P = \phi$), it is obvious that he will lose in the trial stage, because D's presentation is either $r_D = H$ (leading to $(r_P, r_D) = (\phi, H)$) or $r_D = \phi$ (leading to $(r_P, r_D) = (\phi, \phi)$), and P loses in both cases. Thus, P's expected payoff is 0.

If P consults an expert, it is straightforward to check that P wins in the trial stage only if $(r_P, r_D) = (L, \phi)$. In this case, the probability of P's winning $(*)$ depends on P's belief about D's action:

*(a)* If D does not contact an expert ($s_D = 0$), the probability of P's winning $(*)$ is given by

$$\mu e(1-p) + (1-\mu)ep = e\underbrace{(\mu(1-p)+(1-\mu)p)}_{=P(x_P=L)},$$

where $P(x_P = L)$ is the unconditional probability that the hidden information is $L$. Because D does not provide any evidence, there are only two possibilities in the trial stage: $(r_{\rm P}, r_{\rm D}) = (L, \phi)$ or $(r_{\rm P}, r_{\rm D}) = (\phi, \phi)$. That is, P wins if and only if he can obtain and reveal $x_{\rm P} = L$ to J, whose probability is given above. This probability gives us P's expected payoff as proposed if $s_{\rm D} = 0$.

*(b)* If D contacts an expert ($s_{\rm D} = 1$), the probability of P's winning $(*)$ is given by

$$\mu \underbrace{e(1-p)}_{(A1)} \underbrace{(1-ep)}_{(A2)} + (1-\mu)ep(1-e(1-p)).$$

Note that P cannot secure his winning by revealing $x_{\rm P} = L$ in the trial stage, because D can counteract P's evidence by revealing $x_{\rm D} = H$, in which case J's posterior belief is equal to $\mu \geq 1/2$ and therefore D wins. Thus, if $s_{\rm D} = 1$, the probability of P's winning $(*)$ is lower than under $s_{\rm D} = 0$: $(A1)$ is the probability that P obtains $x_{\rm P} = L$ given $t = h$, and $(A2)$ is the probability that D remains silent given $t = h$. Thus, $(A1) \times (A2)$ is the probability that $(r_{\rm P}, r_{\rm D}) = (L, \phi)$ occurs in the trial stage, given $t = h$. The other term can be similarly understood. This probability gives us P's expected payoff as proposed if $s_{\rm D} = 1$.

Thus, P consults an expert if and only if the cost of consulting an expert is less than the net benefit from expert advice:

$$c \leq c_{\rm P}^{\rm P} = \mu e(1-p)(1-ep \cdot s_{\rm D}) + (1-\mu)ep(1-e(1-p) \cdot s_{\rm D}),$$

where *(i)* the subscript P in the threshold $c_{\rm P}^{\rm P}$ indicates that this is the threshold for P, and *(ii)* the superscript P in $c_{\rm P}^{\rm P}$ indicates that the BOP is on P. As shown above, D's counteracting effort reduces P's incentive to consult an expert: $c_{\rm P}^{\rm P}$ is larger when $s_{\rm D} = 0$ than when $s_{\rm D} = 1$. Thus, as D becomes *more* aggressive in consulting an expert, P becomes *less* aggressive.

As the event of D's winning is the complement of P's winning, it is straightforward to calculate D's expected payoff as follows:

$$\underbrace{1 - \{\mu e(1-p) + (1-\mu)ep\} \cdot s_{\rm P}}_{\text{prob. of D's winning}} \text{ if he does not consult an expert, or}$$

$$\underbrace{1 - \{\mu e(1-p)(1-ep) + (1-\mu)ep(1-e(1-p))\} \cdot s_{\rm P}}_{\text{prob. of D's winning}} - c \text{ if he consults an expert,}$$

where $s_{\rm P} = 1$ if P contacts an expert and $s_{\rm P} = 0$ otherwise. Thus, D's behavior can be also summarized by an appropriate threshold $c_{\rm D}^{\rm P}$ such that D consults an expert if and only if $c \leq c_{\rm D}^{\rm P}$, where the superscript and subscript in $c_{\rm D}^{\rm P}$ have the same meaning as before. The table summarizes the simultaneous game that the litigants play in the pretrial stage.

Note that D never consults an expert when P does not, because $c_{\rm D}^{\rm P} = 0$ if $s_{\rm P} = 0$. This finding shows that D's motive for consulting an expert is primarily to counteract his opponent's evidence when he does not bear the BOP. Thus, as P becomes *more* aggressive in consulting an expert, D also becomes *more* aggressive.

*Table*
Payoff Table in Pretrial Stage (BOP on P and $\mu \geq 1/2$)

| | | P: Consult | P: Not |
|---|---|---|---|
| D | Consult | $1-\{\mu e(1-p)(1-ep)+(1-\mu)ep(1-e(1-p))\}-c$ <br> $\mu e(1-p)(1-ep)+(1-\mu)ep(1-e(1-p))-c$ | $1-c$ <br> $0$ |
| | Not | $1-\{\mu e(1-p)+(1-\mu)ep\}$ <br> $\mu e(1-p)+(1-\mu)ep-c$ | $1$ <br> $0$ |

### 3.2.2 Prior in Favor of P

In this subsection, I assume $\mu < 1/2$. It is routine to check that P's expected payoff is given as follows:

$$0 \text{ if he does not consult an expert, or}$$
$$\underbrace{\mu e(1-p)+(1-\mu)ep}_{=eP(x_P=L)=\text{prob. of P's winning}}-c \text{ if he consults an expert.}$$

Note that if P obtains and reveals favorable evidence, he always wins in the trial stage regardless of D's action. In contrast to the previous case, D cannot counteract P's evidence, because P enjoys a favorable prior assessment for his cause: P wins not only under $(r_P,r_D) = (L,\phi)$, but also under $(r_P,r_D) = (L,H)$, because J's posterior belief is equal to $\mu < 1/2$, which leads to P's winning. Thus, P consults an expert if and only if

$$c \leq c_P^P = \mu e(1-p)+(1-\mu)ep.$$

It is also straightforward to obtain D's expected payoff as follows:

$$\underbrace{1-\{\mu e(1-p)+(1-\mu)ep\}\cdot s_P}_{=1-eP(x_P=L)s_P=\text{prob. of D's winning}} \text{ if he does not consult an expert, or}$$
$$\underbrace{1-\{\mu e(1-p)+(1-\mu)ep\}\cdot s_P}_{=1-eP(x_P=L)s_P=\text{ prob. of D's winning}}-c \text{ if he consults an expert.}$$

It is clear that D never wants to consult an expert. Note that D's winning does not depend on his action, but only on P's: whenever P reveals $x_P = L$, P wins regardless of D's presentation (i.e., P wins under $(r_P,r_D) = (L,H)$ and $(r_P,r_D) = (L,\phi)$); and whenever P cannot reveal $x_P = L$, P loses regardless of D's presentation (i.e., P loses under $(r_P,r_D) = (\phi,H)$ and $(r_P,r_D) = (\phi,\phi)$).[14] Thus, D rationally chooses not to consult any expert, leaving the final verdict dependent on P's choice.

[14] Remember that P loses under $(r_P,r_D) = (\phi,\phi)$, because I assume that the BOP is on P.

### *3.3 Equilibrium*

Note that the allocation of the BOP depends on J's belief regarding which litigant consulted an expert. Conversely, when the litigants choose whether to consult an expert, they take the BOP (and therefore J's belief about their own behavior) as given. In an equilibrium, the BOP allocation must be consistent with the litigants' strategies. I now turn to this subject and find the equilibria in DI.

It turns out that there exist two types of equilibria in DI. The first type is called the *P-equilibrium*, and the second type the *D-equilibrium*. In the P-equilibrium the BOP is on P, whereas in the D-equilibrium it is on D. I present the first main result in the following proposition. I omit the D-equilibrium result to save space, considering that it is symmetric.

PROPOSITION 1 *There exist* $\underline{c}$ *and* $\bar{c}$ *such that* $0 < \underline{c} < \bar{c}$ *and the following is true:*

1. *If* $\mu \geq 1/2$*, the P-equilibrium always exists, and*

   $\bar{c} < c$*: neither litigant consults an expert in the P-equilibrium,*
   $c \in (\underline{c}, \bar{c}]$*: only P consults an expert in the P-equilibrium,*
   $c \leq \underline{c}$*: both litigants consult an expert in the P-equilibrium.*

2. *If* $\mu < 1/2$*,*

   $\bar{c} < c$*: the P-equilibrium does not exist,*
   $c \leq \bar{c}$*: the P-equilibrium, in which only P consults an expert, exists if* $\mu$ *is close to* $1/2$ *or e is close to* $1$*.*

The results are intuitive. Consider the first part, in which $\mu \geq 1/2$. When the cost of consulting an expert is large, neither litigant is willing to incur a cost to consult an expert. In the P-equilibrium, this implies that J observes no evidence in the trial stage and, knowing that no expert was involved in equilibrium, rules in favor of D because her posterior belief is equal to $\mu \geq 1/2$. Although P knows that he will surely lose in the trial stage, he refrains from using expert advice, because it is not worth the cost.

As $c$ decreases, litigants are willing to consult an expert in equilibrium, and if $c$ is sufficiently small, both litigants consult an expert for information. Note that P has a higher incentive to use an expert, and therefore only P uses expert advice for the intermediate range of $c$. Because the BOP is on P, there is no chance for P to win if he does not consult an expert, whereas D still has a chance to win without using expert advice. Therefore, expert advice has a larger effect on P's expected payoff, generating the cost range in which only P consults an expert.

On the other hand, the existence of the P-equilibrium is not guaranteed for $\mu < 1/2$, in which case P enjoys a favorable initial assessment toward his claim. Note that, as the analysis of the pretrial stage reveals, D has no incentive to consult an expert in this case, because J's decision does not depend on D's presentation in the trial stage. Thus, either P alone consults an expert for small $c$, or neither litigant uses expert advice for large $c$.

If $c$ is large, no expert is consulted in equilibrium, and J therefore rules in favor of P after observing no evidence, because $\mu(\phi,\phi) = \mu < 1/2$. However, such a posterior belief is not consistent with the BOP's being on P, and therefore the P-equilibrium does not exist in this situation. If $c$ is small, P consults an expert, which increases J's equilibrium posterior belief $\mu(\phi,\phi)$ because J exercises skepticism toward P's silence in the trial stage. Thus, if this increase in belief is sufficiently large, I have $\mu(\phi,\phi) \geq 1/2$, which supports the existence of the P-equilibrium. Observe that this is possible if $\mu$ is large (i.e., $\mu$ is close to $1/2$) or $e$ is large (i.e., $e$ is close to 1). If $\mu$ is close to $1/2$, even a small degree of posterior updating will move J's equilibrium belief beyond $1/2$. If $e$ is close to 1, P's silence is likely to have come from manipulation, which increases J's equilibrium posterior belief by a large amount.

## 4 Centralized Institution

In CI, J makes a decision after directly consulting experts. Because J directly interacts with experts, she observes evidence from experts without any information loss arising from evidence distortion as in DI.[15] In the following analysis, I study J's choice of using expert advice and her final decision at trial under the assumption that $\mu \geq 1/2$. As the analysis for the other case, $\mu < 1/2$, is symmetric,[16] I omit the result to save space and to avoid unnecessary confusion.

First, suppose that J consults two experts. For comparison with DI, I denote the result from the first expert's investigation as $r_{\mathrm{P}}$ and that from the second expert's investigation as $r_{\mathrm{D}}$. The following are the possible situations:

$$(r_{\mathrm{P}}, r_{\mathrm{D}}) = \begin{cases} (H,H) & : \text{D wins}, \\ (H,L) \text{ or } (L,H) & : \text{D wins } (\because \text{ posterior is equal to } \mu \geq 1/2), \\ (L,L) & : \text{P wins}, \\ (H,\phi) \text{ or } (\phi,H) & : \text{D wins}, \\ (L,\phi) \text{ or } (\phi,L) & : \text{P wins}, \\ (\phi,\phi) & : \text{D wins } (\because \text{ posterior is equal to } \mu \geq 1/2). \end{cases}$$

In contrast to DI, there is no indirect evidence that can be collected from the no-evidence event, $(\phi,\phi)$, because it simply indicates that both experts are uninformed. Thus, J has no information under the event $(\phi,\phi)$, and her posterior belief

---

15 Evidence distortion could arise in CI as well. For this possibility, see Dewatripont and Tirole (1999) and the extensions of their model, including Palumbo (2001, 2006), Iossa and Palumbo (2007), Deffains and Demougin (2008), and Kim (2014b), which adopt an incomplete-contract framework.

16 Proposition 2 presents the result for the case of $\mu \geq 1/2$ and is summarized by the thresholds $\underline{c}_{\mathrm{J}}$ and $\bar{c}_{\mathrm{J}}$. The result for $\mu < 1/2$ can also be summarized by appropriate thresholds with the same structure as in Proposition 2.

therefore is equal to her prior belief. Because I assume $\mu \geq 1/2$, D wins under such a situation.

Anticipating these results, J's expected payoff when consulting two experts is

$$\begin{aligned}
\pi_{\mathrm{J}}^2 = & \underbrace{P(H,H)\mu(H,H) + 2P(H,L)\mu + P(L,L)(1-\mu(L,L))}_{\text{exp. payoff from observing both signals}} \\
& + \underbrace{2P(H)\mu(H) + 2P(L)(1-\mu(L))}_{\text{exp. payoff from observing only one signal}} \\
& + \underbrace{(1-e)^2\mu}_{\text{exp. payoff from observing no signal}} \\
& - \underbrace{2c}_{\text{cost of expert advice}} \\
= & \, e^2(\mu p^2 + 2p(1-p)\mu + (1-\mu)p^2) + 2e(1-e)(\mu p + (1-\mu)p) + (1-e)^2\mu - 2c;
\end{aligned}$$

$$\begin{aligned}
P(j,j')&: \text{probability of } (r_{\mathrm{P}}, r_{\mathrm{D}}) = (j,j') \text{ for } j,j' \in \{H,L\}, \\
P(j)&: \text{probability of } (r_{\mathrm{P}}, r_{\mathrm{D}}) = (j,\phi) \text{ for } j \in \{H,L\}, \\
\mu(j,j')&: \text{posterior from } (r_{\mathrm{P}}, r_{\mathrm{D}}) = (j,j') \text{ for } j,j' \in \{H,L\}, \\
\mu(j)&: \text{posterior from } (r_{\mathrm{P}}, r_{\mathrm{D}}) = (j,\phi) \text{ for } j \in \{H,L\}.
\end{aligned}$$

More precisely, consider the first term in J's expected payoff. The probability to observe $(H,H)$ is

$$P(H,H) = e^2(\mu p^2 + (1-\mu)(1-p)^2).$$

Given that the hidden evidence is $(H,H)$, J believes that the probability of $t = h$ is

$$\mu(H,H) = \frac{\mu p^2}{\mu p^2 + (1-\mu)(1-p)^2} > \frac{1}{2}.$$

Thus, J rules in favor of D, expecting to obtain

$$\mu(H,H) \times 1 + (1-\mu(H,H)) \times 0,$$

which is equal to $\mu(H,H)$. Multiplying $P(H,H)$ and $\mu(H,H)$ provides us with the first term, $e^2\mu p^2$. The other terms can be similarly understood.

Second, suppose that J consults only one expert.[17] The following are the possible situations:

$$r_{\mathrm{P}} = \begin{cases} H: & D \text{ wins}, \\ L: & P \text{ wins}, \\ \phi: & D \text{ wins } (\because \text{ posterior is equal to } \mu \geq 1/2). \end{cases}$$

[17] I denote the information from this expert as $r_{\mathrm{P}}$ without loss of generality.

Anticipating these results, J's expected payoff from consulting only one expert is

$$\begin{aligned}\pi_J^1 = &\underbrace{P(H)\mu(H)+P(L)(1-\mu(L))}_{\text{exp. payoff from observing one signal}}\\ &+ \underbrace{(1-e)\mu}_{\text{exp. payoff from observing no signal}}\\ &- \underbrace{c}_{\text{cost of expert advice}}\\ =& \, e(\mu p+(1-\mu)p)+(1-e)\mu-c,\end{aligned}$$

where $P(j)$ and $\mu(j)$ for $j \in \{H,L\}$ are as defined previously.

Finally, if J consults no experts, she simply rules in favor of D according to her prior belief, and therefore her expected payoff is given by

$$\pi_J^0 = \mu.$$

By comparing these expected payoffs, I can identify the conditions under which J consults two, only one, or no experts, which is summarized in the following proposition.

PROPOSITION 2 *There exist* $\underline{c}_J > 0$, $\bar{c}_J > 0$, *and* $\overline{\mu} \in (1/2, p)$ *such that the following is true:*

1. *When* $\mu \in [1/2, \overline{\mu})$, *the optimal number of experts for J is*

   0 *if* $\bar{c}_J < c$,
   1 *if* $c \in (\underline{c}_J, \bar{c}_J] \neq \emptyset$,
   2 *if* $c \leq \underline{c}_J$.

2. *When* $\mu \geq \overline{\mu}$, *the optimal number of experts for J is*

   0 *if* $(\underline{c}_J + \bar{c}_J)/2 < c$,
   2 *if* $c \leq (\underline{c}_J + \bar{c}_J)/2$.

The first part of the proposition presents an intuitive result: as information from experts is valuable, a lower cost induces J to consult more experts. In particular, if the cost lies in the intermediate range, J consults only one expert for information. On the other hand, the second part demonstrates that it is never optimal for J to consult only one expert under certain situations. The intuition is straightforward: if J's prior belief is sufficiently strong, information from only one expert is not persuasive enough, and J therefore wants to hear from at least two experts if she chooses to consult any expert.

## 5 Comparison

In this section, I compare the two institutional arrangements for expert testimony and establish two main results. First, I show that the no-expert cost threshold is higher in DI than in CI. In other words, the litigants consult an expert in DI even when J is reluctant to do so in CI when the cost of consulting an expert is high. This finding supports the claim by Posner (1988), who argues that one of the merits of using the decentralized procedure is the high initiative of the litigants in shaping the fact-finding process. Second, I show that, given the same number of experts consulted under both institutions, the final decision by J is more accurate in CI than in DI. This finding highlights the concerns echoed by Tullock (1988), who criticizes decentralized legal systems for production and presentation of misleading information by the litigants, to the detriment of the final verdict's accuracy.

### *5.1 Incentive to Consult Experts*

The following proposition demonstrates that the no-expert threshold is higher under both types of equilibria of DI than under CI.

PROPOSITION 3 *The no-expert threshold from the P-equilibrium in DI is higher than the thresholds in CI:* $\max\{\underline{c}_J, \bar{c}_J\} < \bar{c}$. *The same result holds for the D-equilibrium in DI and CI.*

To understand the intuition, supposing $\mu \geq 1/2$, it is instructive to compare the net benefit from consulting one expert rather than none under both institutions. In CI, J's net benefit from consulting one expert rather than none is given by[18]

$$e((1-\mu)p - \mu(1-p)). \tag{2}$$

The first term inside the parentheses, $(1-\mu)p$, is the probability of observing the low signal when the true state is low. Because J rules in favor of P upon observing the low signal, this is "good news" leading to correct decision-making. However, the second term inside the parentheses, $\mu(1-p)$, indicates "bad news" leading to an incorrect decision: this is the probability of observing the low signal when the true state is high. Because the low signal induces J to rule in favor of P, it generates errors, which reduces J's incentive to consult an expert.

In contrast, in DI, finding the low signal is always good news for P, whose net benefit from consulting an expert is given by[19]

$$e((1-\mu)p + \mu(1-p)). \tag{3}$$

As is obvious from the expression above, finding the low signal is always good news for P, because the low signal is favorable to his cause and he wants to win regardless of the true state. This effect increases a litigant's incentive to consult an expert relative to J's, and therefore an expert operates under a larger range of the cost parameter in DI than in CI.

[18] In the proof of Proposition 2, J's net benefit from consulting one expert rather than none is given by $\bar{c}_J$. After rearranging terms, $\bar{c}_J$ can be expressed as (2).

[19] In the proof of Proposition 1, P's net benefit from consulting an expert is given by $\bar{c}$. After rearranging terms, $\bar{c}$ can be expressed as (3).

The discussion above suggests that a litigant, who is a *partisan agent*, has a higher incentive to consult an expert than a trier of fact, who is an *impartial agent*. Related results are reported in the literature. In a setting with heterogeneous prior beliefs, Che and Kartik (2009) show that an agent whose prior belief is different from the decision-maker's has a stronger incentive to search for information, which induces the decision-maker to optimally hire such an agent despite communication problems. Whereas their model demonstrates that the decision-maker always prefers a partisan agent to a neutral one, my model identifies the conditions under which using a partisan agent (i.e., using DI) is better than using a neutral agent (i.e., using CI), and vice versa.

Dewatripont and Tirole (1999) ask related questions in a principal–agent setting in which an uninformed principal acquires information through agents before making a decision. Their main results show that using two agents (termed *advocacy*), each collecting information for a competing cause, generates information with lower agency costs than having one agent collect information for both competing causes (termed *nonpartisanship*). As the agents are rewarded based on the principal's final decision in their model (termed *decision-based rewards*), the agent in charge of conflicting tasks is reluctant to provide information for both causes, because if he does so, the two units of conflicting information will lead to the status quo, generating no payment to the agent. The principal does not have such a problem if she hires two agents and makes each agent a "partisan" to a cause, which generates the value of using a partisan agent in their model. Note that the agent under the nonpartisanship in their model is not impartial, in the sense that he wants to move the principal's decision away from the status quo. Thus, their main result is about a comparison between two different types of partisan preferences of the agents induced by the decision-based rewards, whereas Proposition 3 involves a comparison of the partisan and impartial preferences of the agents.

In contrast to these findings, Dur and Swank (2005) demonstrate that the bias of the agent may discourage his search effort in a soft-information framework. This is because when an agent recommends a policy to the decision-maker, a strongly biased agent makes a recommendation following his bias, not his information. Thus, as the bias of the agent increases, he values information less and therefore puts less effort into information collection. Note that they obtain this result because an agent's recommendation can be different from his information, which is possible under a soft-information framework. This finding suggests that the nature of information (i.e., hard versus soft) is an important factor in studying an agent's incentive for information search. For a general discussion regarding information search incentives, see Sobel (2013).

In general, a growing body of literature investigates the trade-off between the collection and communication of information. On the one hand, for better communication between an informed agent and an uninformed decision-maker, it is necessary to reduce the degree of conflict of interest between them. On the other hand, it is often observed that noncongruent preferences create incentives for agents to exert more effort for information. The current paper is in line with the existing literature in that it shows that a partisan agent has a higher incentive to consult an expert than an impartial agent, because the partisan agent's net benefit from additional information is higher.

### 5.2 *Information Loss from Evidence Distortion*

Both legal institutions, DI and CI, generate errors because J faces uncertainty in decision-making. To examine which system is better at reducing mistakes, I formally define the measure of errors as follows:

$$E = \mu\alpha + (1-\mu)\beta, \tag{4}$$

where $\alpha = P(\text{P wins} \mid t = h)$ is the probability that P wins despite $t = h$, and $\beta = P(\text{D wins} \mid t = l)$ is the probability that D wins despite $t = l$. Note that D's winning under $t = l$ and P's winning under $t = h$ are clearly incorrect decisions. In particular, considering $t = h$ as the "null hypothesis" and $t = l$ as the "alternative hypothesis," $\alpha$ and $\beta$ can be interpreted as Type I and Type II errors, respectively. With such an interpretation, the measure in (4) is the average of the two types of errors. In the subsequent analysis, I calculate $E$ from each legal institution and compare them.

Consider the cost range in which only one expert is consulted in both institutions. First, suppose $\mu \geq 1/2$. Then, the error from the P-equilibrium in DI is calculated as[20]

$$\begin{aligned} E^1_{\text{P}} &= \mu\alpha + (1-\mu)\beta \\ &= \mu P(\text{P wins} \mid h) + (1-\mu)P(\text{D wins} \mid l) \\ &= \mu e P(L \mid h) + (1-\mu)(1 - e + eP(H \mid l)) \\ &= \mu e(1-p) + (1-\mu)(1 - e + e(1-p)). \end{aligned}$$

More precisely, $\alpha$ is the probability that J incorrectly rules in favor of P. Note that only P consults an expert, and he wins if and only if he can present favorable evidence for his cause to J. Given $t = h$, such an event occurs with probability $eP(L \mid h)$, which is $\alpha$ in DI. Similarly, given $t = l$, D wins if and only if P cannot present favorable evidence to J. Thus, the probability for such an event is equal to $1 - e + eP(H \mid l)$, which is $\beta$ in DI.

The error in CI is given by

$$\begin{aligned} E^1_{\text{J}} &= \mu\alpha + (1-\mu)\beta \\ &= \mu P(\text{P wins} \mid h) + (1-\mu)P(\text{D wins} \mid l) \\ &= \mu e P(L \mid h) + (1-\mu)(1 - e + eP(H \mid l)) \\ &= \mu e(1-p) + (1-\mu)(1 - e + e(1-p)). \end{aligned}$$

---

20 The subscript $P$ in $E^1_{\text{P}}$ clarifies that this is the error from the P-equilibrium in DI.

By consulting only one expert, J observes $H$, $L$, or $\phi$ as a result of the expert's investigation. Note that D wins under $\phi$ because there is no evidence distortion in CI and therefore J's posterior belief under $\phi$ is equal to $\mu \geq 1/2$. Thus, P wins if and only if J observes $x = L$ from the expert, which implies $\alpha = eP(L \mid h)$ and $\beta = 1 - e + eP(H \mid l)$.

It is interesting to find that the two institutions generate the same number of mistakes, i.e., $E_{\mathrm{P}}^1 = E_{\mathrm{J}}^1$. The intuition is as follows. In DI, P distorts evidence submitted to J by suppressing unfavorable evidence for his cause. Thus, J only observes the low signal ($r_{\mathrm{P}} = L$) or nothing ($r_{\mathrm{P}} = \phi$) from P. If J observes the low signal, she "correctly" – in the sense that her decision is based on all the available evidence – rules in favor of P. If P remains silent, J reasons that there are two possibilities. First, if P is silent due to a manipulation motive (i.e., hiding $x_{\mathrm{P}} = H$), the correct ruling should be to rule in favor of D. Second, if P is silent simply because he is uninformed, J's posterior belief must be equal to $\mu \geq 1/2$, and therefore the correct ruling should be again to rule in favor of D. Thus, in any case, the optimal decision for J under the no-evidence event is to rule in favor of D, which is exactly what J does in the P-equilibrium of DI. This finding demonstrates that evidence distortion is not necessarily detrimental for the decision-making authority, at least when the decision is binary.

Second, suppose $\mu < 1/2$. If the P-equilibrium with P consulting an expert exists, its error takes the same formula as previously calculated. In contrast, the error in CI is given by

$$\begin{aligned} E_{\mathrm{J}}^1 &= \mu\alpha + (1-\mu)\beta \\ &= \mu P(\mathrm{P\ wins} \mid h) + (1-\mu)P(\mathrm{D\ wins} \mid l) \\ &= \mu(eP(L \mid h) + 1 - e) + (1-\mu)eP(H \mid l) \\ &= \mu(e(1-p) + 1 - e) + (1-\mu)e(1-p). \end{aligned}$$

When the prior belief is against D, the no-evidence event induces J to rule in favor of P. Thus, P wins unless J observes the high signal from the expert, which implies $\alpha = eP(L \mid h) + 1 - e$ and $\beta = eP(H \mid l)$. Since it immediately follows that $E_{\mathrm{J}}^1$ is smaller than $E_{\mathrm{P}}^1$ in this case, I obtain the following proposition.[21]

PROPOSITION 4 *Suppose that only one expert is consulted in both institutions.*

1. $\mu \geq 1/2$: *$E_P^1 = E_J^1$, and $E_D^1 > E_J^1$ if the D-equilibrium exists.*
2. $\mu < 1/2$: *$E_D^1 = E_J^1$, and $E_P^1 > E_J^1$ if the P-equilibrium exists.*

Although evidence distortion in the P-equilibrium of DI is not detrimental to the decision-making authority when $\mu \geq 1/2$, it is when $\mu < 1/2$. If P remains silent in the trial stage of DI, the P-equilibrium requires J to rule in favor of D. This decision is not optimal if P is silent due to lack of evidence, because in that case J's

[21] As the analysis for the D-equilibrium part is symmetric, I present the result without the proof.

posterior should be equal to $\mu < 1/2$, leading to P's winning. Thus, the impartiality of CI works to reduce decision-making errors relative to DI in such a situation.

Now consider the cost range in which two experts are consulted in both institutions. For the P-equilibrium (the D-equilibrium), this is possible only when $\mu \geq 1/2$ ($\mu < 1/2$). Let $E_{\mathrm{P}}^2$ ($E_{\mathrm{D}}^2$) and $E_{\mathrm{J}}^2$ denote the errors from the P-equilibrium (the D-equilibrium) in DI and CI, respectively. It turns out that when two experts are consulted in both systems, the decision-making error is always strictly smaller under CI because there is no evidence distortion in the system. To see this more clearly, consider the situations in which there is no direct evidence in the P-equilibrium. The event $(r_{\mathrm{D}}, r_{\mathrm{P}}) = (\phi, \phi)$ occurs under the following four possibilities: $(x_{\mathrm{D}}, x_{\mathrm{P}}) = (L, H)$, $(\phi, H)$, $(L, \phi)$, or $(\phi, \phi)$. For example, the "correct" decision under $(x_{\mathrm{D}}, x_{\mathrm{P}}) = (L, \phi)$ is to rule in favor of P. However, J is induced to rule in favor of D in such a situation because the litigants present $(r_{\mathrm{D}}, r_{\mathrm{P}}) = (\phi, \phi)$, under which D wins in the P-equilibrium. Thus, J cannot optimally make use of the available evidence because of the litigants' evidence distortion, which increases the error under DI.

Proposition 5 *Suppose that two experts are consulted in both institutions. If* $\mu \geq 1/2$, $E_P^2 > E_J^2$. *If* $\mu < 1/2$, $E_D^2 > E_J^2$.

These results suggest that the benefit of DI lies in the interested parties' high initiatives, which induce litigants to use expert information for a larger range of the cost parameter than J does in CI. However, the implicit cost of DI, other than the cost of experts, is an information loss due to evidence distortion by the litigants. If the same number of experts are consulted in both institutions, DI generates more mistakes than CI, due to evidence distortion by the litigants.

## 6 Discussion

### 6.1 Continuous Decision

The binary decision assumption is crucial in simplifying the analysis. If J's decision becomes continuous in DI, an immediate challenge is that checking the consistency of beliefs becomes a daunting task. To describe this point, let us suppose that J's optimal decision $d^*$ under $(r_{\mathrm{P}}, r_{\mathrm{D}})$ is equal to her posterior belief.[22] Then, the following are the four possible situations in the trial stage:

1. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (L, \phi)$: $d^* = \mu(L, \phi)$.
2. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (\phi, H)$: $d^* = \mu(\phi, H)$.
3. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (L, H)$: $d^* = \mu$.
4. $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (\phi, \phi)$: $d^* = \mu(\phi, \phi)$.

Compared to the basic model, there are two main changes in this extended formulation: *(i)* the magnitude of J's posterior belief becomes more important, and

22 That is, I assume that J's objective function takes the form of the quadratic function $-(d-t)^2$, where $d \in \mathbb{R}$ is J's decision and $t \in \{0, 1\}$ is the true state.

*(ii)* J's belief about the litigants' behavior in the pretrial stage becomes more important. For example, consider the first situation, in which J observes $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (L, \phi)$. In the basic model, J rules in favor of P, and her decision does not depend on the magnitude of her posterior belief. In contrast, in this extended formulation, J's decision crucially depends on the strength of her belief about the true state: if J strongly believes that the true state is in favor of P's claim, her decision becomes more favorable toward P. Furthermore, in contrast to the basic model, J's decision depends on J's belief about the litigants' behavior in the pretrial stage: $d^*$ can be high or low, depending on whether D also consulted an expert in the pretrial stage. This second effect was present only under $(r_{\mathrm{P}}, r_{\mathrm{D}}) = (\phi, \phi)$ in the basic model, but it operates under other report profiles as well in this extended formulation. I leave a more careful analysis of this extended model to future research.

### *6.2 Soft Information*

Another important assumption in the current model is that information is hard. Thus, the litigants in DI may conceal evidence if it is harmful to their causes, but they cannot falsify the evidence presented to J. Although models with hard information seem reasonable in a trial setting in which the falsification of evidence imposes large penalties upon the party, an interesting research area is to study the ways in which the possibility of falsification may affect the litigants' strategies along with the trial outcome. For example, see Emons and Fluet (2009a,b), who study a litigation game in which players may falsify their information by paying some cost.

The current model is not well suited to study the effect of soft information, because if information is soft, a litigant has no incentive to consult an expert: a litigant always wants to present favorable information to J in the trial stage, because he wants to win regardless of the true state, and therefore he does not need to consult an expert in the pretrial stage. In order to provide a litigant with an incentive to seek expert advice within the soft-information framework, the model may need to be extended in such a way that the litigant's preference depends on the true state.[23] In such a situation, the litigant wants to obtain knowledge about the true state before presenting any soft information to J, which generates the value of consulting an expert.

It is not clear whether the main results still hold in this soft-information framework. In particular, as discussed in section 5.1, in light of the work by Dur and Swank (2005) it is possible that a litigant's strong preference bias decreases his incentive to consult an expert. If that is so, the degree of verifiability of evidence at trial will be an important factor in the trade-off between the two institutions. A careful analysis of this issue awaits future research.

[23] For example, a litigant may ask for a high decision when the true state is moderate, whereas he may ask for a moderate decision when the true state is low. Such preferences may arise due to a litigant's moral concerns, which keep him from deviating too much from the true state.

### *6.3 Cost and Deterrence*

The focus of the main results in comparing the two institutions is the accuracy of J's decision. However, there are at least two other important characteristics of legal institutions: cost and deterrence.

First, let us consider the cost effect in comparing the institutions.[24] Proposition 3 suggests that for the high cost range, DI is likely to be superior to CI in accuracy because expert advice is utilized only in the former institution. As more expert advice means more information for J's decision-making, leading to higher accuracy, DI is expected to perform better than CI as far as accuracy is concerned. However, as more information from expert advice can be obtained only by spending more resources for consulting an expert, the litigants' strong incentive to obtain information is not necessarily beneficial for society. In light of this trade-off between accuracy and cost, the societal preference over legal outcomes becomes important: if a society attaches more value to accuracy, it may prefer DI to CI; otherwise, it may prefer CI to DI.[25] In contrast, Propositions 4 and 5 suggest that we need not be concerned about such a trade-off for the low cost range. As expert advice is expected to be utilized in both institutions, the evidence distortion problem in DI decreases the system's accuracy relative to CI, in which such a problem does not exist. Thus, if the same amount of expert advice is used in both institutions, CI is superior to DI regardless of the cost consideration, because a higher level of accuracy can be achieved in CI at the same cost as in DI. This discussion suggests that the cost consideration operates in favor of CI in the current model.

A related issue is the effect of the rule that requires the litigants, rather than J, to pay the cost in CI. The main results do not change under this rule if J takes into account the cost borne by the litigants. If J does not consider the costs of expert advice, she will always consult two experts in CI regardless of the cost parameter, because expert advice is free information for J. This change could increase the accuracy of J's final decision at the expense of higher costs borne by the litigants, exhibiting the trade-off discussed above.

Second, let us consider how the two institutions perform differently in terms of deterrence. Deterrence is intimately related to accuracy, because the trial outcomes influence an individual's choice of the primary behavior. Following Kaplow (1994),

[24] Posner argues that accuracy and cost are the two most important criteria in comparing legal systems (Posner, 1999, p.1542).

[25] Thus, the existence of different legal institutions may reflect preference differences across societies. Kaplow (1994, pp. 307–308) notes that "[o]ne might go so far as to say that a large portion of the rules of civil, criminal, and administrative procedure and rules of evidence involve an effort to strike a balance between accuracy and legal costs." Presumably, in pursuit of such a balance, certain societies might have embraced a decentralized way of solving information provision problems, whereas others have adopted a centralized system. Thus, the current form of legal institutions in a society could be indicative of the preference of the society. In this vein, Demougin and Fluet (2005) conclude, studying the variation in the standard of proof across societies, that common-law countries are more concerned with deterrence than accuracy whereas civil-law countries attach a greater weight to accuracy.

who argues that one benefit of accuracy is its deterrence effect, one could argue for a positive association between accuracy and deterrence: a higher level of accuracy is associated with a higher level of deterrence. Then, DI is expected to increase deterrence relative to CI for the high cost range (Proposition 3), whereas CI is more likely to perform better in terms of deterrence for the low cost range (Propositions 4 and 5). However, there is also a possibility of tension in pursuing these two legal outcomes simultaneously. For example, in a series of influential articles, Demougin and Fluet (2005, 2006, 2008) demonstrate that the common-law rules of proof maximize deterrence at the expense of accuracy. Investigating the trade-off among different legal outcomes will be a fruitful future research topic.

### *6.4 Criminal versus Civil Cases*

In the main results, I assume that society is equally averse to both types of errors made by J. In criminal cases, however, society is typically more averse to Type I errors, wrongly convicting the innocent, than to Type II errors, wrongly acquitting the guilty. Thus, in general, the measure of accuracy can be defined as

$$E = \mu\lambda\alpha + (1-\mu)\beta,$$

where $\lambda > 0$ measures the relative weight of Type I errors. In this extended formulation, criminal cases can be identified with $\lambda > 1$.

To understand how this change may affect the main results, consider $E^1_{\mathrm{P}}$ and $E^1_{\mathrm{J}}$. If $\mu \geq 1/2$, I still have $E^1_{\mathrm{P}} = E^1_{\mathrm{J}}$, because Type I errors (and Type II errors as well) under both institutions are the same. However, if $\mu < 1/2$ (assuming the existence of the P-equilibrium in DI), we have

$$\begin{aligned} E^1_{\mathrm{P}} &= \mu\lambda\underbrace{e(1-p)}_{\text{Type I}} + (1-\mu)(\underbrace{1-e}_{(A)} + e(1-p)), \\ E^1_{\mathrm{J}} &= \underbrace{\mu\lambda(e(1-p) + \underbrace{1-e}_{(B)})}_{\text{Type I}} + (1-\mu)e(1-p), \end{aligned}$$

where Type I errors are larger in CI. Thus, if society is sufficiently averse to Type I errors (i.e., $\lambda$ is large), I obtain $E^1_{\mathrm{P}} < E^1_{\mathrm{J}}$, in contrast to the previous result.

It is interesting to find that, when the burden of proof is on P as in typical criminal cases, the decentralized way of providing information to the fact-finder could generate fewer mistakes. This result follows from different BOP allocations across legal institutions. In the P-equilibrium of DI, P loses when his expert has no evidence (with probability $1-e$ in $(A)$), because he has the BOP. In contrast, in CI, it is D who loses when J's expert fails to obtain hard evidence (with probability $1-e$ in $(B)$). Thus, the "implicit" BOP falls on D in CI, although neither litigant explicitly bears the BOP, because J directly interacts with experts.

## 7 Conclusion

Within the framework of a persuasion game with endogenous information, this paper examines the relative merits of two institutions, CI and DI. The main results demonstrate that there is a trade-off: although DI supplies the fact-finder with valuable information more often, it also suffers from an information loss due to its competitive nature.

The analysis suggests that the ranking of the two institutions in terms of accuracy depends on the cost of consulting an expert. If the cost is large, the decision-making accuracy is expected to be higher in DI than in CI, because expert information is utilized only in the former institution. In contrast, CI is expected to be superior when the cost is small: if the same amount of expert information is utilized in the two systems, the decision-making accuracy is expected to be higher in CI because there is no information loss in the system.

Although proponents for policy reforms who encourage the trier of fact to appoint her own experts raise valid concerns, one should keep in mind that the cost of using expertise may affect the system's performance. If it is costly to make use of the knowledge possessed by experts in specific domains, society may observe a *decline* in the usage of expert information in trial courts as a result of policy reforms, which could lead to less-accurate decision-making by judges.

## Appendix

### A.1 Proof of Proposition 1

The proof consists of two steps. First, taking J's equilibrium belief as given, I find the players' equilibrium strategies. Second, I verify whether J's equilibrium belief is indeed consistent with the players' equilibrium strategies found in the first step. As the proof builds on the analysis from sections 3.1 and 3.2, I reproduce the main results of those subsections here as lemmas:

Lemma A1 (section 3.1) *In the trial stage, the following is each player's behavior:*

1. *P only reports $x_P = L$ whenever possible.*
2. *D only reports $x_D = H$ whenever possible.*
3. *J's decision is given by*

$$(r_P, r_D) = \begin{cases} (L,\phi): & P \text{ wins}, \\ (\phi,H): & D \text{ wins}, \\ (L,H): & D \text{ wins if and only if } \mu \geq 1/2, \\ (\phi,\phi): & D \text{ wins if and only if } \mu(\phi,\phi) \geq 1/2. \end{cases}$$

LEMMA A2 (SECTION 3.2) *Assume the BOP falls on P. In the pretrial stage, the following is each player's behavior:*

1. *If $\mu \geq 1/2$, there exists a pair $(c_P^P, c_D^P)$ such that*

   *P consults an expert if and only if $c \leq c_P^P$, and*
   *D consults an expert if and only if $c \leq c_D^P$,*

   *where (i) $c_P^P$ and $c_D^P$ depend on the litigants' choices, and (ii) $c_D^P = 0$ when P does not consult an expert.*
2. *If $\mu < 1/2$, D does not consult an expert, and there exists $c_P^P$ such that P consults an expert if and only if $c \leq c_P^P$.*

### A.1.1 Step 1: Litigants' Equilibrium Strategies in the Pretrial Stage

When the BOP is on P, Lemma A2 demonstrates that three cases are possible: neither litigant consults experts, P alone consults an expert, or both consult experts. In particular, D is never willing to consult an expert alone. The number of consulted experts with the BOP on P depends on parameter values. To simplify the notation, let us define the following quantities:

$$c_1 = \mu e(1-p) + (1-\mu)ep,$$
$$c_2 = \mu e(1-p) + (1-\mu)ep - \{\mu e(1-p)(1-ep) + (1-\mu)ep(1-e(1-p))\},$$
$$c_3 = \mu e(1-p)(1-ep) + (1-\mu)ep(1-e(1-p)),$$

where (considering $\mu \geq 1/2$ for interpretation)[26]

$c_1$ is P's net benefit from expert advice when D does not consult an expert,
$c_2$ is D's net benefit from expert advice when P consults an expert, and
$c_3$ is P's net benefit from expert advice when D consults an expert.

Having defined these quantities, I can rank them according to their magnitudes. It is easy to show $\max\{c_2, c_3\} < c_1$. The following lemma shows $c_2 < c_3$:

LEMMA A3 $c_2 < c_3$.

PROOF Rearranging terms, I obtain

$$c_2 < c_3 \iff e < \frac{\mu - 2\mu p + p}{2p(1-p)} \equiv \hat{e}.$$

[26] When $\mu < 1/2$, P's net benefit from expert advice is $c_1$ regardless of D's choice, and D's net benefit from expert advice is 0 regardless of P's choice.

Observe that $\hat{e}$ is positive. The denominator of $\hat{e}$ is positive because $p \in (1/2, 1)$. The numerator of $\hat{e}$ is also positive because

$$0 \leq (\sqrt{\mu} - \sqrt{p})^2 = \mu - 2\sqrt{\mu p} + p < \mu - 2\mu p + p,$$

where the last inequality holds because $\mu p$ is a fraction.

To prove the lemma, it is sufficient to show $\hat{e} > 1$. To this end, let us define $g(p)$ as

$$\begin{aligned} g(p) &= \underbrace{\mu - 2\mu p + p}_{\text{numerator of } \hat{e}} - \underbrace{2p(1-p)}_{\text{denominator of } \hat{e}} \\ &= 2p^2 - 2\mu p - p + \mu. \end{aligned}$$

This function is an increasing function for $p \in (1/2, 1)$ because $g'(p) > 0$. As $g(1/2) = 0$, I conclude that $g(p) > 0$ for $p \in (1/2, 1)$. This completes the proof.

The litigants' behavior in the pretrial stage depends on the cost of using an expert. Assume $\mu \geq 1/2$. If $c > c_1$, then even when D does not use an expert, P's net benefit from using an expert is less than the cost. Thus, neither litigant consults experts for information. If $c \in (c_2, c_1]$, it is straightforward to show that only P consults an expert. If $c \leq c_2$, both litigants consult an expert: D is willing to consult an expert when P consults an expert; because $c \leq c_2 < c_3$, the cost also rationalizes P's choice of consulting an expert. Thus, both litigants consult an expert when $c \leq c_2$.

Now assume $\mu < 1/2$. In this case, D is never willing to consult an expert. Therefore, the only litigant who may consult an expert is P, and his choice depends on whether the net benefit of consulting an expert is larger than the cost of doing so. Thus, if $c \leq c_1$, P consults an expert, and he does not do so otherwise. The following lemma summarizes these findings.

LEMMA A4 *Suppose that the BOP is on P in equilibrium. Then, the following are the litigants' equilibrium strategies in the pretrial stage.*

*If* $\mu \geq 1/2$,

$c_1 < c$: *neither litigant consults an expert,*
$c \in (c_2, c_1]$: *only P consults an expert,*
$c \leq c_2$: *both litigants consult an expert.*

*If* $\mu < 1/2$,

$c_1 < c$: *neither litigant consults an expert,*
$c \leq c_1$: *only P consults an expert.*

### A.1.2 Step 2: Verifying Consistency of J's Equilibrium Belief

In the following, I examine whether the litigants' equilibrium strategies from Lemma A4 are consistent with the BOP's being on P. As before, I separate the analysis into two parts, $\mu \geq 1/2$ and $\mu < 1/2$. I begin with the first part.

*Prior in Favor of D.* Assume $\mu \geq 1/2$. It turns out that any number of experts consulted by the litigants is consistent with the P-equilibrium, and therefore the P-equilibrium always exists:

1. If neither or both of the litigants consult an expert for evidence, I have $\mu(\phi,\phi) = \mu \geq 1/2$, which is consistent with the BOP's being on P.

2. If only P consults an expert, I have $\mu(\phi,\phi) > \mu \geq 1/2$, which is also consistent with the BOP's being on P. These findings are summarized in the following lemma:

LEMMA A5 *If $\mu \geq 1/2$, the P-equilibrium always exists, and*

> $c_1 < c$*: neither litigant consults an expert in the P-equilibrium,*
> $c \in (c_2, c_1]$*: only P consults an expert in the P-equilibrium,*
> $c \leq c_2$*: both litigants consult an expert in the P-equilibrium.*

Letting $\bar{c} \equiv c_1$ and $\underline{c} \equiv c_2$ proves the first part of Proposition 1.

*Prior in Favor of P.* Now assume $\mu < 1/2$. If $c > c_1$, P and D do not consult an expert for evidence, and therefore they present nothing to J. In equilibrium, J correctly anticipates the litigants' behavior, and this implies that $\mu(\phi,\phi) = \mu < 1/2$. Thus, the BOP cannot fall on P in this case, and therefore there is no P-equilibrium.

If $c \leq c_1$, P consults an expert but D does not. Thus, under the no-evidence event, J's belief is updated upward, i.e., $\mu(\phi,\phi) > \mu$. If the P-equilibrium is to exist, this updating must be enough so that $\mu(\phi,\phi)$ becomes larger than $1/2$ in spite of $\mu < 1/2$. This is possible if $\mu$ is large (i.e., close to $1/2$) or $e$ is large (i.e., close to 1). Note that $\mu(\phi,\phi)$ in this case is given by (1), which can be easily verified to be increasing in $\mu$ and $e$. Because I have $\mu(\phi,\phi) > 1/2$ for $\mu = 1/2$, by continuity I also have $\mu(\phi,\phi) > 1/2$ when $\mu$ is sufficiently close to $1/2$. Also, it is straightforward to obtain $\mu(\phi,\phi) > 1/2$ from (1) for $e = 1$. Thus, again by continuity, I have $\mu(\phi,\phi) > 1/2$ for sufficiently large $e$.

These findings are summarized in the following lemma:

LEMMA A6 *If $\mu < 1/2$,*

> $c_1 < c$*: the P-equilibrium does not exist,*
> $c \leq c_1$*: the P-equilibrium, in which only P consults an expert, exists if $\mu$ is close to $1/2$ or $e$ is close to 1.*

Letting $\bar{c} \equiv c_1$ proves the second part of Proposition 1.

### *A.2 Proof of Proposition 2*

J consults one expert rather than neither if $\pi_J^1 \geq \pi_J^0$, or equivalently, if the cost is less than the net benefit of consulting an expert:

$$c \leq \bar{c}_J = e(p - \mu).$$

Similarly, if the cost is such that

$$c \leq \underline{c}_J = (1-e)\cdot \underbrace{e(p-\mu)}_{(A)>0} + e\cdot \underbrace{ep(1-p)(2\mu-1)}_{(B)\geq 0 \because \mu \geq 1/2},$$

then J consults two experts rather than only one.

Lastly, if the cost is such that

$$c \leq (\underline{c}_J + \bar{c}_J)/2,$$

then J consults two experts rather than none.[27]

Rearranging $\underline{c}_J$ and $\bar{c}_J$, I obtain

$$\underline{c}_J < \bar{c}_J \iff \mu < \frac{2p-p^2}{1+2p(1-p)} \equiv \overline{\mu},$$

where it is straightforward to show $\bar{\mu} \in (1/2, p)$.

First, consider the case of $\underline{c}_J \geq \bar{c}_J$. If $c > \bar{c}_J$, no expert is better than one expert for J. If $c \leq \bar{c}_J$, one expert is better than no expert, and two experts are better than one expert because $c \leq \bar{c}_J \leq \underline{c}_J$. Thus, J never consults only one expert in this case, and therefore the only issue for J is whether to consult two experts or none. Hence, J consults two experts when $c \leq (\underline{c}_J + \bar{c}_J)/2$ and she consults no expert otherwise, which proves the second part of the proposition. The first part can also be proved similarly.

### *A.3 Proof of Proposition 3*

Let us compare the no-expert threshold from the P-equilibrium in DI and the thresholds from CI. Consider $\mu \geq 1/2$. Comparing $\bar{c}$ and $\bar{c}_J$, I obtain

$$\bar{c} = c_1 = e((1-\mu)p + \mu(1-p)) > e((1-\mu)p - \mu(1-p)) = e(p-\mu) = \bar{c}_J.$$

Subtracting $\underline{c}_J$ from $\bar{c}$, I obtain

$$\bar{c} - \underline{c}_J = e(\underbrace{2\mu(1-p)(1-ep)}_{(+)}) + e(\underbrace{p-p^2}_{(+)} + \underbrace{p-\mu}_{(+)}) > 0.$$

Next, consider $\mu < 1/2$. In this case, $\bar{c}_J$ and $\underline{c}_J$ are given by

$$\begin{aligned}
\bar{c}_J &= e(\mu p - (1-\mu)(1-p)), \\
\underline{c}_J &= e^2(\mu p^2 + 2p(1-p)(1-\mu) + (1-\mu)p^2) + 2e(1-e)(\mu p + (1-\mu)p) \\
&\quad + (1-e)^2(1-\mu) - e(\mu p + (1-\mu)p) - (1-e)(1-\mu).
\end{aligned}$$

[27] The derivation of these cutoffs is straightforward, so I omit the details. See the working-paper version of this article for the details.

Subtracting $\bar{c}_J$ from $\bar{c}$, I obtain

$$\bar{c} - \bar{c}_J = e(1-2\mu p) > 0.$$

Subtracting $\underline{c}_J$ from $\bar{c}$, I obtain

$$\bar{c} - \underline{c}_J = e(\underbrace{(1-e)}_{(+)}\underbrace{(1-2\mu p)}_{(+)} + ep^2\underbrace{(1-2\mu)}_{(+)} + e\mu) > 0.$$

Thus, $\bar{c}$ from the P-equilibrium in DI is higher than $\bar{c}_J$ and $\underline{c}_J$ in CI. As the proof for the D-equilibrium part is completely symmetric, this completes the proof.

### *A.4 Proof of Proposition 5*

First, assume $\mu \geq 1/2$. In the P-equilibrium with two experts, the error can be calculated as

$$\begin{aligned} E_P^2 &= \mu P(\text{P wins} \mid h) + (1-\mu)P(\text{D wins} \mid l) \\ &= \mu e(1-p)(1-e+e(1-p)) + (1-\mu)(1-ep(1-e+ep)). \end{aligned}$$

If J consults two experts in CI, the error is given by

$$\begin{aligned} E_J^2 &= \mu P(\text{P wins} \mid h) + (1-\mu)P(\text{D wins} \mid l) \\ &= \mu(e^2(1-p)^2 + 2e(1-e)(1-p)) + (1-\mu)(1-e^2p^2 - 2e(1-e)p). \end{aligned}$$

Then, subtracting $E_J^2$ from $E_P^2$, I obtain

$$E_P^2 - E_J^2 = (p-\mu)e(1-e) > 0.$$

Since the proof for the D-equilibrium part is symmetric, this completes the proof.

## *References*

Adrogué, S., and A. Ratliff (2003), “The Independent Expert Evolution: From the ‘Path of Least Resistance’ to the ‘Road Less Traveled?’ ” *Texas Tech Law Review*, 34(4), 843–897.

Che, Y.-K., and N. Kartik (2009), “Opinions as Incentives,” *Journal of Political Economy*, 117(5), 815–860.

Deffains, B., and D. Demougin (2008), “The Inquisitorial and the Adversarial Procedure in a Criminal Court Setting,” *Journal of Institutional and Theoretical Economics (JITE)*, 164(1), 31–43.

Demougin, D., and C. Fluet (2005), “Deterrence versus Judicial Error: A Comparative View of Standards of Proof,” *Journal of Institutional and Theoretical Economics (JITE)*, 161(2), 193–206.

— and — (2006), “Preponderance of Evidence,” *European Economic Review*, 50(4), 963–976.

— and — (2008), “Rules of Proof, Courts, and Incentives,” *The RAND Journal of Economics*, 39(1), 20–40.

Dewatripont, M., and J. Tirole (1999), “Advocates,” *Journal of Political Economy*, 107(1), 1–39.

Dur, R., and O. H. Swank (2005), “Producing and Manipulating Information,” *The Economic Journal*, 115(500), 185–199.

Emons, W., and C. Fluet (2009a), “Accuracy versus Falsification Costs: The Optimal Amount of Evidence under Different Procedures,” *The Journal of Law, Economics, & Organization*, 25(1), 134–156.

— and — (2009b), “Adversarial versus Inquisitorial Testimony,” Discussion Paper 09-04, Department of Economics, University of Bern.

Froeb, L. M., and B. H. Kobayashi (1996), “Naive, Biased, yet Bayesian: Can Juries Interpret Selectively Produced Evidence?” *The Journal of Law, Economics, & Organization*, 12(1), 257–276.

— and — (2001), “Evidence Production in Adversarial vs. Inquisitorial Regimes,” *Economics Letters*, 70(2), 267–272.

Gross, S. R. (1991), “Expert Evidence,” *Wisconsin Law Review*, 1113–1232.

Hand, L. (1901), “Historical and Practical Considerations Regarding Expert Testimony,” *Harvard Law Review*, 15(1), 40–58.

Hillman, T. (2002), “Using Court-Appointed Experts,” *New England Law Review*, 36(3), 587–591.

Iossa, E., and G. Palumbo (2007), “Information Provision and Monitoring of the Decision-Maker in the Presence of an Appeal Process,” *Journal of Institutional and Theoretical Economics (JITE)*, 163(4), 657–682.

Kaplan, L. A. (2006), “2005 Milton Handler Antitrust Review: Experts in the Courthouse: Problems and Opportunities,” *Columbia Business Law Review*, (2), 247–260.

Kaplow, L. (1994), “The Value of Accuracy in Adjudication: An Economics Analysis,” *The Journal of Legal Studies*, 23(1), 307–401.

Kim, C. (2014a), “Adversarial and Inquisitorial Procedures with Information Acquisition,” *The Journal of Law, Economics, & Organization*, 30(4), 767–803.

— (2014b), “Partisan Advocates,” *Bulletin of Economic Research*, 66(4), 313–332.

— (2015), “An Economic Rationale for Dismissing Low-Quality Experts in Trial,” Working Paper, Munich Personal RePEc Archive (MPRA), https://mpra.ub.uni-muenchen.de/69620/, last accessed June 3, 2016.

— (2016), “Adversarial Bias, Litigation, and the *Daubert* Test: An Economic Approach,” Working Paper, Munich Personal RePEc Archive (MPRA), https://mpra.ub.uni-muenchen.de/69978/, last accessed June 3, 2016.

Milgrom, P., and J. Roberts (1986), “Relying on the Information of Interested Parties,” *The RAND Journal of Economics*, 17(1), 18–32.

Palumbo, G. (2001), “Trial Procedures and Optimal Limits on Proof-Taking,” *International Review of Law and Economics*, 21(3), 309–327.

— (2006), “Optimal Duplication of Effort in Advocacy Systems,” *Journal of Economic Behavior & Organization*, 60(1), 112–128.

Parisi, F. (2002), “Rent-Seeking through Litigation: Adversarial and Inquisitorial Systems Compared,” *International Review of Law and Economics*, 22(2), 193–216.

Posner, R. A. (1988), “Comment: Responding to Gordon Tullock,” in: S. S. Nagel (ed.), *Research in Law and Policy Studies: Volume 2*, JAI Press, Greenwich (CT), pp. 29–33.

— (1999), “An Economic Approach to the Law of Evidence,” *Stanford Law Review*, 51(6), 1477–1546.

Robertson, C. T. (2010), “Blind Expertise,” *New York University Law Review*, 85, 174–257.

Runkle, D. (2001), “Court Appointed Scientific Experts: A Demonstration Project of the American Association for the Advancement of Science,” *Risk Decision and Policy*, 6(2), 139–143.

Sharif, Z., and O. H. Swank (2012), “Do More Powerful Interest Groups Have a Disproportionate Influence on Policy?” Discussion Paper TI 2012-134/VII, Tinbergen Institute.

Shin, H. S. (1998), “Adversarial and Inquisitorial Procedures in Arbitration,” *The RAND Journal of Economics*, 29(2), 378–405.

Sidak, J. G. (2013), “Court-Appointed Neutral Economic Experts,” *Journal of Competition Law & Economics*, 9(2), 359–394.

Sobel, J. (2013), “Giving and Receiving Advice,” in: D. Acemoglu, M. Arellano, and E. Dekel (eds.), *Advances in Economics and Econometrics: Tenth World Congress, Volume I: Economic Theory*, Cambridge University Press, New York, pp. 305–341.

Tullock, G. (1975), “On the Efficient Organization of Trials,” *Kyklos*, 28(4), 745–762.

— (1980), *Trials on Trial: The Pure Theory of Legal Procedure*, Columbia University Press, New York.

— (1988), “Defending the Napoleonic Code over the Common Law,” in: S. S. Nagel (ed.), *Research in Law and Policy Studies*, vol. 2, JAI Press, Greenwich (CT), pp. 3–27.

Chulyoung Kim
School of Economics
Yonsei University
50 Yonsei-ro Seodaemun-gu
Seoul 03722
Republic of Korea
chulyoung.kim@gmail.com

# Prohibition versus Taxation in Corrupt Environments

by

Desiree A. Desierto and John V. C. Nye*

Received November 8, 2013; in revised form December 27, 2015;
accepted January 6, 2016

If corruption is rife and tolerated by society, prohibiting the production of a good with negative social costs may be more efficient at limiting quantity than legalizing and taxing producers. It becomes incentive-compatible for a corrupt government to enforce prohibition and credibly limit supply in order to extract bribes from illegal producers. In equilibrium, total quantity is low. In contrast, when the good is legal and taxed, a corrupt government can extract rents only by expropriating the tax revenues. Thus, it prefers a larger market in order to generate more taxes, and quantity is higher. (JEL: D21, D23, H41)

## *1 Introduction*

Proponents of the legalization of drugs and other contraband argue that social costs are high when such goods are illegal because of two things – the proliferation of the good itself, which generates negative externalities to consumers, and the concomitant corruption and rent-seeking of law enforcers, which inefficiently allocates producers' resources toward the payment of bribes. This paper shows, however, that these two wrongs actually make a right – that corruption in an illegal market helps to keep social costs down, and makes prohibition more effective than taxation in lowering production of the good.

This is because making a good or service illegal gives an additional incentive to a corrupt government to crack down on producers in order to elicit bribes, whereas there is no such incentive when the good is legal, even if the government were to tax producers and expropriate the tax revenues. In that case, the government would actually prefer a larger market, since this would generate more tax revenues

* D. Desierto (corresponding author): Department of Political Science, University of Wisconsin–Madison, U.S.A.; J. Nye: Economics Department, George Mason University, Fairfax (VA), U.S.A., and National Research University – Higher School of Economics, Moscow, Russian Federation. We wish to thank John Nachbar, Norman Schofield, Kenneth Shepsle, Alex Possajennikov, Emmanuel De Dios, Raul Fabella, anonymous referees, and various conference and seminar participants for their comments and suggestions.

*Journal of Institutional and Theoretical Economics* 173, 239–252 – ISSN 0932-4569
DOI: 10.1628/093245616X14618208477808 – 

and allow greater expropriation. If the good is instead prohibited, the government's gains from corruption are contingent on the credible enforcement of prohibition rules – illegal producers will not want to keep paying bribes if they can readily supply the market. The government thus wants to ensure that total quantity is limited.

The crucial assumption is that the government is able to maximize the total bribes it receives from illegal producers, which, in turn, is possible if society tolerates a corrupt government. In this environment, the bribes that producers spend are not a deadweight loss, since they are merely transferred to the government. Even if the good generates negative consumption externalities, total social costs are kept to a minimum, since there are effectively no production externalities.

Glaeser and Shleifer (2001) also provide an argument for prohibition, but without explicitly considering corruption and bribe-taking by law enforcers. They show that prohibition can be more efficient than taxation, essentially because it is easier to detect violators of prohibition rules than of taxation rules. That is, mere possession of the illegal good is already evidence of violation, which allows even ordinary citizens to report such violations, whereas payment of taxes may not be directly verified by those ordinary citizens. In this sense, the enforcement of prohibition rules is less costly than taxation and hence more efficient.

We get to the same result, but by the opposite route, since we focus on the government's private gains from rent-seeking. In corrupt environments, violations of prohibition become *harder* to detect, not only because illegal producers and corrupt enforcers effectively cooperate in exchanging bribes, but also because consumers tolerate such exchanges. When consumers can already legally obtain, and producers legally supply, the good, there is less reason for them not to report instances of rent-seeking by the government. In effect, the expropriation of tax revenues in a legalized market is easier to detect. Because rent-seeking is harder to detect (or easier to hide) in illegal than in legal markets, it is easier for a corrupt government to do its job of enforcing prohibition rules in order to obtain rents.

The foregoing suggests why prohibition, and not legalization with taxation, has been the persistent mode of control of socially undesirable goods. Indeed, Desierto and Nye (2011) note that in a sample of 101 countries, 100 prohibit drugs, 66 prohibit prostitution, and 33 prohibit gambling, and that corruption is higher on average for those countries that prohibit. Yet, until this paper, there has been no positive theory that can explain why states that are more corrupt are the ones more likely to keep undesirable goods illegal, while it is typically advanced states with low corruption that are more likely to move to legalizing and taxing those same products. Our paper precisely shows that corrupt states find it optimal to prohibit such goods.

Nevertheless, there has been considerable work in the literature showing the inefficiency of quantity regulation in comparison with excise taxes, and of the ineffectiveness of prohibition in curtailing consumption (see, e.g., Weitzman, 1974; Miron, 2004, 2008; and Becker, Murphy, and Grossman, 2006 – henceforth, BMG). BMG, in particular, argue that if demand for the good is inelastic, the social costs of enforcing prohibition rules are high, since producers will waste resources on avoiding punishment in order to keep filling the inelastic demand. Taxation would thus be a better way of reducing consumption, as this precludes the *avoidance costs* of enforcement.

Our paper uses the basic BMG framework, which, after all, implicitly allows the avoidance cost of producers to include bribes. The main difference, however, is that we consider situations in which bribe-taking is widely tolerated so that the government can openly maximize total bribes. More specifically, the social welfare that the government maximizes in order to calculate the optimal level of enforcement is composed not just of the private value to producers and the value to consumers (including any externalities), but also of the value of the bribes that accrue to the government.

Thus, when one really considers the true political economy in the market for illegal goods, one can explain why the seemingly second-best alternative of keeping a good illegal and punishing offenders persists as an equilibrium. It may be that what one naively views as a waste of resources is not really a complete waste; that anticrime efforts do pay, albeit some of the benefits are illegally gotten, and relative to these efforts, the legalization alternative may incur even greater waste.

The rest of the paper is organized as follows. Section 2 builds on BMG to take into account the effect of bribe-taking by prohibition enforcers, while section 3 provides an analogous model for a legalized market in which producers are taxed. Section 4 evaluates some existing empirical work on illegal drugs in the light of our model, and section 5 concludes.

## 2 *Prohibition*

The setup of the model follows BMG. Let $E$ be the intensity per unit of output with which prohibition rules are enforced against an illegal producer, and $A$ be the unit avoidance cost – the amount of resources that the producer spends per unit in order to bring the illegal good to the market. The producer chooses $A$ and the government chooses $E$ simultaneously.

Where we initially depart from BMG is in distinguishing between two types of avoidance costs – bribes to the government, and all other kinds of spending to avoid getting caught, like the adoption of systems and technologies, the maintenance of secure and hidden facilities, hiring of legal and other consultants, etc. Thus, let $A = \beta A + (1-\beta)A$, where $0 \leq \beta \leq 1$ is the fraction spent as bribes and, as such, is an indicator of how corrupt the government is.

Taking $E$ as given, the producer is caught with probability $p(E,A)$, which is assumed to increase with $E$ and decrease with $A$. If caught, the producer has to pay a per-unit fine $F$. Letting $c$ denote the cost of producing a unit, the expected unit cost $u$ of bringing an illegal good to the market is thus

$$u = \frac{c + A + p(E,A)F}{1 - p(E,A)}, \tag{1}$$

where $F$ is weighted by $p$, but $A$ is not, since $A$ is incurred for every unit produced, while $F$ is incurred only if the unit is detected.

Using the odds ratio

$$\theta(E,A) = \frac{p(E,A)}{1-p(E,A)},$$

equation (1) can be expressed as

$$u = (c+A)(1+\theta) + \theta F,$$

where $u$ is linear in $\theta$. Minimizing this expected cost and restricting attention to interior solutions give the following first-order condition (FOC):

$$-\frac{\partial\theta}{\partial A}(c+A+F) = 1+\theta, \tag{2}$$

which can be implicitly solved to get the optimal avoidance cost $A^* > 0$.

Following BMG, we assume a competitive market and constant returns to scale (CRS) technology, which implies that the good is sold at a price equal to the minimum expected cost:[1]

$$P(E) = (c+A^*)[1+\theta(E,A^*)] + \theta(E,A^*)F. \tag{3}$$

Meanwhile, taking $A$ as given, the government chooses enforcement $E$. The main difference with BMG is that we capture the case in which corruption is so pervasive that it enjoys the complicity of the entire bureaucratic apparatus.[2] In this environment where all producers and all government agents tolerate corruption, the government can then openly maximize total bribes, along with the total welfare of consumers and producers:

$$\max_E W = V[Q(E)] + \beta A^*(E)Q(E) - P(E)Q(E) - C\{Q(E), E, \theta[E, A^*(E)]\}, \tag{4}$$

where $Q$ is the quantity supplied by illegal producers, which is a function of the strength of enforcement, $V[\cdot]$ the value to consumers (net of all consumption externalities), $\beta AQ$ total bribes, $PQ$ the cost to producers, and $C\{\cdot\}$ the cost the government incurs to enforce prohibition.

1 It would be interesting to consider other forms of industrial organization. For instance, if some avoidance costs were fixed, only a small number of large producers might survive as the cost became prohibitively high. Less competition among producers would lead to higher prices and lower quantity, but the elasticity of quantity with respect to enforcement could also be smaller. We thank an anonymous referee for pointing this out.

2 An example is given by *jueteng*, a form of gambling in the Philippines, which is illegal but rampant. It has been alleged that most (if not all) of jueteng is syndicated, with backing from top government officials who stand to lose significant revenues from legalization, even including the former President, Joseph Estrada. (See Lambsdorff (2007), Pamintuan (2010), and, for an expose by a Philippine senator, http://www.senate.gov.ph/press_release/2010/0922_santiago1.asp.)

As in BMG, we let the total $C\{\cdot\}$ include fixed and variable components:

$$C(Q,E,\theta) = C_1 E + C_2 QE + C_3 \theta Q, \tag{5}$$

which is linear in the strength of enforcement, and also depends on the quantity $Q$ and the probability of catching an illegal producer (through $\theta$).

Equation (4) can then be expressed as

$$\begin{aligned}\max_E W =\ & V[Q(E)] + \beta A^*(E)Q(E) - P(E)Q(E) \\ & - C_1 E - C_2 Q(E)E - C_3\theta[E, A^*(E)]Q(E),\end{aligned}$$

which (focusing on interior solutions) has the following FOC for the $W$-maximizing level of $E$:

$$\begin{aligned}& V_q \frac{dQ}{dE} + \beta\left[A^*\left(\frac{dQ}{dE}\right) + Q\left(\frac{dA}{dE}\right)\right] - MR\frac{dQ}{dE} \\ & \quad - C_1 - C_2\left[Q + \left(\frac{dQ}{dE}\right)E\right] - C_3\left[\theta\left(\frac{dQ}{dE}\right) + Q\left(\frac{\partial\theta}{\partial E} + \frac{\partial\theta}{\partial A}\frac{dA}{dE}\right)\right] = 0,\end{aligned} \tag{6}$$

where $MR \equiv dPQ/dQ$ is the marginal revenue of producers. This can then be implicitly solved for the optimal enforcement $E^* > 0$.

One can also rearrange equation (6) into

$$\begin{aligned}& C_1 + C_2\left[Q + \left(\frac{dQ}{dE}\right)E\right] + C_3\left[\theta\left(\frac{dQ}{dE}\right) + Q\left(\frac{\partial\theta}{\partial E} + \frac{\partial\theta}{\partial A}\frac{dA}{dE}\right)\right] \\ & \quad = V_q\frac{dQ}{dE} - MR\frac{dQ}{dE} + \beta\left[A^*\left(\frac{dQ}{dE}\right) + Q\left(\frac{dA}{dE}\right)\right]\end{aligned} \tag{7}$$

to show the left-hand side as the marginal cost of enforcement, and the right-hand side as the marginal benefit.

The main result is that enforcement is more effective in decreasing the quantity of illegal goods, the larger the extent of corruption.

PROPOSITION 1 *Let $Q(dA/dE) < A|dQ/dE|$. Then $dQ/d\beta < 0$.*

PROOF Note that $dQ/d\beta = (dQ/dE)(dE/d\beta)$. Denote the elasticity of demand as $\epsilon_d$. Then $dQ/dE = \epsilon_d(Q/P)(dP/dE)$, which is negative, since $\epsilon_d < 0$, $Q/P > 0$, and $dP/dE > 0$. (To see the last inequality, differentiate equation (3) with respect to $E$: $dP/dE = (\partial\theta/\partial E)(c + A^* + F) > 0$.) Now, applying the implicit-function theorem to equation (6),

$$\frac{dE}{d\beta} = -\left(\frac{A(dQ/dE) + Q(dA/dE)}{-c_2(dQ/dE)}\right).$$

Thus, $dE/d\beta > 0$ if $Q(dA/dE) < A|dQ/dE|$, since $dQ/dE < 0$ and $dA/dE > 0$. (To see the latter, apply the implicit-function theorem to equation (2) to get

$$\frac{dA}{dE} = -\left(\frac{-\theta(\partial\theta/\partial E)}{-\theta(\partial\theta/\partial A) - (\partial\theta/\partial A)}\right),$$

where

$$\frac{\partial\theta}{\partial E} = \frac{(1-p(E,A))(\partial p/\partial E)+p(E,A)(\partial p/\partial E)}{(1-p(E,A))^2} = \frac{\partial p}{\partial E}\frac{1}{(1-p(E,A))^2},$$

which is positive, since $\partial p/\partial E > 0$ by assumption, while $\partial\theta/\partial A = [\partial p/\partial A][1/(1-p(E,A))^2]$, which is negative, since $\partial p/\partial A < 0$ by assumption. Thus, $dA/dE > 0$.)

The intuition is straightforward. The receipt of bribes incentivizes the government to enforce against illegal goods, which raises the probability of being caught and induces the producer to spend resources to avoid this. The price of the good rises and hence the demand for it decreases.

Using the expressions for $dA/dE$, $dQ/dE$, and $dP/dE$, this effect is seen to be more likely the lower the probability of being caught, the more effective the avoidance in lowering this probability, and the higher the elasticity of demand for the good.

Thus far, we have considered only interior solutions. BMG consider the instance in which there may not exist a nonzero value of enforcement $E$ that maximizes social welfare, depending on the elasticity of demand. In their model, bribes are not included in the social welfare function, so that when the government's marginal cost of enforcement is set to zero, the optimal level of enforcement is given by $V_q = MR \equiv P(1+1/\epsilon_d)$. When demand is inelastic (so that $MR < 0$) and the marginal social value of consumption is nonnegative (i.e., $V_q \geq 0$), enforcement cannot be nonzero. Thus, there is no level of enforcement that is socially optimal, which justifies abandoning prohibition and freeing the market.

For an analogous result for corrupt environments, we also set marginal enforcement costs to zero in equation (7). Denote $G_E \equiv \beta[A^*(dQ/dE)+Q(dA/dE)]$. The optimal value of enforcement $E$ is thus given by

$$V_q + \frac{G_E}{dQ/dE} = MR \equiv P\left(1+\frac{1}{\epsilon_d}\right). \tag{8}$$

This leads to the following result.

Proposition 2 *Let $Q(dA/dE) < A|dQ/dE|$. Assume that the marginal cost of enforcement is zero, and that $V_q \geq 0$. Then if demand is inelastic, enforcement is never optimal. If demand is elastic, there is always a nonzero value of enforcement $E$ that is socially optimal.*

Proof With inelastic demand, $MR < 0$. Thus, equation (8) implies that when $V_q + G_E/(dQ/dE) \geq 0$, that is, $V_q \geq -G_E/(dQ/dE)$, $E$ cannot be nonzero. Since $G_E < 0$ and $dQ/dE < 0$, this inequality is *always* satisfied when $V_q \geq 0$. If demand is elastic, $MR > 0$, which implies that enforcement is not optimal when $V_q \leq -G_E/(dQ/dE)$, which is *never* satisfied when $V_q \geq 0$.

Thus, for both our model and BMG's, prohibition is optimal only when demand is elastic (assuming the marginal social value of consumption is nonnegative). The difference is that with the same $V_q \geq 0$ and $\epsilon_d > 1$ as in BMG, the optimal level of enforcement is larger in a corrupt environment, provided that $G_E < 0$. To see this, note that $G_E/(dQ/dE)$ is positive, which means that $MR$ is higher than in BMG. Since $\epsilon_d > 1$ is the same, it means that the price $P$ is higher in a corrupt environment, which is possible precisely because $E$ and, hence, $\theta(E, A)$ are larger. (Recall equation (3).) The intuition is given by Proposition 1 – with $G_E < 0$, we have $Q(dA/dE) < A|dQ/dE|$, and so each bribe spent by the illegal producer incentivizes the government to enforce even more, which lowers quantity further. In BMG, there is no such incentive, as the government does not include bribes in its maximand.[3]

This key difference is preserved even if the marginal cost of enforcement is nonzero, since (from the left-hand side of equation (7)) the latter is not affected by $\beta$. In this case, denoting the left-hand side of equation (7) as $MC$, the optimal enforcement in BMG is given by

(9) $$V_q - \frac{MC}{dQ/dE} = MR,$$

while the analogous equation for a corrupt environment is

(10) $$V_q + \frac{G_E - MC}{dQ/dE} = MR.$$

Thus, as long as $G_E < 0$, the optimal level of enforcement is still larger, and quantity smaller, in a corrupt environment.

Now suppose that there are other costs of enforcement in a corrupt environment, which depend on the extent $\beta$ of corruption. In particular, let the total cost be given by

$$C(Q, E, \theta, \beta) = C_1 E + C_2 QE + C_3 \theta Q + C_4(\beta) E,$$

where $C_4$ is the additional cost per unit of enforcement, which is a function of $\beta$, but let the total cost in BMG remain as in equation (5). The marginal cost of enforcement in a corrupt environment is now

$$C_1 + C_2\left[Q + \left(\frac{dQ}{dE}\right)E\right] + C_3\left[\theta\left(\frac{dQ}{dE}\right) + Q\left(\frac{\partial\theta}{\partial E} + \frac{\partial\theta}{\partial A}\frac{dA}{dE}\right)\right] + C_4,$$

[3] Note that in their model, illegal producers also incur avoidance costs to prevent being caught, which may include bribes to enforcement officers. It is only that such bribes do not enter the social welfare function. One can then use the BMG model to capture the case of a relatively *clean* bureaucracy, in which only the prohibition agents engage in corrupt practices by accepting bribes.

which means that optimal enforcement is given by

$$V_q + \frac{G_E - MC - C_4}{dQ/dE} = MR. \tag{11}$$

The effect of corruption on the optimal level of enforcement thus depends on $C_4$. If corruption decreases the marginal cost of enforcement (i.e., $C_4 < 0$), the left-hand side of equation (11) is still larger than the left-hand side of equation (9) as long as $G_E < 0$. Thus, for the same values of $V_q \geq 0$, $MC$, and $|\epsilon_d| > 1$ as in BMG, equation (11) implies a higher level of $E$. The same analysis holds when the marginal cost of enforcement increases with corruption (i.e., $C_4 > 0$) as long as $G_E + MC > C_4$, since in this case, $G_E + MC - C_4$ is still negative and $(G_E - MC - C_4)/(dQ/dE)$ still positive. The following thus formalizes the result:

PROPOSITION 3 *Suppose the cost of enforcement in a corrupt environment increases with the extent $\beta$ of corruption by amount $C_4$. Let $V_q \geq 0$, $Q(dA/dE) < A|dQ/dE|$ (so that $G_E < 0$), and $|\epsilon_d| > 1$. Then, for the same values of $V_q$, $MC$, and $\epsilon_d$ as in BMG, the optimal level of enforcement is higher in the corrupt environment, and the quantity smaller, as long as either of the following conditions holds: (i) $C_4 < 0$, or (ii) $C_4 > 0$ and $G_E + MC > C_4$.*

## 3 Taxation

Suppose the market is legalized, and producers now pay excise taxes $\tau$ per unit of output while incurring the same marginal cost $c$. The total tax revenue from the good is thus $\tau Q$. Since the market is now legal, producers do not have to pay bribes and incur avoidance costs. However, if the government remains corrupt, it can expropriate some of the revenues, so that only $\delta\tau Q$ is plowed back to society, with $\delta \in (0,1)$. The extent of corruption is thus now captured by $1-\delta$, which is higher the closer $\delta$ is to zero.

An important assumption is that in a legalized market, producers are no longer complicit with government corruption. In contrast, bribery in the case of prohibition is a two-way transaction, and thus both illegal producers and the government are aware of the corruption. In choosing the optimal level of enforcement, the government can openly maximize its bribe revenues, which is why the latter are included in the total social welfare function. With a legalized market, the government has to hide the expropriation so that, upon observing the tax rate, producers cannot back out the amount of tax revenues that was expropriated. Otherwise, they reveal this information, in which case we simply assume that society immediately demands and obtains the entire tax revenue $\tau Q$ at zero cost. As long as the expropriated tax $(1-\delta)\tau Q$ is *hidden*, or excluded from the government's maximand, a social return of $\delta\tau Q < \tau Q$ could be excused as ordinary public-sector inefficiency and not as evidence of expropriation.

In other words, for corruption to be successfully hidden in a legalized market, the government's observable behavior should be the same as in a clean environment.

The government thus maximizes a social welfare function that excludes its private gain from expropriated tax revenues, which is exactly the same as in BMG's model of taxation:[4]

$$W_m = V(Q) - cQ - \tau Q + \delta\tau Q,$$

where $V(Q)$ is the benefit to consumers, $cQ + \tau Q$ the total cost to producers, and $\delta\tau Q$ the social value from tax revenues. The difference from BMG is that they interpret $\delta$ as an indicator of the extent of efficiency with which the government converts tax revenues into public goods. (Thus, $\delta = 1$ is interpreted as a pure transfer.)

With perfect competition, the price is $P = c + \tau$. We can then plug in $\tau = P - c$ above and rearrange to get

$$W_m = V(Q) - cQ - (1-\delta)[P(Q)Q - cQ)].$$

Let the government choose a target (i.e., optimal) output level, and restrict attention to interior solutions. Then the FOC for $Q$ is

$$V_q - c - (1-\delta)(MR - c) = 0, \tag{12}$$

with $V_q \equiv dV/dQ$. This can be implicitly solved to get the optimal quantity $Q^* > 0$.

As expected, higher tax rates curb the supply:

Proposition 4 *Quantity $Q$ decreases when tax rate $\tau$ increases. Specifically,*

$$\frac{dQ}{d\tau} = \frac{\delta}{(1-\delta)QP\epsilon_d} < 0.$$

Proof Using $c = P - \tau$ and applying the implicit-function theorem to equation (12), we get $dQ/d\tau = \delta/[(1-\delta)(dMR/dQ)]$, which is negative, since $dMR/dQ < 0$. (To see the latter, note that since $MR \equiv P(1 + 1/\epsilon_d) = P(1 + Q/[P(\partial Q/\partial P)])$, the derivative $dMR/dQ = \partial P/\partial Q = (1/\epsilon_d)(P/Q)$ is negative, since $\epsilon_d < 0$.)

Thus, both enforcement of prohibition rules and taxation of a legalized market decrease quantity. However, under certain conditions, prohibition can induce a larger decrease in quantity than taxation:

Proposition 5 *Let $dP/dE < \delta/[(1-\delta)Q^2\epsilon_d^2]$. Then $dQ/dE < dQ/d\tau$.*

Proof Recall that $dQ/dE = \epsilon_d(Q/P)(dP/dE)$ and $dQ/d\tau = \delta/[(1-\delta)QP\epsilon_d]$, which are both negative, since $\epsilon_d < 0$. Thus, $Q$ drops faster when $E$ increases than when $\tau$ increases. This simplifies to $dP/dE < \delta/[(1-\delta)Q^2\epsilon_d^2]$.

[4] If expropriation were not hidden, then $W_m = V(Q) - cQ - \tau Q + \delta\tau Q + (1-\delta)\tau Q = V(Q) - c(Q)$. That is, it would be obvious that no tax revenues are plowed back to society.

Recall that

$$\frac{dP}{dE} = \left(\frac{\partial p}{\partial E}\frac{1}{(1-p(E,A))^2}\right)(c + A^* + F).$$

Proposition 5 then implies that prohibition is likely to be more effective at curbing quantity when avoidance costs, the probability of getting caught, and the responsiveness of the latter to enforcement efforts are high under prohibition. Under these conditions, bribery works better, since illegal producers have to spend a lot to avoid being caught, which drives up price and lowers quantity.

In fact, we find that corruption actually *increases* quantity in a legalized market:

PROPOSITION 6 *Let the extent of corruption be captured by* $1-\delta$. *Then* $dQ/[d(1-\delta)] > 0$.

PROOF Applying the implicit-function theorem to equation (12), we have

$$\frac{dQ}{d(1-\delta)} = -\left(\frac{MR - c}{(1-\delta)(dMR/dQ)}\right),$$

which is positive, since, as shown previously, $dMR/dQ < 0$.

This result is the direct opposite of the case of prohibition. The intuition is that when the good is illegal, a corrupt government has to enforce prohibition rules in order to induce illegal producers to pay bribes. That is, the gains from corruption are contingent on enforcement efforts. When the good is legal, a corrupt government now has to directly extract from the value of the market, which is why it prefers a larger market.

Lastly, we can also analyze corner solutions. Writing out $MR \equiv P(1+1/\epsilon_d)$ in equation (12) and rearranging gives the implied level of taxation that achieves the optimal quantity $Q^*$:

$$\frac{V_q - \delta c}{1-\delta} = P\left(1 + \frac{1}{\epsilon_d}\right). \tag{13}$$

This leads to the following result:

PROPOSITION 7 *There is no nonzero tax rate that is socially optimal if either of the following conditions holds: (i)* $V_q > \delta c$ *and* $|\epsilon| < 1$, *or (ii)* $V_q < \delta c$ *and* $|\epsilon| > 1$.

PROOF With $|\epsilon_d| < 1$, the right-hand side of equation (13) is negative. Thus, an optimal tax rate requires that the left-hand side be negative, which cannot be satisfied if $V_q > \delta c$ (since $\delta c \geq 0$). Analogously, with $|\epsilon_d| > 1$, the right-hand side of equation (13) is positive, which requires the left-hand side to be positive. This cannot be satisfied if $V_q < \delta c$.

Proposition 7 implies that taxation can be optimal for a larger range of values for $V_q$ than enforcement. To see this, note that when demand is inelastic (i.e., $|\epsilon_d| < 1$), taxation becomes nonoptimal only when $V_q > \delta c$. However, from equation (10), enforcement cannot be optimal when $V_q \geq 0$ (assuming, as usual, that $G_E > 0$.) Since $\delta c \geq 0$, taxation may still be optimal when enforcement is not. We can get the same qualitative result in the case of elastic demand. With $|\epsilon_d| > 1$, Proposition 7 establishes that taxation cannot be optimal if $V_q < \delta c$, while equation (10) implies that enforcement cannot be optimal if $V_q < (G_E - MC)/(dQ/dE)$. As long as $(G_E - MC)/(dQ/dE) > \delta c$, there is a larger range of values for $V_q$ for which taxation is optimal. Thus, if enforcement is not optimal, taxation may still be, but the reverse is not true.

This suggests that taxation is more effective than prohibition in *expanding* the market. In the case of inelastic demand, the tax rate could still be further lowered in order to increase quantity when $0 < V_q < \delta c$, while enforcement efforts cannot be decreased further. If demand is elastic, the tax rate can still be lowered, while enforcement cannot be, as long as $(G_E - MC)/(dQ/dE) > V_q > \delta c$.

This follows intuitively from Proposition 5, which establishes that enforcement is more effective than taxation in *decreasing* quantity.

## 4 *The Case of Drugs*

The findings of our model suggest that the prohibition of drugs decreases consumption when corruption is rife and tolerated by society. To date, there is no evidence to prove this, as there are no data on the counterfactual scenario. Virtually no country has legalized drug production. There are, though, a number of countries that have decriminalized drug *use* in varying degrees, e.g., EU countries, and most notably Portugal in 2001.[5] There is some casual evidence that suggests that decriminalization has lowered drug consumption and its concomitant evils, especially in Portugal (see, e.g., Hughes and Stevens, 2010; Greenwald, 2009), but there are also studies (e.g., McKeganey, 2007; Inciardi, 2008; Singer, 2008) that argue that "removing criminal penalties would lead to increased drug use" (Hughes and Stevens, 2010, pp. 999–1000).

There are also studies that conclude that harsher punishment of drug users does not directly affect rates of drug use (see, e.g., Reuter and Stevens, 2007; Degenhardt et al., 2008). Yet even if this is true, it does not readily disprove the results of the model – it does not follow that taking away punishments would lead to a drop in consumption. At most, it only implies that there would be no change in consumption (because punishments have no effect). This result can be approximated by our model if the elasticity of quantity with respect to enforcement is very low because of, say, very low avoidance costs, which decreases bribes. In this case, enforcers would have little incentive to enforce the punishments in the first place, and hence taking away the punishments would matter little.

---

[5] See Hughes and Stevens (2010), Carpenter (2009), and Greenwald (2009) for an overview. "EU nations have adopted what amounts to de facto decriminalization" (Greenwald, 2009, p. 11). Note also that even in the Netherlands, where drug use has been decriminalized, drugs are still illicit (Carpenter).

In fact, countries that have decriminalized actually have low corruption, so it is difficult to disprove the main point of our model that prohibition lowers consumption more the larger the extent of corruption. Since none of these studies have controlled for the level of corruption, one cannot credibly establish the counterclaim that for countries that have similarly decriminalized, those that are more corrupt experienced a smaller increase (or even a decrease), rather than a larger increase, in consumption.

Fundamentally, however, it is doubtful whether such studies on the effects of decriminalization are applicable in analyzing the effect of legalizing drugs. One must distinguish between the degree of enforcement and the mode of enforcement. It can very well be argued that decriminalization is not actually a move towards legalization, but to a more effective kind, or a larger degree, of prohibition. If this is the case, then the fact that drug use has fallen with decriminalization actually illustrates that our model works.

In fact, there is a strong consensus that decriminalization is not akin to legalization, but rather a better mode of combating and reducing drug use, and a more effective way to deter the use of (still) illicit drugs. Note that even for Mexico, which arguably suffers great direct and indirect costs from the drug trade, calls for reform have been in the form of proposing decriminalization, and not legalization, of drugs (see Carpenter, 2009). Greenwald (2009, p. 10) notes that by adopting decriminalization, "the aim in most EU countries is merely to formulate more efficient and proportionate sanctions–not legalize drug use." Also, that "decriminalization was driven *not* by the perception that drug abuse was an insignificant problem, but rather by the consensus view that it was a highly significant problem, that criminalization was *exacerbating* the problem, and that only decriminalization could enable an effective government response" (p. 6). "Put in another way, Portuguese decriminalization was never seen as a concession to the inevitability of drug abuse. To the contrary, it was, and is, seen as the most effective government policy for reducing addiction and its accompanying harms" (p. 10).

In addition, if decriminalization can thus be seen as a form of prohibition, our model suggests that among countries that have decreased drug consumption since decriminalizing, those that are more corrupt would enforce more and, hence, would make decriminalization even more effective. Note indeed that Portugal, which Greenwald shows to have had the lowest consumption of cannabis among EU countries between 2001 and 2005, has a higher level of (perceived) corruption than Denmark, the country in the sample with the highest consumption of cannabis.[6]

Of course, this fact is merely suggestive, and a more rigorous statistical analysis is called for. But the point is that without properly controlling for the effect of corruption, it is difficult to disprove the results of our model. Ideally, one should have sufficient variation in the level of corruption across countries to tease out the effect of corruption on drug consumption when countries move from less to

[6] The 2009 Corruption Perceptions Index (CPI) of Portugal is 5.9, while that of Denmark is 9.3 (see Desierto and Nye, 2011).

more prohibition (or vice versa). To date, however, such a sample of countries is unavailable.

## *5 Conclusions*

This paper highlights a key reality surrounding illegal markets – the fact that illegal producers can and do bribe officials to avoid getting caught and punished. However, perhaps because this fact and the value judgments associated with it are now commonplace, a truly positive economic approach is still lacking in the literature. What we have thus shown is that when one only considers pure efficiency arguments, the gains from bribery can end up providing additional justification for banning a good. If corruption is tolerated by society, a wealth-maximizing state will also maximize the bribes it receives from illegal producers. In this case, bribes do not represent a deadweight loss to society, because they are simply transferred to the state.

An important implication is that even if consumption of the good generates large marginal social value, it is still optimal for a corrupt government to engage in an enforcement war against illegal producers, precisely because producers bribe their way out. The bribes provide an incentive to the government to enforce prohibition, since producers have to keep paying bribes in order to keep supplying the market. In equilibrium, quantity is smaller than in a clean environment in which the government cannot maximize bribes.

We have also shown that the reduction in quantity from the enforcement of prohibition rules may actually be larger than the reduction achieved by legalizing and taxing production. Thus, if the goal is to limit consumption of the good, then making it illegal is likely to be a more effective strategy than imposing taxes. When the good is illegal, the government's gain from corruption is contingent on enforcement against illegal producers – to sustain the payment of bribes, the government has to keep enforcing, and the latter remains credible if the quantity that ends up being supplied to the market is limited. In contrast, when the good is legal, the only way a corrupt government can extract rents is to expropriate tax revenues. Thus, it prefers a larger market, which can generate more taxes.

While our model does not fully settle the question of whether prohibition or legalization is the better policy for dealing with undesirable drugs and other such contraband, it focuses on the role of corruption and thus provides an important and plausible theoretical rationale for prohibition that has not been adequately considered in the literature.

An interesting avenue for further research would be to investigate equivalent considerations for goods with strong positive externalities. An analogy with our model might suggest a rationale for requiring the direct provision of certain public goods (such as public education or vaccines) over mere subsidies. If rents are contingent on the provision of such goods, a corrupt government might actually want to increase quantity. In contrast, it might want to limit subsidies, since the latter would decrease the amount of tax revenues that could be expropriated.

*References*

Becker, G. S., K. M. Murphy, and M. Grossman (2006), "The Market for Illegal Goods: The Case of Drugs," *Journal of Political Economy*, 114(1), 38–60.

Carpenter, T. G. (2009), "The International Drug War," in: *Cato Handbook for Policymakers,* 7th edn., Cato Institute, Washington (DC), pp. 599–608.

Degenhardt, L., W. T. Chiu, N. Samson, R. C. Kessler, J. C. Anthony, et al. (2008), "Toward a Global View of Alcohol, Tobacco, Cannabis, and Cocaine Use: Findings from the WHO World Mental Health Surveys," *PLoS Medicine*, 5(7), e141.

Desierto, D. A., and J. V. C. Nye (2011), "Why do Weak States Prefer Prohibition to Taxation?" in: N. Schofield and G. Caballero (eds.), *The Political Economy of Institutions, Democracy and Voting*, Springer-Verlag, Berlin and Heidelberg, pp. 75–94.

Glaeser, E. L., and A. Shleifer (2001), "A Reason for Quantity Regulation," *The American Economic Review*, Papers and Proceedings, 91(2), 431–435.

Greenwald, G. (2009), *Drug Decriminalization in Portugal: Lessons for Creating Fair and Successful Drug Policies*, Cato Institute, Washington (DC).

Hughes, C. E., and A. Stevens (2010), "What Can we Learn from the Portuguese Decriminalization of Illicit Drugs," *The British Journal of Criminology (BJS)*, 50(6), 999–1022.

Inciardi, J. A. (2008), *The War on Drugs IV: The Continuing Saga of the Mysteries and Miseries of Intoxication, Addiction, Crime, and Public Policy*, Pearson, Boston (MA).

Lambsdorff, J. G. (2007), *The Institutional Economics of Corruption and Reform: Theory, Evidence and Policy*, Cambridge University Press, New York.

McKeganey, N. (2007), "The Challenge to UK Drug Policy," *Drugs: Education, Prevention and Policy*, 14(6), 559–571.

Miron, J. A. (2004), *Drug War Crimes: The Consequences of Prohibition*, The Independent Institute, Oakland (CA).

— (2008), "The Budgetary Implications of Drug Prohibition," Report for the Criminal Justice Policy Foundation and Law Enforcement against Prohibition, Maryland, available at http://inorml.com/download/cost-of-prohibition/Miron-economic-report-2008.pdf, last accessed April 8, 2016.

Pamintuan, J. (2010), "Ang Politika, Kultura at Ekonomiks ng Jueteng (The Politics, Culture, and Economics of Jueteng)," Ph.D. Dissertation, University of the Philippines, Quezon City.

Reuter, P., and A. Stevens (2007), *An Analysis of UK Drug Policy*, UK Drug Policy Commission, London.

Singer, M. (2008), *Drugging the Poor: Legal and Illegal Drugs and Social Inequality*, Waveland Press, Long Grove (IL).

Weitzman, M. L. (1974), "Prices *vs.* Quantities," *The Review of Economic Studies*, 41(4), 477–491.

Desiree A. Desierto
Department of Political Science
University of Wisconsin–Madison
110 North Hall
1050 Bascom Mall
Madison, WI 53706
U.S.A.
desierto@wisc.edu

John V. C. Nye
Economics Department
George Mason University
Carow Hall MSN 1D3
4400 University Drive
Fairfax, VA 22030
U.S.A.
jnye@gmu.edu

# The Economics of Empire-Building: Predatory and Price Competitions

by

Antoine Pietri, Tarik Tazdaït, and Mehrdad Vahabi*

Received July 6, 2015; in revised form February 20, 2016;
accepted March 13, 2016

We build a theoretical framework consistent with historical evidence in which empire-building is explained by price and predatory competitions on the market for protection. We explore how the assets structure possessed by the buyers of protection influences the nature of protection and *in fine* the size of empires. Our main contribution is to introduce a distinction between two types of rent, namely an "absolute" and a "differential" one. The first corresponds to rents extracted by empires using threats and coercion; the second, to economic advantages conferred on subjects of an empire. (JEL: D74, H56, L13)

## 1 Introduction

In the second half of the fifteenth century, most spices were carried by Arab merchants from India to Jeddah in the Red Sea. These merchants were under the protection of the Egyptian Sultan, who exacted a large protection fee. At Alexandria, the spices were sold to the Venetians and other Europeans. When the Portuguese reached India by circumnavigating Africa, the famous Venetian merchant banker, Gerolamo Priuli, predicted that the Portuguese would be able to undersell the Vene-

* Antoine Pietri: University of Paris 1 (CES), Paris; Tarik Tazdaït: C.N.R.S. – E.H.E.S.S. – CIRED, Nogent Sur Marne; Mehrdad Vahabi (corresponding author): Department of Economics, University of Paris 8, Saint-Denis, France. Earlier versions of this paper have been presented at the Max Planck Institute for Tax Law and Public Finance in Munich, January 9th, 2013; in research seminars of the Department of Political Economy, University of Neuchatel, Switzerland, January 17th, and at the annual conference of the 2013 Midwest Political Science Association (MPSA), in Chicago, April 2013. We benefited from the comments of the participants in the aforementioned scientific conferences in preparing the present version of our paper. We would like to thank the participants in these conferences as well as Renaud Bellais, Bertrand Crettez, Jean-Marie Grether, Kai Konrad, Stergios Skaperdas, Milad Zarin-Nejadan, one anonymous referee, and Gerd Muehlheusser, for their inspiring and insightful remarks. Obviously, all the remaining errors are ours.

*Journal of Institutional and Theoretical Economics* 173, 253–278 – ISSN 0932-4569
DOI: 10.1628/093245616X14659946859954 

tians because they would avoid the high taxes in Egypt by bringing the spices around Africa. However, "This is *not* what happened. The Portuguese did *not* set their prices below those common at Venice in the fifteenth century […]. The Portuguese king attempted to prevent by armed force the passage of any spices from India to the Red Sea or Persian Gulf" (Lane, 1979, pp. 15–16). On the one hand, the Republic of Venice used violence as a means of accumulating mercantile profit in quest of economic advantages for Venetian merchants: "Through all the conquest, Venetian commercial interests reigned supreme" (Tilly, 1990, p. 145). On the other hand, the Portuguese strategy of empire-building focused on the increase in wealth of the Crown. Violence was used to threaten and extort from people, partly because private merchants were not influential and the pepper trade was managed by a royal company.

From an economic perspective, these behaviors are traditionally analyzed in terms of a market for protection[1] in which different "predatory governments" or "Leviathans" (Alesina and Spolaore, 2003) care more about their own welfare than about the welfare of their subjects. In this framework empowerment of buyers of protection leads to "democratically constrained Leviathans" characterized by rent dissipation (Spolaore, 2012). Although this view seems correctly fitted to the Portuguese strategy, it fails to explain the rationale of merchant empires. In this paper, we claim that the traditional concept of rent is too narrow to depict the economics of empire-building (and more generally the question of political borders). By "empire" we mean "a specific form of rule" (Spruyt, 2008, p. 291). As a consequence, in our paper, an empire is seen as an "imperial rule" enacted by a central authority in a bounded geographical area. This imperial rule is deeply asymmetrical (Nexon and Wright, 2007) and consists of a system of "contracts" within a context of coercive power depending on its "military integration zone" (Vahabi, 2004). Our original contribution is to introduce a distinction between two different types of rent, namely an "absolute" and a "differential" one. What we call *absolute protection rent* (AR) is a rent extracted from a coerced population by the sellers of protection (kings, lords, vassals). In contrast, the concept of *differential protection rent* (DR) stands for economic advantages provided to the buyers of protection (merchants, bankers, and producers). The distinction between these two types of rents has great significance and allows a more subtle analysis of empire-building (Vahabi, 2016). On the one hand, AR corresponds to private benefits of the rulers, and captures a tyrannical kind of expansion based on coercion. In this case, the economic significance of violence is the extraction of *monopoly rent*. On the other hand, introducing DR entails the understanding of empires promoting *profit-maximizing* activities of merchants and other groups of investors, which is one of the characteristics of merchant empires. The use of violence is no longer the principal source of accumulation; it is rather supportive of profit maximization as a new source of accumulation.

[1] Following Volckart (2002, p. 339), we believe that the economic concept of market could be used in the case of protection mostly because "protection rather had the characteristics of a private than of a public good, being in effect a commodity that could be exchanged like any other marketable good."

In this paper, we conduct a theoretical analysis of these two kinds of economic rents in order to provide a better understanding of empire-building throughout history.

This decomposition is consistent with the different kinds of competition existing in the market for protection. Indeed, it is in line with economic historians like Lane (1973, 1979), North and Thomas (1973), and North (1981), who stressed the importance of *price competition* in the market for protection[2] as a source of Western ascendancy and empire-building. However, theorists of conflictual activity have argued against this form of competition. Conflict models are often inspired by a particular form of European serfdom in the Middle Ages, in which peasants were at the mercy of protector-lords (Findlay, 1996; Skaperdas, 2002; Konrad and Skaperdas, 2012). According to these models, subjects are *immobile*, and "providers of protection, instead of competing on the price of their service, typically compete with their means of violence over turf" (Konrad and Skaperdas, 2012, p. 418). By systematically rejecting the potential mobility of demanders of protection, theoretical models fall short of encompassing the various forms of empire-building.[3] In stark contrast, we argue that mobility of buyers of protection is the key factor explaining behavior of an empire, because it allows subjects to escape from coercion. Taking the example of rackets, Konow (2014, p. 50) provides a salient illustration of this phenomenon: "a gangster's threat of violence to businesses if they do not pay 'protection money' is not coercion if businesses choose to ignore the threat." Applying this idea to the case of empire-building implies that if buyers of protection are fully mobile, they can choose to ignore the empire's threat of coercion by fleeing from its sphere of influence. Consequently, the degree of subjects' mobility determines whether the empire-building is carried through coercion or consent.

Our line of investigation is to take into account the assets structure possessed by subjects. We establish a link between assets redeployability and the degree of mobility of buyers of protection (for a more detailed investigation, see Vahabi, 2016, chapters 6 and 7). The nature of the market for protection drastically depends on this level of mobility. We distinguish three types of market for protection. In the first one, the *sellers' protection market*, buyers possess nonredeployable assets and are at the mercy of the provider of protection. In stark contrast, in a *buyers' protection market* subjects hold perfectly redeployable assets and are able to switch from one provider of protection to another. Third, we analyze a *hybrid protection market* in which the assets are partly redeployable. The distinction between the two types of rents combined with the concept of assets redeployability provides an economic explanation for the different usages of violence previously highlighted in the cases of the Portuguese and Venetian empires.

[2] See also all the papers in the volume edited by Tracy (ed.) (1991) regarding the political economy of merchant empires.

[3] Skaperdas (2003, p. 150) claims that "rulers have typically worried much more about the armies of their competitors across their borders than about how the fiscal policies of their competitors affect the movement of their subjects." However, he concedes that price competition on the market for protection might theoretically be considered if subjects are mobile.

The paper proceeds as follows. The second section highlights the historical relevance of subjects' mobility among different providers of protection. In particular we establish a theoretical linkage between the assets structure and the degree of mobility. Then we explore the economic significance of two different methods of using violence – extracting absolute or differential protection rents – and we identify two types of empires based on these respective objectives. In section 3, we examine dyadic competitions among empires according to the mobility of subjects. We distinguish three cases: perfectly immobile, perfectly mobile, and partially mobile subjects. Finally, section 4 highlights our conclusions and few extensions.

## 2 Population Mobility and Protection Rents

This section focuses on the assumption of mobile population on the protection market, which is the necessary condition allowing price competition. We first discuss the role of assets redeployability as an explanatory factor for the mobility of subjects (section 2.1). Then, we distinguish two ideal types of empire, based on the type of rent (absolute or differential) that they maximize (section 2.2).

### 2.1 Population Mobility and the Nature of the Protection Market

Conflict models based on the assumption of immobile subjects are not inclusive of all historical situations, for at least two reasons. First, there were historical periods in which subjects of an empire had the ability to escape, and therefore could choose their provider of protection. According to Bloch (1966, pp. 85–87), legal definitions of serfdom do not describe serfs as "attached to the soil" before the fourteenth century, and serfs were in practice usually free to move, although maintaining at least in theory a legal bond to their masters.[4] Second, in principle the mobility of subjects should not be reduced to a geographical concept. Indeed, mobility is not systematically associated with a physical move. People can decide to disown a provider of protection in order to be protected by another one without changing their location. In fact such betrayal was not rare throughout history. A salient example is discussed by Douglass North (1981) regarding lower taxes imposed upon border areas in medieval Burgundy due to competition between feudal lords. In this case, peasants in border areas had the possibility to change their allegiance to take advantage of price competition[5] between predator landlords. Germany during the Thirty Years' War, during the higher Middle Ages (from the eleventh to the four-

[4] Interestingly enough, Evsey Domar (1970) also underlined that before 1550 Russian peasants were free men; a hundred years later they were serfs.

[5] Historically, the protection "price" was not a monetary price; it was principally composed of taxes in kind. For example, a certain amount of grain and number of livestock were decreed as the appropriate tax (Levi, 1981, p. 458). *Corvée*, conscription, and other forms of servitude were also included in the protection price. In this context, price competition means paying less in grain or livestock or serving fewer hours of *corvée* for landlords.

teenth centuries), and throughout the nineteenth century is another good example of how several violence-using enterprises can compete in demanding payments for protection in almost the same territory (Lane, 1979, p. 51; Volckart, 2000, p. 267). Hence, even geographically attached people might be mobile in the sense that they have *exit power*: they can choose their provider of protection. In that sense a person is mobile as soon as she/he has the ability to substitute a provider of protection for another. In this context, changing his/her allegiance may be possible. Moreover, radical institutional changes such as the abolition of feudalism radically alter the situation of peasants by increasing their detachment from the land, giving rise to free and mobile labor. Land reform augments peasants' exit power. Thus, assets mobility is not limited to merchants, financiers, or industrialists but also includes free urban and rural wage laborers. However, economic literature, focusing on immobile populations, disregards these situations. In this paper we are interested in building a framework taking into account diverse degrees of exit power.

Undeniably, there is a link between the degree of exit power and the assets structure faced by people. Borrowing Williamson's (1985) asset specificity criteria, asset redeployability corresponds to the cost induced by employing an asset in an alternative use. Adapted to our specific framework, perfectly redeployable assets stand for the fact that the revenue of a person is the same whoever protects him/her: changing allegiance is almost costless and mobility is ensured. Notably, capital is less specific than land in the sense that it can be redeployed elsewhere without losing its value. For example, merchants intensively use capital, which is a highly redeployable asset. Contrary to *nonmovable* assets such as land and buildings, capital is "the most mobile factor of production" (Bates and Lien, 1985, p. 54). It is harder to confiscate or tax capital because of its exit power. As the English fiscal experience proved, the new taxes on trade and movables possessed two significant shortcomings: "They could be easily avoided. And partially as a consequence, they had to be bargained for" (Bates and Lien, 1985, p. 55). Consequently, the bargaining power of proprietors of movable and redeployable assets compelled the predator states to offer competitive prices of protection with regard to their service as violent protectors of their movable assets. As we can see in the table below, this situation gives rise to a *buyers' protection market*: a situation in which subjects have full ability to exit and thus can influence empire-building. Montesquieu had already predicted that the rise of commerce could have political consequences. Discussing the invention of the letter of credit, Montesquieu (1768) wrote:[6] "The Jews invented letters of exchange; commerce, by this method, became capable of eluding violence, and of maintaining everywhere its ground; the richest merchant having none but invisible effects, which he could convey imperceptibly wherever he pleased … From

[6] This passage of Montesquieu (1768) has already been cited by Hirschman (1978, p. 98) as well as by Bates and Lien (1985, p. 60) in the following terms: " 'through this means commerce could elude violence …; for the richest trader had only invisible wealth which could be sent everywhere without leaving any trace. … Since that time, the rulers have been compelled to govern with greater wisdom than they themselves have intended. …' "

this time it became necessary that princes should govern with more prudence than they themselves could even have imagined" (Book 21, chapter XXI-20). High exit power gives rise to the use of price strategies (offering a lower price of protection) to attract mobile subjects. Borrowing De Long and Shleifer's (1993) terminology, we contend that the European *feudal anarchy*, or the presence of competing landlords and the peasant's power to exit from one territory to another, was the source of price competition in the protection market before the fourteenth century.[7]

*Table*
The Nature of the Market for Protection

| Assets structure | Degree of exit power | Nature of the market for protection |
|---|---|---|
| Perfectly redeployable assets | high | buyers' protection market |
| Partly redeployable assets | medium | hybrid protection market |
| Nonredeployable assets | low or nil | sellers' protection market |

In sharp contrast, subjects possessing nonredeployable assets are trapped and have no exit option. In this case, empires use coercive ways to attach and confine subjects to their territory. "From Mesopotamia to China, Egypt, Mesoamerica, or feudal Europe [after the fourteenth century according to Bloch, 1966], serfs were tied to the land and free peasants typically had no outside options, with rulers coming and going without any change in their incentives for production" (Konrad and Skaperdas, 2012, p. 429). In such cases, the supply side is able to dominate the protection market: this is a *sellers' protection market*. This case is the most commonly studied in the empire-building literature.

Last, but not least, there is an intermediate situation between these two polar cases. Indeed, people can possess partially redeployable assets. In this case, exit power exists but is limited by the switching costs induced by the change from one provider to another. Medium exit power involves a protection market in which both merchants (possessing highly redeployable assets) and peasants attached to the soil coexist. As a consequence both strategies are possible: using force to coerce and trap people, or offering a low price of protection to attract demand. We call this specific setting the *hybrid protection market*.

The nature of the protection market is fundamental to understanding the behavior of empires throughout history. Obviously, there exists a continuum of behaviors depending on the institutional, political, or ecological context. The empirical diversity of historical evidence notwithstanding, in the next subsection we distinguish

[7] In fact, De Long and Shleifer (1993, p. 681) acknowledge this type of competition: "In other cases jurisdictions were so small that merchants could flee to feudal domains that provided protection, and competition between petty despots to attract merchants and their commerce constrained arbitrary exactions. In still other cases the most powerful political units in feudal anarchy turned out to be mercantile republics, which owed their self-government to the inability of feudal authorities to enforce commands."

two ideal-type behaviors and empires according to the strategies used on the protection market.

### 2.2 *Absolute and Differential Protection Rents and Empire-Building*

We distinguish two strategies of expansion: use coercive means to fetter people and loot them, or attract them by offering a lower protection price. The first strategy describes the behavior of *absolute-rent-maximizing empires* (hereafter AR-maximizing empires), and the second depicts that of *differential-rent-maximizing empires* (hereafter DR-maximizing empires). In our paper, we consider a dyadic competition between two empires over a given population $N > 0$.

#### 2.2.1 AR-Maximizing Empires

The most studied path of empire-building is defined by the use of threat and violence by the central authority. In this case, assets are mostly nonredeployable and protection is seen as a racket.[8] The territorial expansion is based on coercion of subjects whose nonredeployable assets make them easy targets for looting. An AR-maximizing empire extorts an *absolute protection rent*, which corresponds to the difference between protection revenues and the costs induced by this activity. The absolute protection rent, or *tribute*, is a rent for *sellers* of protection directly derived from the use of coercive means. Following Alesina and Spolaore (2003, p. 69), decisions "are taken by rent maximizing governments who care about their own welfare, and that of their close associates, rather than the welfare of their citizens." Consequently, the aim of an empire of this kind, say Empire 1, is to maximize the following function:

$$AR_1 = N_1\bar{\alpha} - F_1 - S_1, \tag{1}$$

where $N_1 \in [0, N]$ corresponds to the population protected by Empire 1 (its size), $\bar{\alpha}$ represents the price of protection extorted from each of its subjects, and $F_1$ and $S_1$ are the (positive) amounts of resources devoted respectively to fighting effort and safeguarding effort. To draw a parallel with the inputs of production in standard microeconomics, $F_1$ and $S_1$ are the "inputs of protection." The fighting effort, $F_1$, corresponds to the resources allocated to appropriative activities (Hirshleifer, 2000) and represents both efforts to rob subjects and efforts to threaten potential competing empires. In contrast, the safeguarding effort, $S_1$, comprises the resources devoted to the maintenance of the governance structure and the enforcement of the contracts.[9] Both $F_1$ and $S_1$ are required by an AR-maximizing empire in order to

[8] Following Tilly (1985, pp. 170–171), we consider a racketeer as "someone who produces both the danger and, at a price, the shield against the danger." Consequently, an AR-maximizing empire appears to be a racketeer or, in a certain sense, a mafia.

[9] Safeguarding efforts correspond to the resources devoted to ensure internal stability of the empire. For example, they can embrace both efforts to secure property rights of subjects and efforts to eschew a putsch or to enhance the political influence of a country over other countries through diplomatic maneuvers.

rule over a territory $N_1$. Last, it should be noticed that $\bar{\alpha}$ stands for the maximum level of extortion that can be borne by subjects. It is exogenously determined by the characteristics of the population protected and the empire's capacity for raising taxes.

Payoffs received by an AR-maximizing empire for its protection activity, $\pi_i^{AR}$, may be simply defined as

$$\pi_i^{AR}(F,S,\alpha) = \begin{cases} AR_i & \text{if } AR_i \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $F,S,\alpha$ are vectors such that $F = (F_1,F_2)$, $S = (S_1,S_2)$, and $\alpha = (\alpha_1,\alpha_2)$. Payoffs given by (2) highlight a *profitability* condition: if the absolute rent for protection is negative, empires do not provide protection and receive 0.

The Russian Empire is a good example of an AR-maximizing empire: "From top to bottom, the emerging structure of social relations depended on coercion" (Tilly, 1990, p. 141). Other historical examples include the Polish, Hungarian, Serbian, and Brandenburg states. This system involves forced labor, landlord relationships, and the development of the government's armed forces. In contrast, trade routes tend to be thin and to lack capital. This type of hierarchy is tailored to maximize AR. The existence of a coerced subscription of subjects is the cornerstone of AR empire-building. Here, violence is used for plunder and has a *welfare-degrading* effect.

### 2.2.2 DR-Maximizing Empires

In some situations, subjects are not tied to the soil, because of their exit power or the abolition of feudalism. In this case, the use of coercion is not effective, because subjects can flee from one protector to another. In other words, providers of protection are constrained by the preferences of the buyers of protection, due to the existence of competitors (Volckart, 2002). Then, a DR-maximizing empire has to attract mobile subjects by offering a low price of protection in order to extend its territory. Instead of using coercive means, this kind of empire seeks to provide an economic advantage for its subjects. Empire-building is thus mostly based on the consent of the buyers of protection. Another explanation can be found in the *identity* of the holders of the political power. Indeed, due to merchants' political power or their influential position in the state, the ruler may need to furnish the economic conditions favoring commercial profit accumulation (by lowering the protection price) to remain in power. In some cases, as for the Venetian Empire's Council of Doges, merchants are rulers themselves. In other words, a DR-maximizing empire acts in the interests of the "public," or it might be said that the ruler acts as a benevolent dictator,[10] and the expansion of imperial territory is driven by the pursuit of

[10] The benevolence of such an empire is with regard to its own citizens (particularly merchants, industrialists, and bankers) but not with regard to the inhabitants of colonies under the domination of the empire. Thus an empire can be both benevolent within and

higher revenue for its subjects. It gives rise to a protection rent for buyers of protection, capturing the economic advantage of being protected by one empire rather than another. We call this a *differential protection rent (DR)*. This type of rent is almost completely ignored in the literature. The differential protection rent for a subject protected by Empire 1 is given by

$$DR_1 = (\alpha_2 - \alpha_1) + \phi N_1, \tag{3}$$

where for $i = 1,2$, $\alpha_i \in [0,\bar{\alpha}]$ stands for the price of protection that must be paid to be protected by Empire $i$, and $\phi$ is a positive parameter accounting for the benefits of being part of a vast empire. The function (3) identifies two components shaping the economic advantage provided by Empire 1. First, price of protection constitutes a cost borne by a subject. For example, a local merchant who incurs less protection costs obtains a higher profitability in selling his/her products. Therefore, the price difference between providers of protection is a fundamental component of the economic advantage. The existence of different price strategies ($\alpha_1 \neq \alpha_2$) in the market for protection appears here to be a key factor in analyzing empire-building. Second, the economic advantage enjoyed by a subject depends on the size of the empire's territory. We consider that a subject positively values the size of the empire to which she/he belongs ($\phi > 0$). The value of $\phi$ depends on the type of buyers of protection. Typically, a merchant attaches much importance to the size of an empire because it constitutes a domestic market in which he can more easily sell his products (high $\phi$). By contrast, a peasant under feudalism does not attach particular importance to the size of the empire to which s/he belongs (low $\phi$). With the abolition of feudalism, the same peasant, as a rural free laborer, has a stake in the size of the empire, since it offers her/him a larger labor market. One should notice that $\phi$ measures the preference of a DR-maximizing empire for controlling a vast territory. To summarize, DR includes a price component ($\alpha_2 - \alpha_1$) and a size component ($\phi N_1$).

The payoff function of a DR-maximizing empire is

$$\pi_i^{DR}(F,S,\alpha) = \begin{cases} f(DR_i) & \text{if } AR_i \geq 0, \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

where $f(\cdot)$ is a positive and strictly increasing one-argument function capturing the form of the relation between the differential rent of protection provided for its subjects and its own payoff. Indeed, a DR-maximizing empire may be ruled by a benevolent dictator willing to maximize the sum of subjects' economic advantage. The ruler may also be elected and aim at increasing the median subject's DR, or may consist of influential merchants promoting low protection prices and the extension of the domestic market size. The functional form of $f(\cdot)$ depends on the characteristics of the ruler. Finally, it should be noted that the profitability condition

predatory regarding its external colonies. However, the latter aspect is not the focus of our paper.

still holds for DR-maximizing empires: if protection activity is welfare-degrading (i.e., $AR_i < 0$), they will not provide protection and will receive 0 as payoff.

The Venetian merchant empire is a classic example of a DR-maximizing empire: it extended to Cyprus until 1573 and to Crete until 1669. The city's forces launched wars to maintain access to commercial opportunities, and to challenge rivals such as Genoa. However, "more than anything else, its rulers gained reputations from the ability to wage canny and successful sea wars *at relatively low cost to the city's merchants, bankers, and manufacturers*" (Tilly, 1990, p. 147, emphasis added). Venetian empire-building is clearly related to the interests of protected people. Indeed, the Venetian Empire belongs to our second ideal type, namely the *DR-maximizing empire*. In this kind of empire, either there is a ruling group composed of merchants (as in Venice), or merchants have an influential position within the parliament (as in the Dutch and British empires). Similarly, the abolition of feudalism after the French revolution in 1789 and the creation of a nation-state furnished the basis for popular sovereignty. Therefore, the use of violence here is more welfare-enhancing, because it aims at maximizing economic advantage of each domestic merchant or citizen.

## 3 Competitions and Markets for Protection

It should be recalled that conflict models have mostly favored *predatory competition*[11] (Konrad and Skaperdas, 2012), corresponding to a specific kind of market for protection, a *sellers' protection market*. In this section we introduce three structures of protection markets and their implications in terms of competition among empires. These three cases share some common features. First, they assume dyadic competitive relationships between risk-neutral empires. These empires are hierarchical structures, and despite their possible differences in governance, they all have the capacity to guarantee physical and legal protection of their territories by bearing safeguarding costs. Second, the market for protection is characterized by the confrontation of a demand and a supply that determines the size of the empire – which corresponds to its population. In other words, there is a strict equivalence between the population and the size of an empire. Third, we assume that protection is homogeneous and indivisible in the sense that one cannot be partially protected.

### 3.1 Market for Protection with Immobile Subjects

This case corresponds to an extreme situation in which demanders of protection only possess nonredeployable assets. As they are strictly immobile, the price of protection does not play a role in empire-building. Consequently, the only strategy available to expand the empire's territory is to use coercion. We call such a market a *sellers' protection market*.

[11] *Predatory competition* does not pertain to predatory pricing. It describes competition through the use of brute force among AR-maximizing empires.

*Demand for Protection.* Due to the absence of exit power and of redeployable assets, subjects are considered to be immobile. When the protection market is a sellers' one, a subject who buys protection is not just a consumer, because she cannot easily change her supplier of protection. She owes *allegiance* to the empire protecting her. Following Skaperdas (2003), price competition is not conceivable on this kind of market. The size of empires is a result of competition through threats and violence, called *predatory competition* (Konrad and Skaperdas, 2012). As they cannot escape from their providers of protection, they each suffer extortion at a level $\bar{\alpha}$. A sellers' protection market is characterized by the absence of a real demand function: protection appears here as an involuntary transaction dictated by providers of protection.

*Supply of Protection.* This is the classical case explored by the literature: competition over turf by use of violence in which people are tied to their land by law and/or by force. Switching costs borne by subjects are infinite: they are at the mercy of AR-maximizing empires. Each empire $i$, for $i = 1,2$, chooses the level of fighting effort maximizing $\pi_i^{AR}$, taking into account the existence of its opponent. In conformity with conflict theory, the whole territory, $N$, is divided between the two contestants according to a *contest success function* (hereafter CSF). In this paper, we adopt the nonprobabilistic interpretation of a CSF, according to which its outcome can be regarded as a sharing rule for the territory. In particular, we use the form suggested by Grossman and Kim (1995) to determine the share of the population protected by each empire:

$$p_1(F_1,F_2) = \begin{cases} \dfrac{F_1}{F_1+\theta F_2} & \text{if } F_1+F_2>0, \\ \dfrac{1}{2} & \text{if } F_1+F_2=0, \end{cases} \tag{5}$$

where $p_1 \in [0,1]$ is the share of the population protected by Empire 1, with $p_2 = 1 - p_1$, and $\theta$ represents the effectiveness of one unit of fighting effort devoted by Empire 2 relative to Empire 1. A value greater (less) than one indicates that the fighting effort of Empire 2 is more (less) effective.

Intuitively, $p_1$ depends positively on the fighting effort of Empire 1 and negatively on that of Empire 2. However, territorial expansion cannot be infinite, due to organizational difficulties (partly because of the higher costs of managing heterogeneous cultural groups composing a vast empire) as well as protection problems (because borders are strategic elements that must also be defended against external aggressions). Consequently, we consider the existence of an upper bound $Z_i > 0$ for $i = 1,2$, representing the empire's zone of influence (Vahabi, 2004). This zone depends mostly on two elements: the level of safeguarding effort expended, and the efficiency of this effort. First, the greater the amount of resources devoted to safeguarding, the larger is the territory that can be controlled and secured (Findlay, 1996). Second, the zone of influence is defined by the costs associated with the protection of subjects, $e_i > 0$. The level of $e_i$ hinges on topographical characteristics of the territory (it is more costly to ensure protection in mountainous areas, for example) and on the docility of subjects. Consequently the zone of influence of Empire 1 can be defined as $Z_1 = S_1/e_1$, where $1/e_1$ could be seen as a measure of the efficiency of safeguarding by Empire 1. This coefficient captures the organizational difficulties facing an empire in maintaining its internal stability. High efficiency means that small safeguarding efforts result in a large zone of influence. In the remainder of this paper, an empire will be referred to as more efficient than another if the efficiency of its safeguarding efforts is higher. Considering the influence zone's constraint, the size of Empire 1 in a *predatory competition* is

$$N_1^{\mathrm{Pre}} = \min\{p_1 N, Z_1\}.$$

It should be remembered that, to ensure the existence of Empire 1, the profitability condition must be fulfilled (see (2)), which implies $\bar{\alpha} \geq e_1$. Otherwise Empire 1 does not provide protection.

*Equilibrium.* Optimal effort choices are determined by maximizing the payoffs described in (2). As a result, the Nash equilibrium is the vector of fighting efforts $(F_1^{\mathrm{Pre}}, F_2^{\mathrm{Pre}})$ for which each empire maximizes its AR given the strategy of its opponent. Devoting resources to $S_1$ enlarges the zone of influence, but not directly the effective size of the empire. As a matter of fact, the interior solution entails that empires devote the minimum effort in safeguarding that enables them to rule over their territory. Consequently, $\forall i = 1,2$, $S_i = p_i N e_i$. Introducing these levels of safeguarding effort into the equation for AR, we can compute the following reaction functions:

$$F_1 = \sqrt{\theta N F_2(\bar{\alpha} - e_1)} - \theta F_2 \quad \text{and} \quad F_2 = \sqrt{N F_1(\bar{\alpha} - e_2)/\theta} - F_1/\theta.$$

As we consider two AR-maximizing empires, we logically have $\theta = 1$. Considering the profitability condition fulfilled, it can be shown that a Nash equilibrium exists such that the fighting effort for Empire 1 will be

$$F_1^{\mathrm{Pre}} = \frac{N(\bar{\alpha} - e_1)(\bar{\alpha}^2 - \bar{\alpha}(e_1 + e_2) + e_1 e_2)}{(2\bar{\alpha} - e_1 - e_2)^2}, \tag{6}$$

where $F_1^{\mathrm{Pre}}$ stands for the level of resources devoted to fighting by Empire 1 on a sellers' protection market. We deduce from (6) the equilibrium size of Empire 1:

$$N_1^{\mathrm{Pre}} = \frac{\bar{\alpha} - e_1}{2\bar{\alpha} - e_1 - e_2} N. \tag{7}$$

Observation 1 *According to* (7)*, when predatory competition prevails in the market for protection, empire-building depends positively on the level of extortion of the subjects, on the efficiency of safeguarding efforts of the empire, and on the inefficiency of its opponent.*

A salient historical illustration is the emergence of Russia as a leading continental power after her fight against Frederick II's Prussia in the closing stages of the Seven Years' War (1756–1763). Like Russia, Prussia was an AR-maximizing empire, and Russia won its reputation as a great power on the battlefields of that war. In his investigation of the emergence of the eastern powers during 1756–1775, Scott (2001, p. 50) scrutinized what a "great power" meant in the eighteenth century and why Russia had become one after her fight against Prussia: "it was a state which possessed the material and moral resources to fight a major war without outside assistance. Resources and the ability to mobilize these, together with military and, where appropriate, naval potential, were crucial to this assessment. By this yardstick, Russia's military performance after 1756 made clear that she now ranked as one of the leading continental states. There was rather more to it than this, however; […] all the other powers had to acknowledge Russia's enhanced position, […] a process in which Catherine II's personality and ambitions were to prove crucial."

Considering equation (7), Russia satisfied both conditions for winning a predatory competition against Prussia: (1) a high level of extortion from its subjects (high $\bar{\alpha}$); (2) a higher level of efficiency in safeguarding (lower $e_i$ than Prussia).

In our model, difference of efficiency in safeguarding is a major explanatory factor for victory in the predatory competition. The importance of Russia's rise as a *military* power notwithstanding, her emergence as a *great* power was only achieved after the Seven Years' War. While the former achievement should be traced back to Peter I's contribution in modernizing the armed forces and the recruitment system, the latter is totally due to higher efficiency in safeguarding thanks to Catherine II's diplomatic ability. "In the eighteenth century, statesmanship consisted of the ability to see the entire international system and the diplomatic possibilities it offered, rather than one dimension of it, and this conceptual sophistication in turn was a precondition of true great power status. This suggests why Russia only became a leading European power during Catherine II's reign" (Scott, 2001, p. 18). One example of this difference of efficiency in safeguarding under Catherine II's reign (1762–1796) is the way she managed to keep Poland under her political influence after the Seven Years' War. Catherine's diplomatic cooperation with the Prussian king in deciding the destiny of the Polish fief Courland resulted in the extension of Russia's zone of influence over eastern Europe: "Early in 1763, Prussian support had contributed to the first political success of the Empress' reign: the restoration of Ernst Biron as ruler of the Duchy of Courland and the reimposition of the Russian protectorate" (p. 107). Russia acquired all the attributes of a major European power, such as a modern diplomatic service, under Catherine II and started to think and acted like a first-rank state. In this connection, the difference of efficiency in safeguarding should be considered as Russia's second strong point in the predatory competition.

### *3.2 Market for Protection with Fully Mobile Subjects*

In this case, the market for protection is dominated by the demand side, which can freely choose its provider of protection. As a consequence, a strategy of expansion based on coercion is doomed to failure. Indeed, subjects possess a high degree of exit power due to their access to easily redeployable assets, which influences considerably the expansion policy of empires. We call this situation a *buyers' protection market*.

*Demand for Protection.* The demand for protection expresses the population's need to be protected from any aggression, and to establish a higher authority capable of enforcing property rights. However, due to the full redeployability of assets possessed by the buyers of protection, coercion cannot take place, because it is not effective (Konow, 2014). In this form of market, empire-building is based on the consent of subjects and protection is seen as a voluntary transaction. Owing to the homogeneity of protection, price is the only criterion for choosing the provider of protection.

*Supply for Protection.* With fully mobile subjects, we are in the typical case of *price competition*. Regarding the constraint on territorial expansion, we assume that $Z_i \leq \tilde{N} < N$, where $\tilde{N}$ is a positive parameter representing the maximal size of an empire. In other terms, one empire cannot satisfy the whole demand. The demand addressed to Empire 1 is defined as follows:

(8) $$N_1 = \begin{cases} Z_1 & \text{if } \alpha_1 < \alpha_2, \\ \min\{1/2, Z_1\} & \text{if } \alpha_1 = \alpha_2, \\ \min\{1 - N_2, Z_1\} & \text{if } \alpha_1 > \alpha_2. \end{cases}$$

DR-maximizing empires compete over price in order to attract fully mobile subjects. However, there is a lower-bound price below which the empire becomes economically impoverished. As a result, the strategy of Empire 1 in the buyers' protection market is to offer the lowest possible price, corresponding to the one annulling the AR: $\alpha_1^{\text{Pri}} = \{\alpha_1 | AR_1 = 0\}$, where $\alpha_1^{\text{Pri}}$ stands for the price of protection of Empire 1 in a price competition.

*Equilibrium.* The Nash equilibrium corresponds to the levels of prices $(\alpha_1^{\text{Pri}}, \alpha_2^{\text{Pri}})$ simultaneously chosen by empires maximizing their payoffs (4). Given our assumption concerning $f(\cdot)$ – viz. $f > 0$ and $f' > 0$ – this behavior is equivalent to the maximization of the DR. First, we compute the price offered by Empire 1. In order to provide the highest DR, Empire 1 chooses the level of $\alpha_1$ for which $AR_1 = 0$. Considering (1), we can isolate the price of protection on the left-hand side and obtain

(9) $$\alpha_1^{\text{Pri}} = \frac{F_1 + S_1}{N_1}.$$

Proposition 1 *When price competition prevails on the protection market, no resources are spent on fighting effort. This corresponds to a stable full-cooperation equilibrium.*

Proof Using (3) and (9), we can express Empire $i$'s DR as follows: $\forall i, j = 1{,}2$, $i \neq j$, we have $DR_i = ((F_j + S_j)/N_j - (F_i + S_i)/N_i) + \phi N_i$, and equation (8) yields $\partial N_i/\partial F_i = 0$. Following Skaperdas (1992), a full-cooperation equilibrium exists and is stable if, for $(F_1, F_2) = (0{,}0)$, we have $\partial\pi_1^{DR}/\partial F_1 \leq 0$ and $\partial\pi_2^{DR}/\partial F_2 \leq 0$. Using Leibniz's notation, for $i = 1{,}2$, we can write $\partial\pi_i^{DR}/\partial F_i = (d\pi_i^{DR}/dDR_i)\cdot(\partial DR_i/\partial F_i)$. Computing these partial derivatives, we find that $\partial DR_i/\partial F_i = -1/N_i < 0$, and $d\pi_i^{DR}/dDR_i > 0$ holds by definition. Therefore, $\partial\pi_i^{DR}/\partial F_i < 0$ means that devoting resources to fighting effort would necessarily lead to a decrease in the payoffs of DR-maximizing empires. As a result fighting is a strictly dominated option, because it only entails an increase of $\alpha_i$, which reduces $DR_i$ and $\pi_i^{DR}$. Thus, price competition on the protection market systematically leads to a stable equilibrium without investment in fighting effort.

Proposition 1 is founded on an economic intuition: in a buyers' protection market, subjects are perfectly mobile and can easily escape from their protector. Therefore, devoting resources to fighting efforts is pointless and only results in increasing the price of protection, which reduces the attractiveness of the empire. In this kind of market, coercion is useless. Interestingly, this is very close to a predatory competition with "sufficiently ineffective technology" (Skaperdas, 1992, p. 726). Skaperdas call this situation, in which neither of the empires devotes resources to fighting efforts, "full cooperation equilibrium." According to him, low effectiveness of the technology of conflict results from "primitive means of warfare and widely dispersed population" (p. 732). We argue that a high degree of mobility of buyers of protection is a third explanatory variable. Indeed, it tends to make coercion ineffective and makes a situation of unarmed peace between empires more likely. Moreover, Proposition 1 could be seen as an extension of full cooperation: once the technology of conflict is ineffective, suppliers of protection have to find another way to expand their territory, for example by using price strategies. In other words, engaging in price competition may be rational when the technology of conflict is ineffective.

In the same way as fighting efforts, safeguarding efforts raise the price of protection; however, they enable an extension of the influence zone, which is a necessary condition for building and maintaining frontiers of an empire. As a result, DR-maximizing empires should choose the lowest level of safeguarding effort that ensures that the territory is protected: for a given $N_1$ we have $S_1^{\mathrm{Pri}} = N_1 e_1$. Considering the values of efforts ($F_1^{\mathrm{Pri}} = 0$ and $S_1^{\mathrm{Pri}} = N_1 e_1$), we logically deduce from equation (9) the price of protection in a *price competition*:

$$\alpha_1^{\mathrm{Pri}} = e_1. \tag{10}$$

Therefore, the price of protection in a buyers' protection market only depends on the efficiency of safeguarding efforts. Moreover, we notice that equation (10) does not hinge on Empire 2's behavior. Indeed, as we argued earlier, the objective is to furnish the highest possible DR. Accordingly, the best response of Empire 1 would regularly be the lowest price, $\alpha_1^{\mathrm{Pri}}$.

For the clarity of our demonstration, we now consider that Empire 1 is more efficient, so that $e_2 > e_1$. Empire 1 fixes a price lower than its opponent and ensures protection up to the limit of its zone of influence. The problem is to find the size of this zone. According to (3) and (4), we know that providing the largest economic advantage for its subjects involves a territorial dimension (captured by $\phi > 0$). As a consequence, the program of a DR-maximizing empire is to simultaneously offer low price and a large domestic market. The sphere of influence would be extended to its highest level when $S_1/e_1 = \tilde{N}$. We directly derive that $S_1^{\mathrm{Pri}} = \tilde{N} e_1$. Correspondingly, Empire 2 has to satisfy the residual demand $N - \tilde{N}$, allowing safeguarding efforts $S_2^{\mathrm{Pri}} = (N - \tilde{N})e_2$ and offering a price $\alpha_2^{\mathrm{Pri}} = e_2$.

Based on (8), the following equation defines the size of Empire 1 under price competition in the market for protection:

$$N_1^{\mathrm{Pri}} = \begin{cases} \tilde{N} & \text{if } e_2 > e_1, \\ \min\{N/2, \tilde{N}\} & \text{if } e_2 = e_1, \\ \min\{1 - \tilde{N}, \tilde{N}\} & \text{if } e_2 < e_1. \end{cases} \tag{11}$$

Observation 2 *According to equation* (11)*, when price competition prevails in the market for protection, empire-building mostly depends on the efficiency of safeguarding efforts. The more efficient an empire is in establishing its influence zone, the more attractive a price it is able to offer and the bigger its size will be.*

### *3.3 Market for Protection with Both Mobile and Immobile Subjects*

We now consider a segmented demand for protection: both mobile and immobile subjects are present on the market for protection. Accordingly, there are two possible strategies to expand imperial territory: offer a low price of protection to attract mobile subjects, or use coercion to trap subjects without exit power. We call this situation a *hybrid protection market*. In order to avoid unnecessary computational difficulties and keep the treatment tractable, we only consider the symmetric case in which $e_1 = e_2 = e$. For the same reason, we deliberately exclude the case in which empires are not sufficiently efficient in their safeguarding efforts.

*Demand for Protection.* In this setting the demand for protection is heterogeneous and composed of a portion $\lambda$ of mobile subjects, $1 - \lambda$ being the share of immobile subjects, $\lambda \in [0, 1]$. Here $\lambda N$ might be regarded as the number of merchants or other economic agents possessing perfectly redeployable assets. On the other hand, $(1 - \lambda)N$ is the number of peasants completely attached to their land together with all other immobile economic agents. In line with the previous sections, mobile subjects are attracted by the empire offering the lower price of protection, and immobile subjects are divided by coercion according to the CSF described in equation (5). A caveat is warranted with regard to peasants' status. The abolition of feudalism and land reform result in higher mobility of peasants by transforming them into free rural and urban laborers. In this situation, peasants should not be considered as immobile at all times and under all circumstances. Instead of differentiating two specific groups of population as mobile and immobile, we may interpret $\lambda$ as an indicator of the *average* redeployability of assets possessed by demanders of protection.

*Supply of Protection.* Both kinds of empire are able to provide protection. We can distinguish three different setups: competition with two AR-maximizing empires, with two DR-maximizing empires, and with one of each kind. However, the first two lead to an equal division of the territory because of the empires' symmetrical position. Consequently, we focus on the case in which an AR-maximizing empire (say Empire 1) coexists with a DR-maximizing empire (Empire 2). We call it *mixed competition*.

In a mixed competition, empire-building of Empire 1 is driven by the maximization of its payoffs as described by (2). In contrast, Empire 2 aims to maximize (4), which is equivalent to providing the greatest economic advantage for its subjects.

LEMMA *In a symmetrical mixed competition, the AR-maximizing empire never attempts to attract mobile subjects ($\lambda N$).*

PROOF Mobile subjects switch from one provider of protection to another without cost. They choose to be protected by the empire offering the lower price. In other words, to attract mobile subjects Empire 1 has to decrease its price so that $\bar{\alpha} \leq \alpha_2$. Meanwhile, Empire 2 maximizes its DR so that equation (10) holds, i.e., $\alpha_2 = e_2 = e$. As a result, to be attractive, Empire 1 needs to offer $\bar{\alpha} \leq e$. Consider now the function (1) with the minimum level of effort devoted to safeguarding $S_1 = eN_1$ and $\bar{\alpha} \leq e$:

$$AR_1 = N_1\bar{\alpha} - F_1 - N_1 e < 0.$$

As a result, cutting the price of protection is not a viable strategy for an AR-maximizing empire, because it necessarily leads to negative AR.

The Lemma implies that Empire 1 never competes over price, because this would lead to a price war and the disappearance of the AR (or worse). Consequently, the demand for protection addressed to Empire 1 only includes a share $p_1$ of the immobile population, so that $N_1 = F_1(1-\lambda)N/(F_1+\theta F_2)$. Moreover, we have $\bar{\alpha} > \alpha_2$, meaning that Empire 2 attracts all the mobile subjects: $\lambda N$. It is also able to adopt a strategy of expansion based on coercion (especially when $\phi$ is high) to enlarge its territory and thus faces the following demand: $N_2 = \lambda N + \theta F_2(1-\lambda)N/(F_1+\theta F_2)$. In this situation, $\theta$ captures the relative military effectiveness of the DR-maximizing empire over the AR-maximizing one. We assume that $\theta > 1$: fighters of the DR-maximizing empire are more effective on the battlefield.

To justify the assumption $\theta > 1$, we should start by asking about the nature of the army in an AR-maximizing empire in which soldiers have very low (if any) exit option. In an economy with a peasant majority, soldiers often consist of peasant conscripts paying their tributes to the king or lord. On the abolition of feudalism, peasants acquire an exit option and the army has to be restructured. The alternative is to have a volunteer army. This was the principle advocated initially by French revolutionaries and applied systematically in 1791 and 1792 (Forrest, 1990, chapter 3). But the practicality of this principle depended on a sufficient number of young men eager to defend the revolutionary cause. It was assumed that peasant boys who had declined to serve in the armies of the *ancien régime*, and who had gone into hiding rather than submit themselves to the ballot or the militia, would respond to the nation's call. But as Forrest (1990, 2002) shows, this is a risky assumption. Although the ideal of the "nation in arms" was enshrined in the Revolution's legal codes, its implementation in practice was not so easy. A pragmatic way was found by trial and error in 1793 by Jacobins: soldiers got *voice* (Hirschman, 1970, 1978) – political rights as citizens. In the *ancien régime*, the army was aristocratic. Soldiers as subjects had no rights; the nobility had all the rights. Given the higher exit option of liberated peasants from the yoke of feudalism, the army had to provide them voice so that they could find their place in a modernized, reformed army as citizens, not subjects. In other words, major institutional changes such as the abolition of feudalism led not only to higher economic performance but also to higher fighting efforts through corresponding institutional changes in the army.

The military effectiveness of a DR-maximizing empire is augmented because of the higher level of mobility (exit option) for soldiers and their increased political rights as citizens within the army: "The very fact that the soldier was addressed and treated as a privileged member of French society gave him a status and a self-respect which had been denied to the armies of the *ancien régime*. The most critical change sought by the Revolution, however, concerned the political status of the soldier. If he were to be a citizen first and soldier second, he must include political rights, rights which were universally denied to serving soldiers elsewhere in Europe" (Forrest, 1990, p. 193). The structure of the revolutionary armies was more democratic than any in Europe at the time, with noble privilege terminated and military careers opened to talent. The infantryman enjoyed the same rights as his officers. The creation of a mass army instead of an army of elites hinges upon this increased role of soldiers. The French revolutionary and Napoleonic wars ushered in an era of total wars. This explains why in the long term the lessons learned from fighting the French were to bring reforms to every land army in Europe.

As a consequence, $\theta > 1$ seems to be a reasonable assumption. Assuming $\theta > 1$, one could argue that this improved military effectiveness of a DR-maximizing empire is due to higher mobility of the population. Based on the evidence that military effectiveness of an empire is partly related to the potential losses borne by a subject (now a *citizen*) in case of a switch in the provider of protection, we have $d\theta/d\lambda > 0$.

Indeed, the higher the average asset redeployability, the larger are the costs of being protected by the AR-maximizing empire. Accordingly, this difference tends to reinforce the desire (respectively, aversion) to be part of a DR-maximizing (respectively, an AR-maximizing) empire.

*Equilibrium.* Under a mixed competition, each empire pursues simultaneously its own objective described by (2) and (4). Put another way, Empire 1 maximizes its AR and Empire 2 maximizes its DR. Following the Lemma, Empire 1 does not attract any mobile subjects, because price-cutting strategies are not compatible with the maximization of its AR, and it offers a price $\bar{\alpha}$. Consequently, the Nash equilibrium corresponds to the vector of fighting efforts $(F_1^{\text{Mix}}, F_2^{\text{Mix}})$ maximizing the objectives of both empires. First, we compute the reaction function of Empire 1, considering the technology of conflict described by equation (5). Following the same methodology as in section 3.1, it can be directly derived that $F_1 = \sqrt{\theta F_2 (1-\lambda) N (\bar{\alpha} - e)} - \theta F_2$.

Second, we focus on the behavior of Empire 2. In order to avoid computational difficulty, we assume that an increase in fighting effort leads to a constant increase in the price of protection. Accordingly, we consider $\partial \alpha_2 / \partial F_2 = \psi > 0$; $\psi$ captures how one unit of resource spent on fighting effort affects the price offered by the DR-maximizing empire. In other terms, $\psi$ corresponds to the opportunity costs of one unit of resources devoted to fighting effort. Logically, with a high value of $\psi$, warlike strategies are more costly in terms of DR. Based on this assumption, we are able to identify the reaction function of Empire 2: $F_2 = \sqrt{\phi F_1 (1-\lambda) N / (\psi \theta)} - F_1/\theta$. The Nash equilibrium is characterized by the intersection of the two reaction functions

$$F_1^{\text{Mix}} = \frac{(1-\lambda) N \phi (\bar{\alpha} - e)^2 \psi \theta}{[\psi(\bar{\alpha} - e) + \phi\theta]^2}, \qquad F_2^{\text{Mix}} = \frac{(1-\lambda) N \phi^2 \theta (\bar{\alpha} - e)}{[\psi(\bar{\alpha} - e) + \phi\theta]^2}. \tag{12}$$

Using the optimal level of fighting efforts given by equation (12), we deduce the size of each kind of empire:

$$N_1^{\text{Mix}} = \frac{(1-\lambda) N}{1 + \left(\dfrac{\phi}{\psi}\right) \dfrac{\theta}{\bar{\alpha} - e}}, \qquad N_2^{\text{Mix}} = \lambda N + \frac{(1-\lambda) N}{1 + \left(\dfrac{\psi}{\phi}\right) \dfrac{\bar{\alpha} - e}{\theta}}. \tag{13}$$

PROPOSITION 2 *The level of assets redeployability is positively related to the territorial domination of DR-maximizing empires over AR-maximizing ones.*

PROOF According to equation (13), the territorial domination of the DR-maximizing empire over the AR-maximizing empire is captured by the ratio $N_2^{\text{Mix}} / N_1^{\text{Mix}} = (N - N_1^{\text{Mix}}) / N_1^{\text{Mix}}$. We compute the following first derivative:

$$\frac{\partial N_1^{\text{Mix}}}{\partial \lambda} = -N \left[ \frac{\left( \dfrac{\phi}{\psi(\bar{\alpha} - e)} \right) \left( \theta + (1-\lambda) \dfrac{d\theta}{d\lambda} \right)}{\left( 1 + \left( \dfrac{\phi}{\psi} \right) \dfrac{\theta}{\bar{\alpha} - e} \right)^2} + 1 \right] < 0.$$

Accordingly, we find that the size of an AR-maximizing empire is negatively related to the degree of assets redeployability possessed by the buyers of protection. By construction, we have $N = N_1^{\text{Mix}} + N_2^{\text{Mix}}$, and we logically deduce that an increase of $\lambda$ raises the territorial domination of the DR-maximizing empire over its opponent.

Proposition 2 finds justifications through two main channels. The first, a direct one, is the positive link between assets redeployability and population mobility. Indeed, higher mobility leads to lower effectiveness of coercion. As a consequence, the AR-maximizing empire is not able to trap a higher share of the population, and its size decreases. A second, indirect channel is the interaction between the level of assets redeployability of subjects ($\lambda$) and the relative effectiveness of the DR-maximizing empire's fighters ($\theta$). Indeed, an increase in assets redeployability raises the cost borne by a subject to be protected by the AR-maximizing empire. As a consequence, fighters protected by the DR-maximizing empire have more to lose if they change their provider of protection. They are encouraged to fight harder, and the relative advantage of the DR-maximizing empire – captured by $\theta$ – increases. If the DR-maximizing empire is more effective in fighting, the AR-maximizing empire cannot conquer a large territory (lower $N_1^{\text{Mix}}$). This is the impact of major institutional changes (for example, the abolition of feudalism) on the economic performance (higher average level of mobility) as well as on the fighting effort of soldiers through accompanying institutional changes at the army level (the increasing role of soldiers as citizens).

The French revolutionary wars (1792–1799) might be considered as a salient illustration of wars between AR-maximizing empires and a DR-maximizing empire. We exclude Napoleonic wars after the proclamation of Napoleon as emperor (1804–1814), since they were wars of conquest or imperialist wars as opposed to revolutionary wars aiming at defending and consolidating the revolution: "[T]he revolution had begun to clash with established European interests and was being perceived as a threat to the international order. For that matter, war for conquest and war for the consolidation of the revolution were held to be different affairs in Paris" (Rothenberg, 1988, p. 206). Undoubtedly, Forrest (2002, pp. 10–11) is correct in arguing that the dividing lines between revolutionary wars and Napoleonic imperial wars were blurred in reality: "In 1790 the revolutionaries may have promised to liberate the other peoples of Europe, but by 1795 their goal was already territorial conquest, the creation of a buffer of friendly states to France's east, and beyond that, expansion into Germany, Italy and central Europe." It is also true that French armies looted and pillaged. For instance, in 1795 around one-third of Belgium's food crops were seized to help feed the army, while paintings and other valuables were shipped back to Paris, often to adorn the newly created national museum in the Louvre (p. 11). However, the main driving force behind the revolutionary wars was to defend the revolution against the reactionary conservative forces of Europe (and could thus be considered as safeguarding efforts). The vast voluntary national mobilization in France in 1791 and 1792 for the cause of revolution, as reported by Forrest (1990) himself, confirms the fundamental difference between the revolutionary wars and the Napoleonic wars. Moreover, even in Napoleon's army, the tradition of revolutionary wars persisted with respect to the role of soldiers in the army and meritocracy in the promotion of officers. To put it differently, the French Republic (1789–1804) was behaving as a DR-maximizing empire menaced by an alliance of reactionary forces of AR-maximizing empires, whereas France under the reign of the new emperor Napoleon (1804–1814) did not stand for a typically AR-maximizing empire. It experienced a *transition* from a DR- to an AR-maximizing empire in an effort to conquer other European empires by transforming the army into a new privileged caste.

In our theoretical framework, both the initial French victories during the revolutionary republican wars and its final defeats at the end of Napoleonic wars (1810–1814) can be explained in terms of a change in the strategy of empire-building. While the victories and the extension of the empire were related to a DR-maximizing strategy, the retreat and final defeat in the Napoleonic wars were caused by an AR-maximizing strategy. At the start, the French were considered to be liberators; in the end, they were perceived as new imperialist conquerors.[12]

Not only were the French radical revolutionaries, such as the Jacobin Robespierre, initially against external wars,[13] but also "Great Britain, the traditional rival, was resolved on a policy of peace. It regarded all wars as inimical to trade and British financial interests" (Rothenberg, 1988, p. 206). Although the revolutionary and Napoleonic wars involved the whole of Europe for two decades (1792–1815), from 1815 until the First World War the British Empire was the least costly one in the world. Compared to Russia as an AR-maximizing empire, Great Britain was a DR-maximizing one. The difference between the two empires with regard to defense costs was colossal: "by 1913, the average Russian had 50 per cent more of his income appropriated by the state for current defense than did the average Englishman, even though the average Russian's income was only 27 per cent of that of his British contemporary" (Lieven, 1983, pp. 13–14, quoted by Paul Kennedy, 1989, p. 304). The Manchester school preached peace, minimal government expenditures (especially on defense), and reduction of state controls over the economy and the

[12] This change is reflected in Ludwig van Beethoven's decision regarding his initial dedication of his Third Symphony (Sinfonia Eroica) to Napoleon Bonaparte, whom he believed a hero of the French Revolution (1789–1799). He withdrew the dedication when Napoleon proclaimed himself Emperor on May 14, 1804 (see the testimony of Ferdinand Reis, Beethoven's secretary, in George, 1998).

[13] In fact, despite the Girondins' preference for launching war against Austria and other allied reactionary forces, the National Assembly explicitly opposed the initiation of war: Convinced that the "victory of liberty over despotism" spelled an end to wars, the National Assembly resolved in May 1790 that the "French nation renounces the initiation of war for the purposes of conquest," and Victor Comte de Mirabeau, its most influential early leader, proclaimed in August that "the moment is not far off when liberty will acquit mankind of the crime of war." Finally, the French Constitution of September 1791 incorporated the renunciation of "war for the purpose of conquest" in Article 6 (Rothenberg, 1988, p. 205).

individual. Therefore, according to Paul Kennedy (1989, p. 195), the modernization that occurred in British industry and communications was not paralleled by improvements in the army. The economic dominance of the British was not quite reflected in their military power. However, the British navy and British colonialism were more or less unchallenged. Between 1815 and 1880, the British Empire existed in a power-political vacuum, which explains why the colonial army could be kept relatively small (pp. 198–199). This was a source of permanent tension between local British governors in colonies and political and economic decision-makers in Britain.

Investigating British imperial policy during the first three quarters of the nineteenth century, John Galbraith (1960) noted that the Indian policy was largely made by local English governors in Calcutta rather than by "Whitehall" and "the City" in London or Manchester. In the eyes of the British statesmen (in both the Liberal and Conservative Parties) as well as the directors of the East India Company, commerce required peace, and war was costly. Consequently, British governors had to devote their attention to maintaining tranquility within their borders. In contrast to this viewpoint, local British governors were partisans of an expansionist policy, since the commerce of British India could not be secure as long as there were military powers on its frontiers.

The notion of "democratic peace" reflects this liberal doctrine of "peaceful trade and industry" or the maximization of differential rents for the buyers on the protection market. This idea prevailed after the Second World War among major Western democratic countries. Interestingly, while citizens of the latter countries could extensively enjoy the exit option, the Soviet-type regimes of the Eastern bloc were deprived of it. Hence, the distinction between empires according to the costs of exit is not only relevant in early modernity, but also applies to the whole Cold War period. The fall of the Berlin wall in 1989 marked a historical turning point in the demise of the Soviet empire with no exit option.

The same story rings true for the 1917 Russian revolution. In October 1917, the Bolshevik party toppled Kerensky's government under the slogan "Peace, land, bread" aimed at the peasants and workers, and the Russian revolution seemed to be a new liberating force during the "ten days that shook the world" (Reed, 1919). However, Stalin's forced collectivization of agriculture in the thirties ushered in a new era in which peasants lost their mobility.[14] The difference between the American empire and the Russian empire after the Second World War might also be expressed as that between an AR-maximizing empire with no exit option and a DR-maximizing empire with a high exit option.

[14] Interestingly, Trotsky (1935) characterized Stalin's regime in the thirties as "Bonapartism."

## *4 Conclusion*

In this paper we have analyzed empire-building through the economic concept of market for protection. In particular, we shed light on the key role played by the assets structure in the market for protection. Indeed, the level of asset redeployability determines the degree of exit power possessed by subjects protected by the empire, which allows them to influence territorial expansion – or not. Consequently, we identify two different paths in empire-building. First, an empire can be built on the basis of the allegiance of trapped subjects attached to their land. Such an empire can expand its territory as long as the tribute extracted from fettered subjects is positive. This is the *AR-maximization strategy*. Second, in the presence of mobile subjects, an empire might attract subjects by *maximizing the DR*. In compliance with historical evidence, we analyze two principal modes of competition: price competition and predatory competition. Our framework fills a gap in that traditional conflict models systematically focus on situations in which different powerful rulers are involved in warfare against each other. Consequently, they dismiss the distinction between absolute and differential protection rents and ignore the existence of price competition in the market for protection. Therefore, there is a glaring gap that does not incorporate the logic of territorial expansion by merchant empires or empires that pioneered the abolition of feudalism. For example, conflict models do not provide a theoretical explanation for the frontiers of the Venetian empire ruled by the Council of Doges. Considering three kinds of market for protection, we offer in this paper a more complete explanation of the expansion rationale of empires, and we highlight two propositions.

Our first proposition could be seen as an analytic continuation of the conflict theory's literature. Indeed, traditional frameworks suppose that the effectiveness of the technology of conflict is negatively related to the probability of a full and unarmed peace. In this paper we argue that the redeployability of assets possessed by the buyers of protection is a key factor in shaping this relation. In fact, we establish a clear link between assets structure and population mobility, which, in turn, diminishes the effectiveness of coercion (low $\partial p_i / \partial F_i$). As a result, high assets redeployability promotes the existence of a price competition with no investment in fighting efforts (i.e., it promotes the existence of an unarmed peace).

In addition, we believe that Proposition 2 presented in this paper is consistent with historical evolution regarding *institutional change*. In our framework, institutional change is measured in terms of factor mobility. Accordingly, the transition from feudalism to capitalism is marked by the demise of landed property and the ascendancy of merchant, financial, and industrial capital as well as the establishment of free rural and urban labor (the domination of what we call *DR-maximizing empires*). This implies a decline in the importance of immobile wealth and a growing share for the most mobile factor of production, namely, capital. In this situation, there is a link between assets redeployability (assets mobility) and the development of capitalism. Since capitalism involves more redeployable assets (high $\lambda$), the modern empires should behave more and more as DR-maximizing empires to attract mobile assets ($N_2^{\mathrm{Mix}}$ becomes larger). Consequently, the *hybrid protection market* depicted in this paper seems particularly relevant regarding historical evidence. Historically speaking, the economic significance of violence changes in accordance with the institutional change from a *sellers'* to a *buyers' protection market* due to the development of capitalism (Vahabi, 2016, chapter 7). While violence is the source of tribute under an AR-maximizing empire, in a DR-maximizing one it is only a means to provide an economic advantage for buyers of protection. In this kind of market for protection, profit-making – capturing larger shares of markets by offering lower prices for local merchants – is more advantageous than rent-seeking – conquering territory by using coercive means. The mobility of factors as an indicator of institutional change explains the relative importance of predatory and price competition in empire-building.

## *References*

Alesina, A., and E. Spolaore (2003), *The Size of Nations*, The MIT Press, Cambridge (MA).

Bates, R. H., and D.-H. D. Lien (1985), "A Note on Taxation, Development, and Representative Government," *Politics & Society*, 14(1), 53–70.

Bloch, M. (1966), *French Rural History: An Essay on its Basic Characteristics*, transl. by J. Sondheimer, University of California Press, Berkeley and Los Angeles (CA).

De Long, J. B., and A. Shleifer (1993), "Princes and Merchants: European City Growth before the Industrial Revolution," *The Journal of Law & Economics*, 36(2), 671–702.

Domar, E. D. (1970), "The Causes of Slavery or Serfdom: A Hypothesis," *The Journal of Economic History*, 30(1), 18–32.

Findlay, R. (1996), "Towards a Model of Territorial Expansion and the Limits of Empire," in: M. R. Garfinkel and S. Skaperdas (eds.), *The Political Economy of Conflict and Appropriation*, Cambridge University Press, Cambridge, pp. 41–56.

Forrest, A. (1990), *Soldiers of the French Revolution*, Duke University Press, Durham and London.

— (2002), *Napoleon's Men: The Soldiers of the Revolution and Empire*, Hambledon Continuum, London and New York.

Galbraith, J. S. (1960), "The 'Turbulent Frontier' as a Factor in British Expansion," *Comparative Studies in Society and History*, 2(2), 150–168.

George, C. T. (1998), "The Eroica Riddle: Did Napoleon Remain Beethoven's 'Hero'?" *Napoleonic Scholarship*, 1(2), available at http://www.napoleonicsociety.com/english/scholarship98/c_eroica.html.

Grossman, H. I., and M. Kim (1995), "Swords or Plowshares? A Theory of the Security of Claims to Property," *Journal of Political Economy*, 103(6), 1275–1288.

Hirschman, A. O. (1970), *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*, Harvard University Press, Cambridge (MA).

— (1978), "Exit, Voice, and the State," *World Politics*, 31(1), 90–107.

Hirshleifer, J. (2000), "The Macrotechnology of Conflict," *Journal of Conflict Resolution*, 44(6), 773–792.

Kennedy, P. (1989), *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000*, Fontana Press, London.

Konow, J. (2014), "Coercion and Consent," *Journal of Institutional and Theoretical Economics (JITE)*, 170(1), 49–74.

Konrad, K. A. and S. Skaperdas (2012), "The Market for Protection and the Origin of the State," *Economic Theory*, 50(2), 417–443.

Lane, F. C. (1973), *Venice, a Maritime Republic*, The Johns Hopkins University Press, Baltimore (MD) and London.

— (1979), *Profits from Power: Readings in Protection Rent and Violence-Controlling Enterprises*, State University of New York Press, Albany (NY).

Levi, M. (1981), "The Predatory Theory of Rule," *Politics & Society*, 10(4), 431–465.

Lieven, D. C. B. (1983), *Russia and the Origins of the First World War*, St. Martin's Press, New York.

Montesquieu, C. de Secondat (1768), *The Spirit of Laws*, Electronic Text Center, University of Virginia Library, Charlottesville (VA).

Nexon, D. H., and T. Wright (2007), "What's at Stake in the American Empire Debate," *American Political Science Review*, 101(2), 253–271.

North, D. C. (1981), *Structure and Change in Economic History*, W. W. Norton & Co., New York.

— and R. P. Thomas (1973), *The Rise of the Western World: A New Economic History*, Cambridge University Press, Cambridge.

Reed, J. (1919), *Ten Days that Shook the World*, Boni and Liveright, New York.

Rothenberg, G. E. (1988), "The Origins, Causes, and Extensions of the Wars of the French Revolution and Napoleon," in: R. I. Rotberg and T. K. Rabb (eds.), *The Origin and Prevention of Major Wars*, Cambridge University Press, New York, pp. 199–221.

Scott, H. M. (2001), *The Emergence of the Eastern Powers, 1756–1775*, Cambridge University Press, New York.

Skaperdas, S. (1992), "Cooperation, Conflict, and Power in the Absence of Property Rights," *The American Economic Review*, 82(4), 720–739.

— (2002), "Warlord Competition," *Journal of Peace Research*, 39(4), 435–446.

— (2003), "Restraining the Genuine *Homo Economicus*: Why the Economy Cannot be Divorced from its Governance," *Economics & Politics*, 15(2), 135–162.

Spolaore, E. (2012), "The Economics of Political Borders," Working Paper 3854, CESifo, Munich, available at SSRN: http://ssrn.com/abstract=2094955.

Spruyt, H. (2008), " 'American Empire' as an Analytic Question or a Rhetorical Move?" *International Studies Perspectives*, 9(3), 290–299.

Tilly, C. (1985), "War Making and State Making as Organized Crime," in: P. Evans, D. Rueschemeyer, and T. Skocpol (eds.), *Bringing the State Back In,* Cambridge University Press, Cambridge, pp. 169–187.

— (1990), *Coercion, Capital and European States: AD 990–1992*, Wiley-Blackwell, Cambridge (MA).

Tracy, J. D. (ed.) (1991), *The Political Economy of Merchant Empires: State Power and World Trade 1350–1750*, Cambridge University Press, New York.

Trotsky, L. (1935), "The Soviet Union Today," *The New International*, 2(4), 116–122.

Vahabi, M. (2004), *The Political Economy of Destructive Power*, Edward Elgar Publishing, Cheltenham.

— (2016), *The Political Economy of Predation: Manhunting and the Economics of Escape*, Cambridge University Press, New York.

Volckart, O. (2000), "State Building by Bargaining for Monopoly Rents," *Kyklos*, 53(3), 265–293.

— (2002), "No Utopia: Government without Territorial Monopoly in Medieval Central Europe," *Journal of Institutional and Theoretical Economics (JITE)*, 158(2), 325–343.

Williamson, O. E. (1985), *The Economic Institutions of Capitalism*, The Free Press, New York.

Antoine Pietri
University of Paris 1
Centre d'Économie de la Sorbonne
Maison des Sciences Economiques
106–112 Boulevard de l'Hôpital
75013 Paris
France
antoine.pietri@univ-paris1.fr

Tarik Tazdaït
C.N.R.S. – E.H.E.S.S. – CIRED
Jardin Tropical
45 bis, avenue de la Belle Gabrielle
94736 Nogent Sur Marne cedex
France
tazdait@centre-cired.fr

Mehrdad Vahabi
Department of Economics
University of Paris 8
2 Rue de la Liberté
93526 Saint-Denis
France
mehrdad.vahabi@wanadoo.fr

# Labor Unionization Structure, Innovation, and Welfare

by

Tien-Der Han and Arijit Mukherjee*

Received October 28, 2014; in revised form January 5, 2016;
accepted April 28, 2016

We show the effects of cooperation among the labor unions with complementary workers on innovation, consumer surplus, and welfare. Although cooperation among the unions reduces wages, it may either increase or decrease the firm's incentive for innovation, and may also make the consumers and the society worse off by reducing innovation. While cooperation (compared to noncooperation) among the unions makes the workers better off, it may not make all final-goods producers better off. (JEL: D43, J51, L13, O31)

## 1 Introduction

What is the effect of labor unionization structure on innovation? This issue is gaining popularity in recent years due to the diversity of unionized labor markets across countries.[1] The effects of firm-specific labor unions are here compared with those of an industry-wide labor union. Under firm-specific labor unions, a firm deals with a labor union that is associated with that firm, whereas under an industry-wide labor union, the labor union deals with all firms in that industry.

In a patent-race model, Haucap and Wey (2004) show that if an industry-wide labor union charges a uniform wage to all firms, the incentive for labor-saving innovation is higher under such a union. However, if the industry-wide labor union can charge different wages to different firms, the incentive for innovation is higher

* Tien-Der Han: School of Business and Economics, Loughborough University, United Kingdom. Arijit Mukherjee (corresponding author): Nottingham University Business School, United Kingdom; CESifo, Munich; INFER, RWTH Aachen, Germany; Global Research Unit, City University of Hong Kong. We thank two anonymous referees for helpful comments and suggestions. The usual disclaimer applies.

[1] See, e.g., Calmfors and Drifill (1988), Moene and Wallerstein (1997), and Flanagan (1999) for the difference in labor unions with respect to the degree of wage-setting centralization. Decentralized wage setting is often contrasted with centralized wage setting. Under decentralized wage setting, wages are set between employers and firm-specific unions, while under centralized wage setting, an industry-wide union negotiates wages with all firms (Haucap and Wey, 2004).

*Journal of Institutional and Theoretical Economics* 173, 279–300 – ISSN 0932-4569
DOI: 10.1628/093245616X14785139493983 – 

under firm-specific labor unions. In a model with R&D competition, Calabuig and Gonzalez-Maestre (2002) show that the incentive for a labor-saving innovation is higher under firm-specific labor unions for nondrastic innovations, while the incentive for innovation can be higher under an industry-wide labor union for drastic innovation. Manasakis and Petrakis (2009) show that, under noncooperative R&D, the incentive for a labor-saving innovation is higher under firm-specific labor unions if knowledge spillover is high, but the incentive for innovation is always higher under firm-specific labor unions under cooperative R&D. They also show that welfare is higher under firm-specific unions than under an industry-wide labor union. Considering an innovating firm and a noninnovating firm, Mukherjee and Pennings (2011) show the implications of the licensing of innovation and the unions' preferences for wage compared to employment in determining the effects of the unionization structure on a labor-saving innovation.

Although the above-mentioned papers provide important insights, they all consider perfectly substitutable workers, while it is often found that labor unions exist for workers providing complementary services. For example, as mentioned in Aghadadashli and Wey (2015, p. 667), in Germany, "hospital doctors are (mainly) represented by the *Marburger Bund* (a craft union), while the remaining workers are represented by the German trade union (*Verdi*). […] The *Deutsche Bahn* (the dominant railway operator) must bargain with the German train drivers union (*Gewerkschaft Deutscher Lokomotivführer*; GDL) and the railway and transport union (*Eisenbahn- und Verkehrsgewerkschaft*; EVG). […] the GDL is a craft union, which is complementary to workers represented by the EVG. The former takes care of the train drivers' employment conditions, and the latter represents the remaining railway workers' interests. Other examples include airlines (where pilots are represented by the *Vereinigung Cockpit*) and airports (where air traffic controllers are organized in the *Gewerkschaft der Flugsicherung*)." Hence, employers are exposed to the unions representing complementary workers.

Different workers organized under different labor unions can be observed in other countries also. For example, different types of workers are organized in different unions in Sweden (see, e.g., Kjellberg, 2014, for a detailed discussion on the labor unionization structure in Sweden).

Upmann and Müller (2014) also provide evidence for labor unions with complementary workers and different types of wage bargaining, viz., separate wage bargaining, where different labor unions with different types of workers bargain separately with firms, and wage bargaining by an encompassing labor union, where a union bargains with firms for all types of workers. There is also other evidence showing that wages set by the unions can be applicable to the entire industry. For example, as mentioned in Haucap, Pauly, and Wey (2001, p. 288), "a common feature of many labor market systems in continental Europe are coverage extension rules. Under these rules, the coverage of collectively negotiated wage contracts can be extended to entire industries through legal means. With coverage extension, some or all employment terms are made generally binding not only for the members of unions and employers' associations, but for all industry participants. In Germany, for example, collective wage agreements between a union and an employers' association can be made compulsory even for independent employers through the so-called *Allgemeinverbindlicherklärung* (AVE), a legal instrument provided for in § 5 *Tarifvertragsgesetz* (TVG). The Ministry of Labor can, on application of either unions or employers' associations, use an AVE to make some or all terms of a collectively negotiated employment contract generally binding for an entire industry, where otherwise only those unions, employers, and employers' associations that have actually negotiated and signed the contract would be directly bound by it (§ 3 I TVG)." They also provide evidence for the coverage extension rules in other countries and mention that the number of AVEs increased from 448 in 1975 to 588 in 1998. Venn (2009) provides evidence of bargaining between the employers and the employees at the industry level for the OECD and selected non-OECD countries. Evidence of industry-wide wage bargaining can be found also in Carluccio, Fougère, and Gautier (2015) for France and in Lamarche (2013) for Argentina.

Given this background, the purpose of this paper is to show the effects of cooperation among the labor unions of complementary workers on innovation and welfare.[2] In what follows, in a model with an innovating and multiple noninnovating firms, we show in section 2 that an encompassing labor union of complementary workers may decrease (increase) the incentive for innovation compared to separate labor unions if the technological improvement through innovation is large (small). Hence, our analysis follows the literature considering competition between innovating and noninnovating firms (see, e.g., Mookherjee and Ray, 1991; Gallini, 1992; Ray Chowdhury, 1995; Mukherjee, 2003; Mattoo, Olarreaga, and Saggi, 2004; and Mukherjee and Pennings, 2004). We show in appendix A.7 that our result holds even if more than one (but not all) firms are innovators.

As discussed in the following analysis, the raising-rivals'-cost[3] motive may explain why an encompassing union (compared to separate unions) may either increase or decrease the incentive for innovation. Under the raising-rivals'-cost motive a firm can take an action that increases the costs of its competitors. Even if that action increases the cost of the concerned firm, the competitors' cost-increase can be significantly more important, so that on balance, this action allows the concerned firm to increase its profit by acquiring a larger market share. We show that innovation by a firm may either increase or decrease the marginal costs of the noninnovating firms. If innovation reduces the marginal costs of the noninnovating firms, it benefits them and may discourage the innovator from innovating in order to raise the costs of its rivals. Although this effect remains under both encompassing and separate unions, the raising-rivals'-cost motive is stronger under the former than the latter, since an encompassing union (compared to separate unions) benefits the innovator with a lower marginal cost by reducing the complements problem discussed below.

[2] Although we consider the input suppliers as labor unions, our analysis is applicable if the input suppliers are profit-maximizing firms charging linear input prices.

[3] See Williamson (1968), Salop and Scheffman (1983, 1987), and Haucap, Pauly, and Wey (2001) for some earlier papers explaining the raising-rivals'-cost effect.

We also show that although an encompassing union reduces wages compared to separate unions, it may make the consumers and the society worse off by reducing innovation. While an encompassing union makes the workers better off than do separate unions, it may not make all final-goods producers better off.

Our paper contributes to the literature following Horn and Wolinsky (1988), Shapiro (2000), Mukherjee and Pennings (2011), Upmann and Müller (2014), and Aghadadashli and Wey (2015). Shapiro (2000) shows that while choosing the prices of complementary inputs noncooperatively, input suppliers do not internalize the negative external effects of their pricing on other complementary input suppliers' revenues, thus creating the complements problem.[4] Cooperation among the complementary input suppliers solves the complements problem and reduces the input prices, thus increasing the profits of the input suppliers and making the consumers better off by reducing the prices of the final goods. We show that although cooperation among the complementary input suppliers (labor unions in our case) reduces the input price (wage in our analysis) by solving the complements problem, it may do harm by reducing innovation. If the latter effect dominates the former, cooperation among the complementary workers may not benefit the consumers and the society.

Our framework is similar to the no-technology-licensing case of Mukherjee and Pennings (2011) with the exception that we consider complementary workers instead of substitutable workers. It follows from Mukherjee and Pennings (2011) that, in the absence of technology licensing, a final-goods producer's incentive for innovation is higher under cooperation among the unions of substitutable workers. In contrast, cooperation among the unions in our analysis may either increase or decrease innovation. This difference is attributable to the different effects of cooperation among the labor unions on wages.

Horn and Wolinsky (1988), Upmann and Müller (2014), and Aghadadashli and Wey (2015) consider the effects of cooperation among labor unions with complementary workers. Horn and Wolinsky (1988) show that the workers are better off (worse off) under cooperation among the unions if the workers are substitutable (complements). In contrast, we show that complementary workers can be better off under cooperation. This difference occurs because, unlike them but like other papers mentioned above, we allow the firms to determine workers after wage determination. Further, unlike Horn and Wolinsky (1988), we consider product-market competition among the final-goods producers and show the effects of cooperation among the unions on innovation and welfare, which is our main focus.

Both Aghadadashli and Wey (2015) and Upmann and Müller (2014) consider that a firm and two labor unions bargain over wage and employment, thus considering an efficient-bargaining model, and show that an encompassing labor union makes complementary workers worse off than do separate unions. In contrast, like all other papers mentioned above, we consider a *right-to-manage* model where the unions (or bargaining between the firms and the unions) determine wage and the

[4] This is also called "royalty stacking" (Gilbert and Katz, 2011).

firms determine employment.[5] Hence, the bargaining structure considered in our paper is different from theirs. Further, unlike them, we consider product-market competition and the effects through innovation. We show that an encompassing labor union makes the complementary workers better off. We also show the effects of different unionization structures on innovation and welfare, which is our main focus.

The remainder of the paper is organized as follows. Section 2 describes the model and derives the results. Section 3 discusses the implications of some of our assumptions. Section 4 concludes. We relegate many mathematical details to the appendix.

## 2 *The Model and the Results*

Assume that there are one innovating firm (firm 1) and $n-1$ noninnovating firms (firms $2,\ldots,n$), where $n \geq 2$. These firms compete like Cournot oligopolists with homogeneous goods, facing the inverse market demand function $P = 1-q$, where $P$ is price and $q$ is the total output. Production of the final goods requires two types of workers, $x$ and $y$, which are unionized. We assume that the workers $x$ and $y$ are perfect complements. The reservation wages of all workers are $c$, which is normalized to zero for simplicity. We consider a right-to-manage model of labor unions where the unions determine the wages and the firms hire workers according to their requirements.

Like Horn and Wolinsky (1988), we consider two types of labor unions: First, separate labor unions, where union $X$ (union $Y$) organizes workers of type $x$ (type $y$) and the unions determine wages $w_x$ and $w_y$ simultaneously to maximize their own utilities. Second, an encompassing labor union, where a single union organizes both types of workers and determines wages $w_x$ and $w_y$ simultaneously to maximize the total utilities of all workers. We assume that an encompassing union can charge different wages for different types of workers but it charges the same wage to different firms for the same type of worker. Charging the same wage to different firms for the same type of workers is in line with many previous works such as Haucap and Wey (2004), Manasakis and Petrakis (2009), and Mukherjee and Pennings (2011), and can be motivated by the empirical evidence cited in the introduction for industry-wide wage bargaining.

To start with, we assume that all firms require one unit of each type of worker to produce one unit of the final good. However, the innovating firm, firm 1, can invest an amount $k$ in R&D to reduce its labor coefficients for both worker types to $s$, where $s < 1$. Hence, to show our results in the simplest way, we consider that innovation is not biased towards any type of worker and creates a neutral technological progress.[6] Thus, innovation increases labor productivities in firm 1 from 1

[5] See, Layard, Nickell, and Jackman (1991) for arguments in favor of the right-to-manage models.

[6] We discuss the implications of this assumption in section 3.

to $1/s$. It is intuitive that if $s$ is very large, the noninnovating firms will go out of the market. In order to ensure that all final-goods producers always hire workers, we assume that $2/(n+2) < s$ (see appendix A.1 for details). Hence, we consider that $s \in (2/(n+2), 1)$. We consider the binary choice for firm 1's R&D decision for analytical convenience. As we discuss below, our main result holds even if firm 1 chooses the extent of technological improvement through R&D. We consider the following game. Conditional on the unionization structure, at stage 1, firm 1 decides whether or not to innovate. At stage 2, wages are determined by the unions. At stage 3, the final-goods producers (firms $1,\dots,n$) determine their outputs simultaneously, and the profits are realized. We solve the game through backward induction.

### 2.1 Separate Labor Unions

If the separate unions $X$ and $Y$ charge $w_x$ and $w_y$ as wages for workers $x$ and $y$ respectively, then firm 1 and the $i$th firm, $i = 2,\dots,n$, determine their outputs by maximizing the following expressions, respectively:

$$\max_{q_1}[1-q-t(w_x+w_y)]q_1 - f, \qquad \max_{q_i}[1-q-(w_x+w_y)]q_i,$$

where $t = 1$ and $f = 0$ under no innovation by firm 1, and $t = s$ and $f = k$ under innovation by firm 1.

The equilibrium outputs of firm 1 and the $i$th firm, $i = 2,\dots,n$, can be found as

$$(1) \quad q_1 = \frac{1-nt(w_x+w_y)+(n-1)(w_x+w_y)}{n+1}, \quad q_i = \frac{1-2(w_x+w_y)+t(w_x+w_y)}{n+1}.$$

It is clear from (1) that a lower wage decreases the output of firm 1, i.e., $q_1$, if $t < (n-1)/n$. Since, in our analysis, the profit of firm 1 is equal to $q_1^2 - f$, a lower wage decreases the profit of firm 1 for $t < (n-1)/n$.

The demand for workers faced by both unions $X$ and $Y$ is $q_x = q_y = tq_1 + \sum_{i=2}^{n} q_i$. Unions $X$ and $Y$ determine their wages by maximizing the following expressions:

$$\max_{w_x} w_x \left( tq_1 + \sum_{i=2}^{n} q_i \right), \quad \max_{w_y} w_y \left( tq_1 + \sum_{i=2}^{n} q_i \right).$$

The equilibrium wages are

$$(2) \qquad w_x^{nc} = w_y^{nc} = \frac{n-1+t}{6(t-1)+3n(2-2t+t^2)}.$$

If $t < 1$, the equilibrium wages fall as the number of firms (i.e., $n$) increases.[7] If $t < 1$, more firms increase the elasticity of demand for workers and reduce the equilibrium wages.

[7] We get that $\partial w_x^{nc}/\partial n = \partial w_y^{nc}/\partial n = (-(2-t)(1-t)t)/(3[-2(1-t)+n(2-2t+t^2)]^2) < 0$.

We obtain from (1) and (2) that the equilibrium outputs of firm 1 and firm $i$, $i = 2,\ldots,n$, are

$$q_1^{nc} = \frac{-4(1-t)+2n^2(1-t)+n(2-2t+t^2)}{(n+1)[-6(1-t)+3n(2-2t+t^2)]}$$

and

$$q_i^{nc} = \frac{-2(1-t^2)+n(2-4t+3t^2)}{(n+1)[-6(1-t)+3n(2-2t+t^2)]},$$

respectively. The total equilibrium outputs are

$$q^{nc} = q_1^{nc} + \sum_{i=2}^{n} q_i^{nc} = \frac{-2n(1-t)-2(1-t)^2+n^2(4-6t+3t^2)}{(n+1)[-6(1-t)+3n(2-2t+t^2)]}.$$

The equilibrium profits of firm 1 and firm $i$, $i = 2,\ldots,n$, are $\pi_1^{nc} = (q_1^{nc})^2 - f$ and $\pi_i^{nc} = (q_i^{nc})^2$, respectively.

LEMMA 1 *Under separate labor unions, firm 1 innovates for*

$$k < \left[\frac{-4(1-s)+2n^2(1-s)+n(2-2s+s^2)}{(n+1)[-6(1-s)+3n(2-2s+s^2)]}\right]^2 - \frac{1}{9(n+1)^2} \equiv k^{nc}.$$

The expression $k^{nc}$ shows firm 1's maximum willingness to pay for innovation under separate unions. Innovation increases firm 1's product-market profit compared to no innovation. However, innovation also imposes a cost on firm 1. Hence, firm 1 innovates if the cost of innovation is not very high. For the proof of Lemma 1, see appendix A.2.

### 2.2 An Encompassing Labor Union

If the encompassing union charges $w_x$ and $w_y$ as wages for workers $x$ and $y$ respectively, then the equilibrium outputs of firms 1 and 2 are given by (1) and the demand for workers is given by $q_x = q_y = tq_1 + \sum_{i=2}^{n} q_i$. The wages are determined by maximizing the following expression:

$$\max_{w_x, w_y} (w_x + w_y)\left(tq_1 + \sum_{i=2}^{n} q_i\right).$$

The equilibrium wages are

$$w_x^c = w_y^c = \frac{n-1+t}{-8(1-t)+4n(2-2t+t^2)}. \tag{3}$$

As under separate unions, we get under an encompassing union that, if $t < 1$, as the number of firms (i.e., $n$) increases, the elasticity of demand for workers increases and the equilibrium wages fall.[8]

[8] We get that $\partial w_x^c/\partial n = \partial w_y^c/\partial n = (-(2-t)(1-t)t)/(4[-2(1-t)+n(2-2t+t^2)]^2) < 0$.

We obtain from (1) and (3) that the equilibrium outputs of firm 1 and firm $i$, $i = 2,\ldots,n$, are

$$q_1^c = \frac{-3(1-t)+n^2(1-t)+n(2-2t+t^2)}{(n+1)[-4(1-t)+2n(2-2t+t^2)]}$$

and

$$q_i^c = \frac{-(2-t-t^2)+n(2-3t+2t^2)}{(n+1)[-4(1-t)+2n(2-2t+t^2)]},$$

respectively. The total equilibrium outputs are

$$q^c = q_1^c + \sum_{i=2}^{n} q_i^c = \frac{-2n(1-t)-(1-t)^2+n^2(3-4t+2t^2)}{(n+1)[-4(1-t)+2n(2-2t+t^2)]}.$$

The equilibrium profits of firm 1 and firm $i$, $i = 2,\ldots,n$, are $\pi_1^c = (q_1^c)^2 - f$ and $\pi_i^c = (q_i^c)^2$, respectively.

LEMMA 2 *Under an encompassing labor union, firm 1 innovates if*

$$k < \left[\frac{-3(1-s)+n^2(1-s)+n(2-2s+s^2)}{(n+1)[-4(1-s)+2n(2-2s+s^2)]}\right]^2 - \frac{1}{4(n+1)^2} \equiv k^c.$$

The expression $k^c$ shows firm 1's maximum willingness to pay for innovation under an encompassing labor union. The intuition for Lemma 2 is similar to that for Lemma 1. For the proof of Lemma 2, see appendix A.3.

### *2.3 Comparison between Separate and Encompassing Labor Unions*

#### 2.3.1 The Effects of an Encompassing Labor Union on Wage and the Final-Goods Producers

The comparison of (2) and (3) gives the following result immediately.

PROPOSITION 1 *Wages are higher under separate labor unions than under an encompassing labor union, irrespective of firm 1's R&D decision.*

Since the workers are complements, an encompassing labor union reduces wages compared to separate unions, thus solving the complements problem and reducing the marginal costs of all final-goods producers compared to separate labor unions.

Since the encompassing labor union reduces the marginal costs of all final-goods producers compared to separate unions, it is immediate that if firm 1 does not innovate (i.e., $t = 1$), it increases the profits of all final-goods producers compared to separate unions. However, as we will show below, this may not be the case if firm 1 innovates.

PROPOSITION 2 *If firm 1 innovates, an encompassing labor union decreases (increases) the profit of firm 1 compared to separate labor unions for* $s \in (2/(n+2), (n-1)/n)$ ($s \in ((n-1)/n, 1)$)*, but it increases the profits of other final-goods producers.*

Although an encompassing union (compared to separate unions) reduces the marginal costs of all final-goods producers, the marginal-cost saving is higher for the noninnovators than for the innovator. It follows from the discussion after (1) that a lower wage reduces the output and profit of firm 1 if $s < (n-1)/n$. Hence, if the technological difference between the innovator and the noninnovators is large (i.e., $s \in (2/(n+2), (n-1)/n)$), the marginal-cost saving under an encompassing union (compared to separate unions) is significantly higher for the noninnovators than for the innovator, thus reducing the profit of the innovator under an encompassing union. For the proof of Proposition 2, see appendix A.4.

### 2.3.2 The Effect of an Encompassing Labor Union on Innovation

PROPOSITION 3 *If $s \in (2/(n+2), s^*)$ ($s \in (s^*, 1)$), where $s^* = (3+4n-7n^2+\sqrt{9-58n^2+49n^4})/4n$, then an encompassing labor union decreases (increases) firm 1's incentive for innovation compared to separate labor unions for $k \in (k^c, k^{nc})$ (for $k \in (k^{nc}, k^c)$).*

The reason for the above result (for the proof, see appendix A.5) is as follows. Firm 1's incentive for innovation depends on the difference in its profit between innovation and no innovation. On one hand, an encompassing union (compared to separate unions) tends to reduce firm 1's incentive for innovation by increasing its profit under no innovation. On the other hand, an encompassing union (compared to separate unions) tends to increase (decrease) firm 1's incentive for innovation by increasing (decreasing) its profit under innovation if its technological improvement through innovation is small (large). It follows from the discussion after (1) that firm 1's profit decreases (increases) with a lower wage if $s <(>)\ (n-1)/n$. Hence, if firm 1's technological improvement through innovation is large, both the above-mentioned effects reduce firm 1's incentive for innovation under an encompassing union compared to separate unions. However, if firm 1's technological improvement through innovation is small, the above-mentioned second effect can dominate the first effect, and an encompassing union can increase firm 1's incentive for innovation.

The above discussion suggests that the raising-rivals'-cost[9] motive is behind the result shown in Proposition 3. Innovation by firm 1 reduces its marginal cost but it may either increase or decrease the marginal costs of the noninnovating firms. If $s$ is sufficiently small (i.e., $s < s^*$), innovation by firm 1 reduces the marginal costs of the noninnovating firms, implying that firm 1 cannot capture the entire benefit from innovation, which benefits also the noninnovating firms. Hence, the raising-rivals'-cost motive may discourage firm 1 from innovating if $s$ is sufficiently small. Although this effect remains under both encompassing and separate unions, the effect is stronger under the former than the latter unionization structure, since an encompassing union benefits firm 1 from a lower marginal cost than with separate

[9] See Salop and Scheffman (1983, 1987) for earlier works explaining the raising-rivals'-cost effect.

unions. Hence, firm 1's motive for raising the rivals' cost is higher under an encompassing union than under separate unions, and its incentive for innovation is lower under an encompassing union than under separate unions if $s$ is sufficiently small.

The above argument suggests that competition in the product market may play an important role for the innovation-reducing effect of an encompassing union. In other words, an encompassing union may not reduce firm 1's incentive for innovation compared to separate unions if firm 1 is a monopolist producer of the product. We show that this is indeed the case. It is easy to check that if $n = 1$ then $s^* < 2/(n+2)$, implying that the range $(2/(n+2), s^*)$ is empty. Hence, an encompassing union does not reduce firm 1's incentive for innovation compared to separate unions if firm 1 is a monopolist producer of the product.

In the above analysis, we have considered an innovator and $n-1$ noninnovators to show the innovation-reducing effect of an encompassing union in the simplest way. We show in appendix A.7 that this result holds if there are multiple innovators but not all firms are innovators. If all firms innovate under an encompassing union, we show in appendix A.7 that the encompassing union does not reduce innovation compared to separate unions. This happens because the raising-rivals'-cost effect mentioned above is absent in this situation.

We have done our analysis under the assumption that the unions set the same wage to different firms, which gets significant support from the empirical evidence. If the unions set different wages to different firms, the incentive for innovation is higher under an encompassing union than under separate unions, again due to the absence of the raising-rivals'-cost effect.

#### 2.3.3 The Effect of an Encompassing Labor Union on the Total Output

It follows from Propositions 1 and 3 that an encompassing union has two opposing effects on the total outputs. On one hand, it tends to reduce wages, and on the other hand, it may reduce innovation by firm 1. The following proposition shows that, depending on the extent of firm 1's technological improvement through R&D, an encompassing union may either increase or decrease the total outputs produced by all firms, and thus may have an ambiguous effect on consumer surplus.

PROPOSITION 4 *Assume that $s \in (2/(n+2), s^*)$ and $k \in (k^c, k^{nc})$. An encompassing labor union decreases (increases) the total outputs of the final-goods producers compared to separate labor unions for $s \in (2/(n+2), s^{**})$ ($s \in (s^{**}, s^*)$), where*

$$\frac{2}{n+2} < s^{**} = \frac{-4+n+3n^2-\sqrt{3(-n^2+n^4)}}{-4+3n^2} < s^*.$$

Proposition 4 (for the proof, see appendix A.6) suggests that although an encompassing labor union creates a beneficial wage effect, which solves the complements problem, its adverse effect on firm 1's innovation may dominate the beneficial wage effect, thus reducing the total outputs of the final-goods producers under an encompassing union, compared to separate unions. Since consumer surplus in our analysis is $q^2/2$, the above result implies that an encompassing labor union makes the consumers worse off compared to separate unions if the technological improvement through R&D is large.

Proposition 3 shows that an encompassing union increases firm 1's incentive for innovation for $s \in (s^*, 1)$ and $k^{nc} < k < k^c$. In this situation, the total outputs of the final-goods producers are lower under separate labor unions with no R&D by firm 1 than under an encompassing labor union with R&D by firm 1, implying that consumer surplus is higher under an encompassing union than under separate unions. The positive wage effect as well as the positive innovation effect helps to reduce the marginal costs of final-goods production, thus making the consumers better off under an encompassing union.

We have considered a binary choice for firm 1's R&D decision (i.e., firm 1 innovates or does not innovate). As a result, firm 1 may not innovate under an encompassing union if the technological improvement through R&D is not small. However, the situation of no innovation by firm 1 under an encompassing union is an extreme one and is an artifact of the binary choice. If firm 1 could choose the extent of technological improvement through R&D, say, by investing $F(s) = s^2/2$ to reduce its labor coefficient by $s$, it would innovate under an encompassing union, but the extent of technological improvement could be lower under an encompassing union than under separate unions. Hence, even if firm 1's R&D decision is not a binary choice and firm 1 can choose the extent of technological improvement, the adverse effect of an encompassing union on firm 1's innovation remains, which, in turn, may also make the consumers worse off under an encompassing union than under separate unions.[10] The consideration of a binary choice for firm 1's R&D decision helps us to prove our point in the simplest way.

### 2.3.4 The Effect of an Encompassing Labor Union on the Union Utilities

So far, we have done the analysis under separate and encompassing unions. However, it is important to see whether the workers have the incentive to form an encompassing union.

The utilities of the unions are

$$\pi_x^{nc} = \pi_y^{nc} = \frac{(n-1+t)^2}{9(n+1)[-2(1-t)+n(2-2t+t^2)]}$$

and

$$\pi_x^{c} = \pi_y^{c} = \frac{(n-1+t)^2}{8(n+1)[-2(1-t)+n(2-2t+t^2)]}$$

under separate unions and an encompassing union, respectively. It is immediate that if firm 1 either innovates or does not innovate irrespective of the unionization

[10] The intuition follows from Marjit and Mukherjee (2008), which shows in a different context that an input price reduction may either increase or decrease investment in innovation.

structure, the utilities of the unions are higher under an encompassing union than under separate unions.

Now consider the situation where firm 1 innovates under separate unions but it does not innovate under an encompassing union, which occurs for $s \in (2/(n+2), s^*)$ and $k^c < k < k^{nc}$. In this situation, the utilities of the unions are

$$\pi_x^{nc,rd} = \pi_y^{nc,rd} = \frac{(n-1+s)^2}{9(n+1)[-2(1-s)+n(2-2s+s^2)]}$$

and

$$\pi_x^{c,nrd} = \pi_y^{c,nrd} = \frac{n}{8(n+1)}$$

under separate unions and an encompassing union, respectively. Straightforward comparison shows that the union utilities are higher under an encompassing union than under separate unions in this situation.

Finally, consider the case where firm 1 innovates under an encompassing union but it does not innovate under separate unions, which occurs for $s \in (s^*, 1)$ and $k^{nc} < k < k^c$. In this situation, the utilities of the unions are

$$\pi_x^{nc,nrd} = \pi_y^{nc,nrd} = \frac{n}{9(n+1)}$$

and

$$\pi_x^{c,rd} = \pi_y^{c,rd} = \frac{(n-1+s)^2}{8(n+1)[-2(1-s)+n(2-2s+s^2)]}$$

under separate unions and an encompassing union, respectively. We get that the union utilities are higher under an encompassing union than under separate unions in this situation.

The following proposition is immediate from the above discussion.

PROPOSITION 5 *An encompassing labor union increases the utilities of the unions compared to separate labor unions, irrespective of its effect on innovation by firm 1.*

### 2.3.5 The Effect of an Encompassing Labor Union on Social Welfare

Finally, we want to see the effects of an encompassing union on social welfare, which is the sum of the union utilities, the net profits of the final-goods producers, and consumer surplus. If firm 1 either innovates or does not innovate irrespective of the unionization structure, welfare is higher under an encompassing union than under separate unions. Given the technology level, the lower wage under an encompassing union than under separate unions helps to create a higher welfare under the former than the latter unionization structure.

Now consider the case where firm 1 innovates only under separate unions, i.e., when $s \in (2/(n+2), s^*)$ and $k^c < k < k^{nc}$. We will see that an encompassing union may reduce welfare in this situation. Due to the complicated welfare expression, we will consider two numerical examples to show that whether an encompassing union reduces welfare in this situation depends on the product-market competition, given by $n$.

Assume that $n = 2$, $s \in (1/2, s^* = (-17 + \sqrt{561})/8)$, and $k^c < k < k^{nc}$. We get, in this situation, that welfare is higher under an encompassing labor union and no innovation than under separate labor unions and innovation, even if we consider maximum welfare under separate unions, which occurs at the cost of innovation $k^c$. However, if we consider that $n = 20$, $s \in (1/11, s^* = (-2717 + \sqrt{7816809})/80)$, and $k^c < k < k^{nc}$, we get that welfare in this situation is lower (higher) under an encompassing labor union and no innovation than under separate labor unions and innovation, for $s \in (1/11, 13/100)$ (for $s \in (13/100, (-2717 + \sqrt{7816809})/80)$) when we consider minimum welfare under separate unions, which occurs at $k^{nc}$.

The reason for the above result is as follows. If an encompassing union reduces innovation, it creates two opposing effects on welfare. On one hand, it tends to increase welfare by reducing wage. On the other hand, it tends to reduce welfare by reducing innovation. We have seen that more firms reduce the equilibrium wage. Since more firms create significantly lower wage under separate unions, further benefit from an encompassing union due to lower wage is not significant if the product-market competition is significant. Hence, if the number of firms and the technological improvement through innovation are large, the loss from an encompassing union due to lower innovation dominates the gain created by the encompassing union through lower wage, thus creating lower welfare under an encompassing labor union and no innovation than under separate labor unions and innovation.

However, if the product market is highly concentrated (i.e., $n$ is small), the gain from an encompassing union due to a lower wage can outweigh the negative effect of an encompassing union on innovation. In this situation, an encompassing union increases welfare even if it reduces innovation.

Finally, consider the case where firm 1 innovates only under an encompassing union, which occurs for $s \in (s^*, 1)$ and $k^{nc} < k < k^c$. We get in this situation that welfare is higher under an encompassing labor union and innovation than under separate labor unions and no innovation, even if we consider minimum welfare under an encompassing union, which occurs at the cost of innovation $k^c$. The positive effects of both lower wage and innovation following an encompassing union are responsible for this result.

The following result summarizes the above discussion.

PROPOSITION 6 *If an encompassing labor union reduces innovation compared to separate labor unions, social welfare may be lower under an encompassing labor union and no innovation than under separate labor unions and innovation if the product market is sufficiently competitive and the technological improvement through R&D is large.*

An encompassing union solves the complements problem, but it may reduce innovation by firm 1. As mentioned above, this trade-off is responsible for the above result. Since the benefit from an encompassing union due to a lower wage depends on the number of final-goods producers, product-market competition plays an important role for the above result.

As already mentioned, absence of innovation under an encompassing union is an artifact of the binary choice for firm 1's R&D decision, and we consider this binary choice to prove our point in the simplest way. If we consider a nonbinary choice for firm 1's R&D decision and firm 1 can choose the extent of technological improvement, firm 1 would innovate under an encompassing union, but the extent of technological improvement chosen by firm 1 could be lower under an encompassing union than under separate unions. Hence, even if we allow firm 1 to choose the extent of technological improvement, an encompassing union creates an adverse effect on innovation, which, in turn, may reduce social welfare compared to separate unions.

## *3 Discussion*

We now discuss the implications of some of our assumptions.

Like the related literature mentioned in the introduction, we have considered that labor is the only factor of production. This helped us to compare and contrast our results with those of the previous literature in the simplest way by showing the trade-off created by the complements problem and the raising-rivals'-cost effect on innovation. However, one can extend the analysis by incorporating nonlabor factors of production. The raising-rivals'-cost effect shown in our analysis would remain in this extended model with labor and nonlabor factors of production if the innovation were labor-saving, but that effect would not occur if innovation saved the nonlabor factors of production. If the production process involves both labor and nonlabor factors of production, a firm may also have a choice regarding the type of innovation (i.e., labor-saving and/or non-labor-saving), depending on the labor unionization structure. We leave this issue for future research.

To show the trade-off created by the complements problem and the raising-rivals'-cost effect on innovation in the simplest way, we assume that innovation allows the innovator to reduce the use of different types of labor in the same way. Hence, the innovation is not biased towards any type of worker and creates a neutral technological progress. This makes the analysis simple by creating symmetric behavior of different labor unions. However, it is needless to say that if innovation reduces the use of different types of labor differently, our main result showing the adverse effects of an encompassing labor union on innovation and welfare remains.

Finally, to show the adverse effects of an encompassing labor union on innovation and welfare, we considered that the unions have full bargaining power in determining wages. However, it is easy to understand that if the firms have bargaining power, it will reduce wages and would affect the equilibrium outputs and profits, yet the trade-off created by the complements problem and the raising-rivals'-cost effect remain. As long as the unions have the bargaining power, the complements problem creates a lower wage under an encompassing union than under separate unions. Hence, under no innovation, the profit of firm 1 is higher under an encompassing union than under separate unions, reducing firm 1's benefit from innovation under an encompassing union. On the other hand, under innovation, firm 1's profit is lower (higher) under an encompassing union than under separate unions if $s <(>) (n-1)/n$, reducing (increasing) firm 1's benefit from innovation under an encompassing union for $s <(>) (n-1)/n$. These are similar to the effects discussed after Proposition 3. Hence, even if there is bargaining between the firms and the unions, an encompassing union reduces (may increase) the incentive for innovation compared to separate unions if the technological improvement through innovation is large (small).

## 4 Conclusion

While firms use complementary workers in reality, the existing literature examining the effects of the labor unionization structure on innovation considered only substitutable workers and did not pay attention to complementary workers. This paper fills that gap in the literature.

We show that cooperation among the labor unions (or an encompassing labor union) of complementary workers may either increase or decrease a final-goods producer's incentive for innovation compared to noncooperation among the labor unions (or separate labor unions). Although cooperation among the unions solves the complements problem, it may have an adverse effect on the final-goods producer's technological improvement. We show that the adverse effect on the technological improvement may dominate the beneficial wage effect of cooperation among the unions, thus making the consumers and the society worse off. While cooperation among the unions makes the workers better off, it may not make all final-goods producers better off. Thus, our results provide new insights into the literature on labor unionization structure and innovation, and suggest that whether the workers are substitutes or complements is an important factor to consider.

## Appendix

### A.1 The Restriction on s to Ensure that All Final-Goods Producers Always Hire Workers

We show here that if the final-goods producers differ in their technologies, all final-goods producers hire workers under separate and encompassing unions if $2/(n+2) < s$. If the unions want to charge the wage in a way so that it is not profitable for all firms to hire workers at that wage, it is easy to understand that the unions can prevent the noninnovating firms from hiring workers but cannot prevent only the innovating firm from hiring workers. This happens because the outputs of the innovating firm are always positive whenever the outputs of the noninnovating firms are positive, and the unions cannot charge a wage that will induce only the noninnovating firms to hire workers.

*Separate Labor Unions.* First, consider the case of separate labor unions and the equilibrium with symmetric wages. If the unions want to provide workers only to the technologically superior final-goods producer (i.e., to firm 1, which innovates a new technology and creates technological differences between the final-goods producers), wages need to be such that it is not profitable for the technologically inferior noninnovating firms to hire workers. If the unions provide workers to firm 1 only, the demand for workers is $q_x = q_y = s[1 - s(w_x + w_y)]/2$ and the equilibrium wages are $w_x^{nc,m} = w_y^{nc,m} = 1/3s$. The outputs of the noninnovating firms are zero at these wages if $s \leq 4/5$. If $4/5 < s$, the equilibrium wages need to be $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ to prevent the noninnovating firms from hiring workers. Since $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ is the constrained wage, it is immediate that the equilibrium union utilities are lower from charging the wage $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ than from charging the wage $w_x^{nc,m} = w_y^{nc,m} = 1/3s$.

If the workers are hired only by firm 1 at the wages $w_x^{nc,m} = w_y^{nc,m} = 1/3s$, which can happen for $s \leq 4/5$, the equilibrium union utilities are $\pi_x^{nc,m} = \pi_y^{nc,m} = 1/18$. We get that, if $2/(n+2) < s$, then $\pi_x^{nc,m} = \pi_y^{nc,m} = 1/18$ are lower than $\pi_x^{nc} = \pi_y^{nc} = (n-1+s)^2/9(n+1)[-2(1-s)+n(2-2s+s^2)]$, which are the union utilities when all final-goods producers hire workers, as considered in the text. Since $\pi_x^{nc,m} = \pi_y^{nc,m} = 1/18$, i.e., the union utilities under the unconstrained wages $w_x^{nc,m} = w_y^{nc,m} = 1/3s$, are lower than the union utilities from providing workers to all final-goods producers, it is immediate that if $4/5 < s$ and the unions charge $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ to provide workers to only firm 1, the union utilities are lower from providing workers to only firm 1 than from providing workers to all final-goods producers. Hence, the separate unions provide workers to all final-goods producers for $2/(n+2) < s$, as considered in the text.

*An Encompassing Labor Union.* Now consider an encompassing labor union and the equilibrium with symmetric wages. If the union provides workers to firm 1 only, then the demand for workers is $q_x = q_y = s[1 - s(w_x + w_y)]/2$ and the equilibrium wages are $w_x^{c,m} = w_y^{c,m} = 1/4s$. The outputs of the noninnovating firms are zero at these wages if $s \leq 2/3$. If $2/3 < s$, the equilibrium wages need to be $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ to prevent the noninnovating firms from hiring workers. Since $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ is the constrained wage, it is immediate that the equilibrium union utility is lower from charging the wage $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ than from charging the wage $w_x^{c,m} = w_y^{c,m} = 1/4s$.

If the workers are hired only by firm 1 at the wages $w_x^{c,m} = w_y^{c,m} = 1/4s$, which can happen for $s \leq 2/3$, the equilibrium union utilities are $\pi_x^{c,m} = \pi_y^{c,m} = 1/16$. We get that, if $2/(n+2) < s$, then $\pi_x^{c,m} = \pi_y^{c,m} = 1/16$ are lower than $\pi_x^c = \pi_y^c =$

$(n-1+s)^2/8(n+1)[-2(1-s)+n(2-2s+s^2)]$, which are the union utilities when all final-goods producers hire workers, as considered in the text. Since $\pi_x^{c,m} = \pi_y^{c,m} = 1/16$, i.e., the union utilities under the unconstrained wages $w_x^{c,m} = w_y^{c,m} = 1/4s$, are lower than the union utilities from providing workers to all final-goods producers, it is immediate that if $2/3 < s$ and the unions charge $\widehat{w}_x^m = \widehat{w}_y^m = 1/2(2-s)$ to provide workers to firm 1 only, the union utilities are lower from providing workers to firm 1 only than from providing workers to all final-goods producers. Hence, an encompassing labor union provides workers to all final-goods producers for $2/(n+2) < s$, as considered in the text.

### *A.2 Proof of Lemma 1*

Under separate labor unions, firm 1's profit under innovation is

$$\pi_1^{nc,rd} = \left[\frac{-4(1-s)+2n^2(1-s)+n(2-2s+s^2)}{(n+1)[-6(1-s)+3n(2-2s+s^2)]}\right]^2 - k,$$

while its profit under no innovation is $\pi_1^{nc,nrd} = 1/9(n+1)^2$. Firm 1 innovates if $\pi_1^{nc,rd} > \pi_1^{nc,nrd}$, which gives the result. *Q.E.D.*

### *A.3 Proof of Lemma 2*

Under an encompassing labor union, firm 1's profit under innovation is

$$\pi_1^{c,rd} = \left[\frac{-3(1-s)+n^2(1-s)+n(2-2s+s^2)}{(n+1)[-4(1-s)+2n(2-2s+s^2)]}\right]^2 - k,$$

while its profit under no innovation is $\pi_1^{c,nrd} = 1/4(n+1)^2$. Firm 1 innovates if $\pi_1^{c,rd} > \pi_1^{c,nrd}$, which gives the result. *Q.E.D.*

### *A.4 Proof of Proposition 2*

If firm 1 innovates under both unionization structures, its profit is

$$\pi_1^{nc,rd} = \left[\frac{-4(1-s)+2n^2(1-s)+n(2-2s+s^2)}{(n+1)[-6(1-s)+3n(2-2s+s^2)]}\right]^2 - k$$

under separate labor unions and

$$\pi_1^{c,rd} = \left[\frac{-3(1-s)+n^2(1-s)+n(2-2s+s^2)}{(n+1)[-4(1-s)+2n(2-2s+s^2)]}\right]^2 - k$$

under an encompassing labor union. We get that $\pi_1^{nc} > (<)\ \pi_1^{c}$ for $s \in (2/(n+2), (n-1)/n)$ (for $s \in ((n-1)/n, 1)$).

If firm 1 innovates under separate and encompassing labor unions, the profit of firm $i$, $i = 2,\ldots,n$, is

$$\pi_i^{nc} = \left[\frac{-2(1-s^2)+n(2-4s+3s^2)}{(n+1)[-6(1-s)+3n(2-2s+s^2)]}\right]^2$$

under separate labor unions and

$$\pi_i^c = \left[\frac{-(2-s-s^2)+n(2-3s+2s^2)}{(n+1)[-4(1-s)+2n(2-2s+s^2)]}\right]^2$$

under an encompassing labor union. We get that $\pi_i^{nc} < \pi_i^c$ for $s \in (2/(n+2),1)$.

*Q.E.D.*

### A.5 *Proof of Proposition 3*

We obtain that $k^{nc} >(<)\ k^c$ for $s \in (2/(n+2),s^*)$ (for $s \in (s^*,1)$), where $s^* = (3+4n-7n^2+\sqrt{9-58n^2+49n^4})/4n$.

If $s \in (2/(n+2),s^*)$ and $k^c < k < k^{nc}$, firm 1 innovates only under separate unions. In this situation, an encompassing union reduces firm 1's incentive for innovation compared to separate unions. However, the unionization structure does not affect firm 1's incentive for innovation if either $k < k^c < k^{nc}$ (where firm 1 innovates irrespective of the unionization structures) or $k^c < k^{nc} < k$ (where firm 1 does not innovate irrespective of the unionization structure).

If $s \in (s^*,1)$ and $k^{nc} < k < k^c$, firm 1 innovates only under an encompassing union. In this situation, an encompassing union increases firm 1's incentive for innovation compared to separate unions. However, the unionization structure does not affect firm 1's incentive for innovation if either $k < k^{nc} < k^c$ (where firm 1 innovates irrespective of the unionization structure) or $k^{nc} < k^c < k$ (where firm 1 does not innovate irrespective of the unionization structure). *Q.E.D.*

### A.6 *Proof of Proposition 4*

If $s \in (2/(n+2),s^*)$ and $k \in (k^c,k^{nc})$, firm 1 innovates only under separate unions. The total outputs of the final-goods producers under separate labor unions and innovation by firm 1 and under an encompassing labor union and no innovation by firm 1 are

$$q^{nc,rd} = \frac{-2n(1-s)-2(1-s)^2+n^2(4-6s+3s^2)}{(n+1)[-6(1-s)+3n(2-2s+s^2)]}$$

and

$$q^{c,nrd} = \frac{n}{2(n+1)},$$

respectively. We get that $q^{nc,rd} >(<)\ q^{c,nrd}$ if $s \in (2/(n+2),s^{**})$ (if $s \in (s^{**},s^*)$), where

$$\frac{2}{n+2} < s^{**} = \frac{-4+n+3n^2-\sqrt{3(-n^2+n^4)}}{-4+3n^2} < s^*.$$

*Q.E.D.*

### *A.7 The Case of Multiple Innovators*

We show here that an encompassing labor union may reduce innovation compared to separate labor unions even if there are multiple innovators. Since the calculations are straightforward but cumbersome, we skip the mathematical details.

As in the text, we assume that there are $n$ firms in the industry. However, we now assume that all firms can innovate to improve labor productivities.

First, consider the case of separate labor unions. If $m-1$ firms have invested in innovation, the $m$th firm invests in innovation if its equilibrium net profit[11] from innovation (implying that $m$ firms invest in innovation) is higher than its equilibrium profit from no innovation (implying that $m-1$ firms invest in innovation) if

$$\begin{aligned} k &< \frac{(m^2(1-s)^2+n(-3-2n(1-s)+2s)+m(1-s)(3+n+s+ns))^2}{9(1+n)^2(n-m(1-s)(1+m-n(1-s)+s-ms))^2} \\ &\quad - \frac{(m(1-m+n)+2(m-1)(m-2(1+n))s-(m-1)(-4+m-3n)s^2)^2}{9(1+n)^2(m(1-m+n)+2(m-1)(-1+m-n)s+(m-1)(2-m+n)s^2)^2} \\ &\equiv k^{nc}(m). \end{aligned} \tag{A1}$$

We can find that a higher $m$ corresponds with a lower $k^{nc}(m)$, implying that if the cost of innovation increases, it reduces the number of firms undertaking innovation under separate unions.

Now consider the case of an encompassing labor union. If $m-1$ firms have invested in innovation, the $m$th firm invests in innovation if

$$\begin{aligned} k &< \frac{(m^2(1-s)^2+n(-2-n(1-s)+s)+m(1-s)(2+s+ns))^2}{4(1+n)^2(n-m(1-s)(1+m-n(1-s)+s-ms))^2} \\ &\quad - \frac{(m(1-m+n)+(m-1)(2m-3(1+n))s-(m-1)(-3+m-2n)s^2)^2}{4(1+n)^2(m(1-m+n)+2(m-1)(-1+m-n)s+(m-1)(2-m+n)s^2)^2} \\ &\equiv k^{c}(m). \end{aligned} \tag{A2}$$

We can find that a higher $m$ corresponds with a lower $k^c(m)$, implying that if the cost of innovation increases, it reduces the number of firms undertaking innovation under an encompassing union.

Evaluating (A1) and (A2) at $m=1$ and comparing them gives us Proposition 3. Now we consider other cases.

If $k < k^{nc}(n)$ and $k < k^c(n)$, all firms innovate under both unionization structures. However, we get that $k^{nc}(n) < k^c(n)$, suggesting that if the cost of innovation is such that all firms innovate under an encompassing union, the incentive for innovation is not lower under an encompassing union than under separate unions. This happens because the raising-rivals'-cost motive, as discussed after Proposition 3, does not work in this situation.

[11] When determining the equilibrium profits, we have considered the corresponding equilibrium wages.

Now consider the case where the costs innovation are not small enough to make innovation by all firms profitable under an encompassing union. Given (A1) and (A2), we cannot compare $k^{nc}(m)$ and $k^{c}(m)$ generally. Hence, we use numerical examples to show that the number of innovating firms may be lower under an encompassing union if all firms do not find innovation profitable under an encompassing union.

As an example, consider that $n = 5$ and $m = 2$. We get that

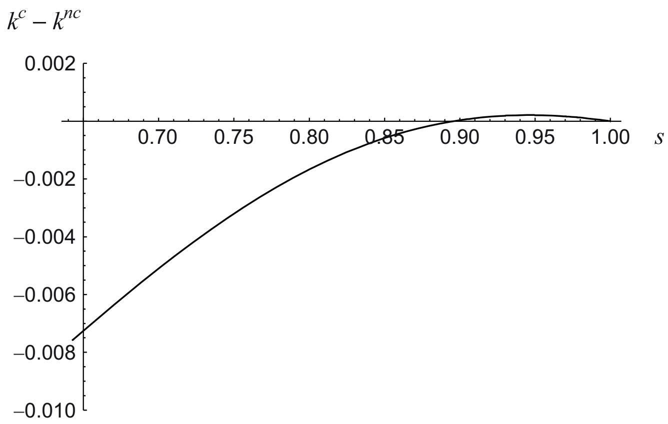
$$k < k^{nc}(2) = -\frac{2(-5184+20664s-37338s^2+38151s^3-21030s^4+2539s^5+4606s^6-3112s^7+704s^8)}{27(72-168s+205s^2-124s^3+40s^4)^2}$$

and

$$k < k^{c}(2) = \frac{3456-13752s+25053s^2-25893s^3+14550s^4-1981s^5-3073s^6+2152s^7-512s^8}{12(72-168s+205s^2-124s^3+40s^4)^2}.$$

We plot $k^{c}(m = 2,\ n = 5) - k^{nc}(m = 2,\ n = 5)$ in the figure and find that $k^{c}(m = 2,\ n = 5) < k^{nc}(m = 2,\ n = 5)$ for $0.6 < s < 0.9$ (approx.).[12]

*Figure*
$k^{c}(m = 2,\ n = 5) - k^{nc}(m = 2,\ n = 5)$




If $0.6 < s < 0.9$ (approx.) and $k^{c}(m = 2,\ n = 5) < k < k^{nc}(m = 2,\ n = 5)$, an encompassing labor union reduces the number of innovating firms compared to separate labor unions.

It is now easy to understand that if an encompassing labor union reduces the number of innovators compared to separate labor unions, it may reduce consumer surplus and welfare compared to separate labor unions even if it solves the complementary problem.

[12] If there are $k$ innovating and $n-k$ noninnovating firms, in an equilibrium with symmetric wages, the unions provide workers to all firms for $[2k+n(k-1)]/k(2+n) < s$. Hence, if $n = 5$ and $m = 2$, the unions provide workers to all firms for $0.6 < s$, and we restrict our attention to $0.6 < s$.

*References*

Aghadadashli, H., and C. Wey (2015), "Multiunion Bargaining: Tariff Plurality and Tariff Competition," *Journal of Institutional and Theoretical Economics (JITE)*, 171(4), 666–695.

Calabuig, V., and M. Gonzalez-Maestre (2002), "Union Structure and Incentives for Innovation," *European Journal of Political Economy*, 18(1), 177–192.

Calmfors, L., and J. Drifill (1988), "Bargaining Structure, Corporatism and Macroeconomic Performance," *Economic Policy*, 3(6), 13–61.

Carluccio, J., D. Fougère, and E. Gautier (2015), "Trade, Wages and Collective Bargaining: Evidence from France," *The Economic Journal*, 125(584), 803–837.

Flanagan, R. J. (1999), "Macroeconomic Performance and Collective Bargaining: An International Perspective," *Journal of Economic Literature*, 37(3), 1150–1175.

Gallini, N. T. (1992), "Patent Policy and Costly Imitation," *The RAND Journal of Economics*, 23(1), 52–63.

Gilbert, R. J., and M. L. Katz (2011), "Efficient Division of Profits from Complementary Innovations," *International Journal of Industrial Organization*, 29(4), 443–454.

Haucap, J., U. Pauly, and C. Wey (2001), "Collective Wage Setting when Wages Are Generally Binding: An Antitrust Perspective," *International Review of Law and Economics*, 21(3), 287–307.

— and C. Wey (2004), "Unionisation Structures and Innovation Incentives," *The Economic Journal*, 114(494), C149–C165.

Horn, H., and A. Wolinsky (1988), "Worker Substitutability and Patterns of Unionisation," *The Economic Journal*, 98(391), 484–497.

Kjellberg, A. (2014), "Union Density and Specialist/Professional Unions in Sweden," Mimeo, Lund University.

Lamarche, C. (2013), "Industry-Wide Work Rules and Productivity: Evidence from Argentine Union Contract Data," *IZA Journal of Labor & Development*, 2(Article 11), 1–25.

Layard, R., S. Nickell, and R. Jackman (1991), *Unemployment, Macroeconomic Performance and the Labour Market*, Oxford University Press, Oxford.

Manasakis, C., and E. Petrakis (2009), "Union Structure and Firms' Incentives for Cooperative R&D Investments," *Canadian Journal of Economics / Revue canadienne d'Economique*, 42(2), 665–693.

Marjit, S., and A. Mukherjee (2008), "International Outsourcing and R&D: Long-Run Implications for Consumers," *Review of International Economics*, 16(5), 1010–1022.

Mattoo, A., M. Olarreaga, and K. Saggi (2004), "Mode of Foreign Entry, Technology Transfer, and FDI Policy," *Journal of Development Economics*, 75(1), 95–111.

Moene, K. O., and M. Wallerstein (1997), "Pay Inequality," *Journal of Labor Economics*, 15(3), 403–430.

Mookherjee, D., and D. Ray (1991), "On the Competitive Pressure Created by the Diffusion of Innovations," *Journal of Economic Theory*, 54(1), 124–147.

Mukherjee, A. (2003), "Does Society Prefer Small Innovation?" *Economics Letters*, 78(3), 301–307.

— and E. Pennings (2004), "Imitation, Patent Protection, and Welfare," *Oxford Economic Papers*, 56(4), 715–733.

— and — (2011), "Unionization Structure, Licensing and Innovation," *International Journal of Industrial Organization*, 29(2), 232–241.

Ray Chowdhury, P. (1995), "Technological Asymmetry and Joint Product Development," *International Journal of Industrial Organization*, 13(1), 23–39.

Salop, S. C., and D. T. Scheffman (1983), "Raising Rivals' Costs," *The American Economic Review*, Papers and Proceedings, 73(2), 267–271.

— and — (1987), "Cost-Raising Strategies," *The Journal of Industrial Economics*, 36(1), 19–34.

Shapiro, C. (2000), "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting," in: A. B. Jaffe, J. Lerner, and S. Stern (eds.), *Innovation Policy and the Economy*, Vol. 1, The MIT Press, Cambridge (MA), pp. 119–150.

Upmann, T., and J. Müller (2014), "The Structure of Firm-Specific Labor Unions," *Journal of Institutional and Theoretical Economics (JITE)*, 170(2), 336–364.

Venn, D. (2009), "Legislation, Collective Bargaining and Enforcement: Updating the OECD Employment Protection Indicators," OECD Social, Employment and Migration Working Paper 89, OECD Directorate for Employment, Labour and Social Affairs, Paris.

Williamson, O. E. (1968), "Wage Rates as a Barrier to Entry: The Pennington Case in Perspective," *The Quarterly Journal of Economics*, 82(1), 85–116.

Tien-Der Han
School of Business and Economics
Loughborough University
Loughborough, Leicestershire
LE11 3TU
United Kingdom
t.han2@lboro.ac.uk

Arijit Mukherjee
Nottingham University
Business School
Jubilee Campus, Wollaton Road
Nottingham
NG8 1BB
United Kingdom
arijit.mukherjee@nottingham.ac.uk

# Market Power in Interactive Environmental and Energy Markets: The Case of Green Certificates

by

Eirik S. Amundsen and Gjermund Nese*

Received October 10, 2014; in revised form December 13, 2015;
accepted June 19, 2016

A market for tradable green certificates (TGCs) is strongly interwoven in the electricity market in that the producers of green electricity are also the suppliers of TGCs. Therefore, strategic interaction may result. We formulate an analytic equilibrium model for simultaneously functioning electricity and TGC markets, and focus on the role of market power (i.e., Stackelberg leadership). One result is that a certificate system faced with market power may collapse into a system of per-unit subsidies. Also, the model shows that TGCs may be an imprecise instrument for regulating the generation of green electricity. (JEL: C7, Q28, Q42, Q48)

## 1 Introduction

Along with the pursuance of targets for renewable energy production, many developed economies (e.g., Norway, Sweden, UK, U.S.) have implemented systems of tradable green certificates (TGCs).[1] In brief, a TGC market consists of sellers and buyers of TGCs. The sellers are the producers of electricity using renewable sources (green electricity). These producers are each issued a number of TGCs

* Eirik S. Amundsen (corresponding author): Department of Economics, University of Bergen, Norway, and Department of Food and Resource Economics, University of Copenhagen, Denmark; Gjermund Nese: Norwegian Competition Authority, Bergen, Norway. Financial support from the Nordic Energy Research (SNF-Projects No. 3156/3157), the Research Council of Norway: the Research Foundation/SAMSTEMT (SNF-Projects No. 3082/3135 and RCN-Project No. 138903/730 and No. 145734/730), and the Academia Agreement, UiB – Statoil, is gratefully acknowledged. The authors thank members of the Nordic Energy Research Program (NEMIEC) for valuable comments. The paper benefited from a presentation at various conferences, including presentations at Aberdeen University, University of Copenhagen, and the Norwegian School of Economics and Business Administration. The usual disclaimer applies.

[1] These are also referred to as Renewable Obligation Certificates (UK) or Renewable Portfolio Standards (U.S.).

*Journal of Institutional and Theoretical Economics* 173, 301–321 – ISSN 0932-4569
DOI: 10.1628/093245616X14690820714372 – © 2017 Mohr Siebeck

corresponding to the amount of electricity they feed into the network. The purchasers of certificates are consumers/distribution companies that are required by the government to hold a certain percentage of TGCs (the *percentage requirement*) corresponding to their total consumption (end-use deliveries) of electricity.[2] The TGCs are then seen as permits for consuming electricity. Accordingly, this system implies that the producers of green electricity receive both the wholesale price and the value of a TGC for each kWh fed into the electricity network. In this manner, the TGC system is supposed to stimulate new investments in green electricity generation.

One major implication of the TGC system is that the percentage requirement functions as a check on total electricity consumption, as the total number of TGCs available is constrained by the total capacity of renewable technologies.[3] For instance, a requirement of 20 percent implies that total consumption can be no larger than five times the electricity produced from renewable sources, unless the price of certificates tends to increase above an upper price bound specified by the regulatory authorities. This price bound then functions as a penalty that the consumers must pay if they do not fulfill the percentage requirement. Also, the TGC system may include a lower price bound, at which level the authorities guarantee to purchase any excess supply of TGCs. The percentage requirement is thus seen as a policy parameter affecting the relative scarcity of green electricity, and in this way regulating the capacity of green electricity generation.[4]

Up until now several aspects of the general functioning of TGC markets have been investigated. For example, problems relating to the TGC market as an instrument for inducing new capacity for green electricity production and problems related to the TGC markets acting in concert with electricity markets and $CO_2$ markets have been studied; see, e.g., Amundsen and Mortensen (2001, 2002), Bye (2003), Butler and Neuhoff (2008), Traber and Kemfert (2009), Fischer (2009), and Böhringer and Rosendahl (2010). Along with this, also the question of market power has been dealt with; see, e.g., Montero (2009) and Amundsen and Bergman (2012). However, yet another problematic feature related to market power needs to be investigated. Namely, a problem emerges in that electricity producers possessing market power take account of the joint functioning of the electricity market and the TGC market. As will be shown later, this may result in a collapse of the pricing mechanism of the TGC system, as the TGC price cannot be established between

[2] Italy is an exception in this respect: the Italian system is supposed to put the purchase obligation on the producers.

[3] However, in many countries, windmills constitute a significant part of the green production technology. The electricity production from windmills will typically vary significantly, giving rise to considerable annual variations in the total production of green electricity and therefore, also, of TGCs issued.

[4] The Swedish TGC system became effective on May 1, 2003; the Norwegian–Swedish TGC system, on January 1, 2012. The Swedish percentage requirement for 2012 was set at 17.9 percent; the Norwegian one, at 3 percent. In 2020 both percentages are to be set close to a maximum around 18 percent. Thereafter, the percentage requirements will fall towards zero in 2035.

the price bounds; i.e., if an equilibrium exists, it must be at either the stipulated upper or lower bound. Similar ideas on the exercise of market power through interactive markets are found in papers by Kolstad and Wolak (2003) and Chen and Hobbs (2005) concerning the joint functioning of the electricity and the $NO_x$ permit market.[5]

In a competitive setting, the TGC system may function as an ordinary market determining TGC prices somewhere intermediate between the upper and lower price bounds. The same may be true for a pure monopoly where the single producer generates both green and black electricity. However, this may no longer be so in the face of market power, where companies specialize in either green or black electricity. Hence, in this setting, if major electricity producers conjecture the effect on the TGC price of their production decisions in the electricity market and take account of this, then the TGC pricing mechanism may break down. By withholding electricity delivered to the wholesale market, the electricity producer can exercise market power by forcing the TGC price to either the upper or the lower price bound (either may be optimal for the producer) at its convenience. Basically, what is happening is that either excess demand for or excess supply of TGCs is created (leading to a price at the upper price bound or the lower price bound, respectively, with corresponding opposite effects on the wholesale prices). These results are valid irrespective of whether it is the producers of green or black electricity (electricity based on nonrenewable sources), or both, that possess market power. Thus, the TGC market may collapse altogether into a system of fixed TGC prices instead of endogenously determined intermediate prices.[6] In that case the TGC system may equally well be replaced by a plain subsidy scheme for green electricity, with presumably much lower transaction costs and more precise effects on green power capacity construction.

The problem of interactive power and TGC markets is then germane, since the TGC market in many countries is related directly to the electricity market, with identical suppliers and consumers to those of the electricity market. Thus, the effect on the TGC price of changing electricity production can hardly be ignored by a major electricity producer knowing that the end-user price of electricity for a large part is composed of the wholesale price and a fraction (e.g., 20 percent) of the TGC price. Hence, the revenue of a major producer of green electricity stems from both markets (i.e., from the electricity wholesale price and the TGC price), and the marginal reduction of green electricity production influences both markets (viz., through a reduction of the supply of electricity and a reduction of the supply of TGCs). Furthermore, a major producer of black electricity knows (even though not directly involved in TGC trade) that a marginal reduction of the electricity supply

[5] In particular, Chen and Hobbs (2005) show that endogenous treatment of the $NO_x$ and electricity market with conjectured price responses may have a substantial influence on $NO_x$ permit prices, and that the price of the permits thereby influences electricity generation.

[6] It is interesting to note that during the first year of the Swedish TGC system, TGCs frequently were traded at prices equal to the upper price bound; see STEM (2005).

will lead to a higher end-user electricity price, hence reduced total consumption, and hence a reduced demand for TGCs.

Market power in electricity generation is likely to exist in many economies. In Denmark, for example, the production of green electricity (notably from windmills) is very concentrated: in the Jutland–Fuen price area of Nord Pool only a single producer is currently active (Olsen, Amundsen, and Donslund, 2006). Hence, the possible malfunctioning of the pricing mechanism pointed to above should be given serious consideration in the discussions and development of alternative TGC systems.

In the following, we formulate an analytic equilibrium model for a TGC system and consider three main cases: (a) perfect competition in both the electricity market and the TGC market, (b) pure monopoly with joint generation of green and black electricity, and (c) a Stackelberg setting consisting of a leader specialized in the generation of black electricity and a follower specialized in the generation of green electricity. The next section of the paper presents the model. The subsequent sections present and analyze the equilibrium solutions for the cases listed above. The final section summarizes and concludes the paper.

## 2 *The Model*

The following model is designed to capture a setting of simultaneously functioning electricity and TGC markets. We will use the following symbols for the variables involved.

*Table*
Definition of the Variables

| Symbol | Definition |
|---|---|
| $p$ | consumer price of electricity |
| $s$ | price of TGCs |
| $\overline{s}$ | upper price bound of TGCs |
| $\underline{s}$ | lower price bound of TGCs |
| $q$ | wholesale price of electricity |
| $x$ | total consumption of electricity |
| $y$ | generation of black electricity |
| $z$ | generation of green electricity, equal to the number of TGCs issued |
| $\alpha$ | percentage requirement of green-electricity consumption |
| $g_d$ | demand for TGCs |

The inverse demand function is assumed given by[7] $p = p(x)$, with $\partial p(x)/\partial x < 0$. The intermediate or long-run industry cost function for black electricity is assumed given by[8] $c = c(y)$, with $\partial c(y)/\partial y = c'(y) > 0$ and $\partial^2 c(y)/\partial y^2 = c''(y) > 0$.

The rationale for choosing a marginal cost function that is increasing for this industry is that the expansion of output may drive up the price of $CO_2$ emission permits or $CO_2$ taxes to comply with national $CO_2$ emission constraints. The corresponding industry cost function for green electricity is assumed given by[9] $h(z)$, with $\partial h(z)/\partial z = h'(z) > 0$ and $\partial^2 h(z)/\partial z^2 = h''(z) > 0$.

The rationale for choosing a marginal cost function that is increasing for this industry is that good sites for generation technologies such as windmills may be in scarce supply, wherefore an expansion of green electricity generation implies increasing costs. On the other hand, learning-by-doing effects may well lead to reduced generation costs for green electricity over time (see Söderholm and Sundqvist, 2007), wherefore this assumption may not seem so realistic after all. However, the specified cost function may be seen as relevant for the medium term, as the full result of learning-by-doing effects will only materialize in the longer term.

## *3 Perfect Competition*

The electricity producers supply a common wholesale market within which a single wholesale electricity price is established. Retailers purchase electricity on the wholesale market, and TGCs on the TGC market. The electricity is distributed to end users, and a single end-user price is established. It is assumed that perfect competition prevails in all markets, with many producers of black and green electricity, many retailers, and many end users of electricity. Hence, all agents treat the various prices as given by the markets.

The producers act as if they jointly maximize[10]

$$\Pi(x) = q(z,y)y + [q(z,y) + s(z,y)]z - c(y) - h(z).$$

---

[7] The industry cost function is derived by "horizontal addition" of the individual cost functions; i.e., the cost of aggregate market supply is minimized. Using the industry cost function avoids using messy notation to describe individual decisions, and our prime interest is in the equilibrium market solution, not individual decisions. However, little detail is lost by this approach, as the individual first-order conditions for electricity producers correspond directly to those derived in the analysis.

[8] For a short-run version of the competitive model, see Amundsen and Mortensen (2001).

[9] In the short run with sunk-cost capital equipment, the marginal cost of green electricity may be close to zero; see, e.g., Amundsen and Mortensen (2001). In the intermediate or long-run situation considered here, however, capital costs are included.

[10] To simplify the presentation we suppress subscripts whenever confusion may be avoided.

The first-order condition for black electricity generation is

$$q = c'(y).$$

The first-order condition for green electricity generation is

$$q + s = h'(z).$$

We assume that a TGC is measured in the same units as electricity (i.e., MWh). With the given percentage requirement $\alpha$, retailers have to purchase a share $\alpha$ of a TGC for each unit of electricity (whether black or green) delivered to the end users. Thus, total demand for TGCs is given by $g_d = \alpha x$, whereas total supply of TGCs is equal to the amount of green electricity generated, $z$. For each unit of electricity (i.e., each MWh) purchased in the wholesale market and sold on to end users, retailers have to pay the wholesale price plus a share $\alpha$ of the TGC price. For simplicity, electricity distribution is assumed to be costless. With a large number of retailers, the equilibrium established in the market (i.e., the competitive equilibrium) must be characterized by

$$p(x) = q(z, y) + \alpha s(z, y),$$

where $x = z + y$.

### *3.1 Equilibrium under Perfect Competition*

The consumption of electricity, and its composition of black and green electricity in equilibrium (denoted by $^*$ and subscript $C$), vary according to whether the price of TGCs in equilibrium, $s_C^*$, is within the specified price interval (i.e., $\underline{s} < s_C^* < \overline{s}$) or on either the upper or the lower price bound. If the price of TGCs is within the interval, then the percentage requirement is fulfilled and the total consumption of electricity is given by $x_C^* = (z_C^*/\alpha)$ (the *allowable* consumption). If the price of TGCs is at the lower bound, i.e., $s_C^* = \underline{s}$, then the demand for TGCs is less than $z_C^*$, and the excess supply of TGCs is bought by the state. In this case the percentage requirement is more than fulfilled. If the price of TGCs in equilibrium is equal to the upper price bound $\overline{s}$, the demand for TGCs exceeds the maximum possible supply. In this case, the retailers/consumers are allowed to buy more black electricity if they pay a "fine" equal to $\overline{s}$ per unit of extra electricity consumption. The equilibrium conditions under perfect competition are

(1) $p(x_C^*) = q_C^* + \alpha s_C^*,$

(2) $x_C^* = y_C^* + z_C^* < \dfrac{z_C^*}{\alpha}$ or $x_C^* = y_C^* + z_C^* = \dfrac{z_C^*}{\alpha}$ or $x_C^* = y_C^* + z_C^* > \dfrac{z_C^*}{\alpha},$

(3) $q_C^* + s_C^* = h'(z_C^*),$

(4) $q_C^* = c'(y_C^*).$

From (2), if there is an excess supply of TGCs (i.e., $\alpha x_C^* < z_C^*$), then $s_C^* = \underline{s}$, and if there is an excess demand for TGCs (i.e., $\alpha x_C^* > z_C^*$), then $s_C^* = \overline{s}$. Otherwise – if TGC demand is equal to TGC supply (i.e., $\alpha x_C^* = z_C^*$) – then $\underline{s} < s_C^* < \overline{s}$. Basically, the quantity constraint implied by the percentage requirement drives a wedge equal to $\alpha s_C^*$ between the electricity price and the marginal cost of electricity generation. The system thus involves a transfer of consumer and producer surplus from black electricity generation to a subsidy of green electricity generation. Furthermore, by substituting (2), (3), and (4) into (1), we find that $p(x_C^*) = (1-\alpha)c'(y_C^*) + \alpha h'(z_C^*)$; i.e., in the competitive equilibrium, the consumer price of electricity is equal to a linear combination of the marginal costs of black and green electricity with the percentage requirement as a weight.

### *3.2 Analysis*

In the TGC system, the percentage requirement is perceived as a policy instrument affecting the level of green electricity in end-use consumption. Unlike price fixation (with quantity as an endogenous variable) or quantity fixation (with price as an endogenous variable), the percentage requirement fixes neither price nor quantity, and both variables are endogenously determined. The following proposition shows that in general it is erroneous to believe that a harsher percentage requirement necessarily will result in an increased capacity of green electricity generation. It does, however, lead to reduced generation of black electricity, and therefore – from (4) – a reduced wholesale price of electricity. As the effect on green electricity is indeterminate, the effect on total consumption and end consumer price is also indeterminate. Note that the TGC system specifies the share and not the absolute amount of green electricity in end-use consumption. Hence, if the effect on end-use consumption of electricity of an increase of $\alpha$ is negative, the percentage requirement may be fulfilled even if the generation of green electricity is reduced.[11]

PROPOSITION 1 *Under perfect competition in the electricity and the certificate markets, the percentage requirement, $\alpha$, has the following effects on the total electricity consumption $x_C^*$ and the green electricity generation $z_C^*$: (i) if $\underline{s} < s_C^* < \overline{s}$, then $dy_C^*/d\alpha < 0$ while* $\text{sign}(dz_C^*/d\alpha)$ *and* $\text{sign}(dx_C^*/d\alpha)$ *are indeterminate, and (ii) if $s_C^* = \overline{s}$ or $s_C^* = \underline{s}$, then $dz_C^*/d\alpha < 0$, $dy_C^*/d\alpha < 0$, $dx_C^*/d\alpha < 0$.*

As shown in Proposition 1, the effect on total electricity consumption of changing the percentage requirement is generally indeterminate (for the proof, see the appendix, section A.1). However, if the marginal cost of black electricity is constant (i.e., $c''(y) = 0$), we find that $dx_C^*/d\alpha < 0$. Thus, an increase of the percentage requirement will lead to a reduction of total electricity consumption. However, the effect on green electricity generation remains indeterminate. In addition, the effects depend on the level of the percentage requirement, $\alpha$. For example, if $\alpha = 0$, then $dz_C^*/d\alpha > 0$, whereas $dx_C^*/d\alpha$ is indeterminate.

[11] This is a generalization of results obtained in Amundsen and Mortensen (2001, 2002).

## *4 Monopoly*

As another reference case, in addition to the case of pure competition, we consider the case of a pure monopoly with a single producer generating both green and black electricity. We assume that the monopolist seeks to maximize the following objective function:

$$\Pi(z,y) = q(z,y)x + s(z,y)z - h(z) - c(y).$$

While recognizing that $q(z,y) = p(x) - \alpha s(z,y)$, we arrive at the following first-order conditions:

$$\frac{\partial \Pi}{\partial z} = \frac{\partial p}{\partial z}x - (\alpha x - z)\frac{\partial s}{\partial z} + q + s - h'(z) = 0$$

and

$$\frac{\partial \Pi}{\partial y} = \frac{\partial p}{\partial y}x - (\alpha x - z)\frac{\partial s}{\partial y} + q - c'(z) = 0.$$

Observe that the second term to the right of the first equality sign in each of these two expressions is always zero. If $\alpha x > z$, then $s = \overline{s}$, and if $\alpha x < z$, then $s = \underline{s}$. As $\overline{s}$ and $\underline{s}$ are constants, we have[12] $\partial\overline{s}/\partial z = \partial\overline{s}/\partial y = \partial\underline{s}/\partial z = \partial\underline{s}/\partial y = 0$. If $\alpha x = z$, then these expressions are also equal to zero.

The equilibrium conditions for a pure monopoly (key variables denoted by $^*$ and subscript $M$) are as follows:

(5) $p(x_M^*) = q_M^* + \alpha s_M^*$,

$$x_M^* = y_M^* + z_M^* < \frac{z_M^*}{\alpha} \quad \text{or} \quad x_M^* = y_M^* + z_M^* = \frac{z_M^*}{\alpha} \quad \text{or} \quad x_M^* = y_M^* + z_M^* > \frac{z_M^*}{\alpha},$$

(6) $\dfrac{\partial p(x_M^*)}{\partial x}x_M^* + q_M^* + s_M^* = h'(z_M^*)$,

(7) $\dfrac{\partial p(x_M^*)}{\partial x}x_M^* + q_M^* = c'(y_M^*)$.

The following proposition (for the proof, see the appendix, sections A.1 and A.2) states that an equilibrium TGC price may be established at an intermediate level

[12] Note that $\partial s/\partial y = 0$ and $\partial s/\partial z = 0$ at the TGC price bounds requires that the quantities of black and green electricity be sufficiently above or below the limits, leading to either an excessive or a deficient supply of green electricity, i.e., $\alpha x < z$ or $\alpha x > z$. If this is not the case, then a marginal increase in $z$ or $y$ will induce a jump either up or down between the price bounds. Hence, if there were a sufficiently small excess supply of TGCs, thus giving rise to $s^* = \underline{s}$, then a marginal reduction of $z$ would induce a jump of the TGC price from $\underline{s}$ to $\overline{s}$, and $\partial s/\partial y$ and $\partial s/\partial z$ would not be defined, as the marginal revenue would be discontinuous at this point. Throughout our analysis we will assume that the quantity of green electricity produced when $s^* = \underline{s}$ or $\overline{s}$ is such that a marginal change in the supply of either black or green electricity will not induce such a change from deficient to excessive supply of green electricity, or vice versa.

between the price limits when a single producer generates both green and black electricity. It also states that the effects of a change in the percentage requirement on electricity generation, in general, are all indeterminate under monopoly.

PROPOSITION 2 *Assume that a monopolist generates both green and black electricity. Then – in equilibrium – the TGC price may be established at (i) an intermediate level, i.e.,* $\underline{s} < s_M^* < \overline{s}$*, or at (ii) either of the price bounds, i.e.,* $s_M^* = \underline{s}$ *or* $s_M^* = \overline{s}$*. Furthermore, (iii) the effects of a change in the percentage requirement,* $\alpha$*, on total electricity consumption* $x_M^*$*, green electricity generation* $z_M^*$*, and black electricity generation* $y_M^*$ *are generally indeterminate, but equal to the effects under perfect competition if* $(dp(x_M^*)/dx) + (d^2p/dx^2)x_M^* < 0$ *(thus covering the case of a linear demand function).*

The reason why the existence of an intermediate TGC price under monopoly is stressed is that it runs counter to the cases where the producers possessing market power are specialized in either green or black electricity generation. This is considered in the next main section.

## 5 Stackelberg Game with Interactive Electricity and TGC Markets

In this section we consider the case of market power in the TGC market. Such a case may arise if one producer (or a few producers) has exclusive access to particularly good sites for green electricity generation (e.g., water power or wind power).[13] As the producer of green electricity is also the only supplier of TGCs, the producer thus possesses market power in the TGC market. We consider a case where the green producer only generates green electricity and not black electricity, while the black producer only generates black electricity and not green electricity. For the generation of black electricity one may for instance consider the case of a competitive fringe, a Nash–Cournot (NC) oligopoly, or a Stackelberg leadership model. The objective of this paper is to investigate the effects of market-power exertion, and in particular a setting where also a producer of black electricity recognizes that his actions in the electricity market have an effect in the TGC market. As this is the objective, it seems reasonable to assume that the producer of black electricity also recognizes that he can influence the green producer's decisions in the electricity market. In accordance with this, we shall therefore consider a Stackelberg game. We thus consider a standard Stackelberg model, where the producer of black electricity is the leader and the producer of green electricity is the follower.

Hence, in the following we take the interaction between the TGC market and the electricity market into account, i.e., we assume that the producers may take account of the effects on both markets of decisions made in the electricity market. The producer of green electricity is assumed to recognize that a reduction of green electricity also implies a reduction of the number of TGCs issued and consequently

[13] An example of this could be Dong Energy in the Jutland–Fuen price area, which has exclusive access to the wind power sites in the Nord Pool market.

that both markets may be affected by the reduction of green electricity generated. Likewise, a producer of black electricity possessing market power is assumed to recognize that a reduction of black electricity generation will affect both the electricity market and the TGC market through the percentage linkage of demand.[14]

We start by considering the optimal behavior of the green producer acting as a follower. In accordance with standard assumptions, we assume that the green producer follows an NC strategy and maximizes profit while considering the quantity of black electricity as given. Hence (for the moment suppressing the subscript indicating market form), the green producer is assumed to maximize the following objective function, with $y$ as given:

$$\Pi(z,y) = q(z,y)z + s(z,y)z - h(z).$$

The first-order condition of this maximization problem is given by

$$\frac{\partial \Pi}{\partial z} = \frac{\partial (q+s)}{\partial z} z + q + s - h'(z) = 0,$$

or, as $p(z+y) = q(z,y) + \alpha s(z,y)$,

$$\frac{\partial \Pi}{\partial z} = \left[ \frac{\partial p}{\partial z} + (1-\alpha) \frac{\partial s}{\partial z} \right] z + q + s - h'(z) = 0.$$

The first-order condition implicitly defines a reaction function $z = R(y)$ for the producer of green electricity.

Next, we consider the producer of black electricity acting as a Stackelberg leader. In accordance with standard assumptions, we assume that the leader maximizes profit while taking the reaction function of the follower, $R(y)$, as given. In doing this the leader will also consider the effects of his quantity decision on the TGC price, because the TGC price affects the wholesale price of electricity through the relation $q(z,y) = p(z+y) - \alpha s(z,y)$. Thus, the producer of black electricity is assumed to maximize the following objective function:

$$\Pi(y, R(y)) = q(R(y), y)y - c(y).$$

The first-order condition of this maximization problem is given by

$$\frac{\partial \Pi}{\partial y} = \left[ \frac{\partial q}{\partial z} \frac{dR}{dy} + \frac{\partial q}{\partial y} \right] y + q - c'(y) = 0,$$

or equivalently

$$\frac{\partial \Pi}{\partial y} = \left[ \frac{\partial p}{\partial x} \left( \frac{dR}{dy} + 1 \right) - \alpha \left( \frac{\partial s}{\partial z} \frac{dR}{dy} + \frac{\partial s}{\partial y} \right) \right] y + q - c'(y) = 0.$$

[14] This is different from a standard Cournot setting where the TGC price would have been treated as exogenous by both the producers of black and green electricity, i.e., neither of the producers would realize that their quantity decisions in the electricity market would affect the TGC price and thereby the resulting wholesale price of electricity through the interaction between the electricity and the certificate market.

As stated earlier, a marginal change in the generation of electricity, both black and green, may affect the wholesale price through both the electricity market and the TGC market. The effect through the electricity market stems from an ordinary effect on the consumer price, while the effect through the TGC market stems from a change induced by the demand/supply of TGCs. Hence, an increase in the generation of black electricity by one unit will, in equilibrium, imply an increased consumption of electricity by one unit and increased demand for certificates by $\alpha$ units, thus giving an upward pressure on the TGC price. Correspondingly, an increase in the generation of green electricity by one unit delivered to the market will also increase the consumption of electricity by one unit, and increase the demand for TGCs by $\alpha$ units, but also increase the number of TGCs by one unit. As the increase of TGC demand is only a fraction $\alpha$ of the increase of TGC supply, the net effect is a downward pressure on the TGC price. Here, it is important to stress that the demand for TGCs is a derived demand equal to a given percentage of the demand for electricity, i.e., a fixed linkage, meaning that the two demand functions are not independent.

### *5.1 Equilibrium under the Stackelberg Game*

The subscript $S$ is used to identify the case of market power in interactive electricity and power markets. We then have the following equilibrium conditions (key variables denoted by $^*$):[15]

$$p(x_S^*) = q_S^* + \alpha s_S^*, \tag{8}$$

$$x_S^* = y_S^* + z_S^* < \frac{z_S^*}{\alpha} \quad \text{or} \quad x_S^* = y_S^* + z_S^* = \frac{z_S^*}{\alpha} \quad \text{or} \quad x_S^* = y_S^* + z_S^* > \frac{z_S^*}{\alpha}, \tag{9}$$

$$\left[\frac{\partial p(x_S^*)}{\partial x} + (1-\alpha)\frac{\partial s(z_S^*, y_S^*)}{\partial z}\right] z_S^* + q_S^* + s_S^* = h'(z_S^*), \tag{10}$$

$$\left[\frac{\partial p(x_S^*)}{\partial x}\left(\frac{dR}{dy}+1\right) - \alpha\left(\frac{\partial s(z_S^*, y_S^*)}{\partial z}\frac{dR}{dy} + \frac{\partial s(z_S^*, y_S^*)}{\partial y}\right)\right] y_S^* + q_S^* = c'(y_S^*). \tag{11}$$

The possibility of affecting the TGC price depends, however, on whether the TGC price is either at the upper or lower price bound or between the upper and lower price bounds. If the TGC price is at either of the price bounds, the effect on the TGC price of a marginal change of the generation of black or green electricity (i.e., $\partial s/\partial y$ or $\partial s/\partial z$) is equal to zero, just as for a pure monopoly.[16] In these cases the wholesale price can only be affected through the electricity market (i.e., the ordinary price effect). If, however, the TGC price were between the price bounds, the producers could also influence the wholesale price through the TGC market. For this case, the marginal effect on the TGC price (i.e., $\partial s/\partial y$ or $\partial s/\partial z$) would

[15] For the cases of $s_S^* = \underline{s}$ and $s_S^* = \overline{s}$, we have $\partial s(x_S^*)/\partial z = \partial s(x_S^*)/\partial y = 0$. Thus, (10) and (11) are reduced to $(\partial p(x_S^*)/\partial x)z_S^* + q_S^* + s_S^* = h'(z_S^*)$ and $(\partial p(x_S^*)/\partial x)(dR/dy+1)y_S^* + q_S^* = c'(y_S^*)$, respectively.

[16] See footnote 12.

not be defined, as the marginal revenue is discontinuous at this point. Accordingly, the producer of black electricity could induce a reduction of demand for TGCs and thus create an excess supply of TGCs by marginally reducing the generation of black electricity. The consequence of this would be a drop of the TGC price to its lower bound and a corresponding upward jump of the wholesale price. The composition of the end-user electricity price would thus be changed to the benefit of the producer of black electricity. Likewise, the producer of green electricity might use the market power to reduce the generation of green electricity and TGCs marginally and thus create an excess demand for TGCs. This would lead to a jump of the TGC price to its upper bound and a corresponding reduction of the wholesale price.

The possibility of creating price jumps in the TGC market (due to the fixed linkage with the electricity market) implies that the green producer (the last mover) always has the option of profitably generating an upward jump of the TGC price, if an intermediate TGC price should emerge as a result of the quantity decision made by the Stackelberg leader. Hence, an intermediate TGC price cannot be an equilibrium price. This does not mean, however, that the equilibrium TGC price is always at the upper price bound. The optimal quantity response of the green producer may well result in an equilibrium TGC price at the lower price bound. This is foreseen by the Stackelberg leader through the knowledge of the reaction function of the follower. Hence, these relationships imply that the TGC market collapses in the sense that the TGC price will never be established at an intermediate level. These results are stated in Proposition 3 (for the proof, see the appendix, sections A.1 and A.2).[17]

PROPOSITION 3 *Assume that the producer of black electricity acts as a Stackelberg leader and the producer of green electricity acts as an NC-playing follower in interactive electricity and TGC markets. Then – in equilibrium – the TGC price will be equal to either the lower or the upper price bound and never lie between the two bounds, i.e.,* $s_S^* = \underline{s}$ *or* $s_S^* = \overline{s}$.

### 5.2 Illustrations

Figure 1 and Figure 2 illustrate the profit curves of the producer of green electricity under a setting of Stackelberg leadership. The figures are based on a simple numerical model satisfying the assumptions of the model; see appendix, section A.2. The profit curve of the green producer (the follower) is generated assuming that the quantity of black electricity is fixed at the profit-maximizing level of the producer of black electricity (the leader). The producer of green electricity generates an amount corresponding to the maximum profit, as foreseen by the producer of

[17] The results of assuming that the producers of black electricity behave like a competitive fringe or an NC oligopoly, while the producers of green electricity behave like an NC-playing group, are easily deducible from the analysis of the Stackelberg model. In fact, all cases give the same result: the TGC price will always be established at either the upper or the lower price bound. Proofs may be obtained from the authors upon request.

*Figure 1*
Profit of Green Electricity Generation, Stackelberg Equilibrium at Upper TGC Price

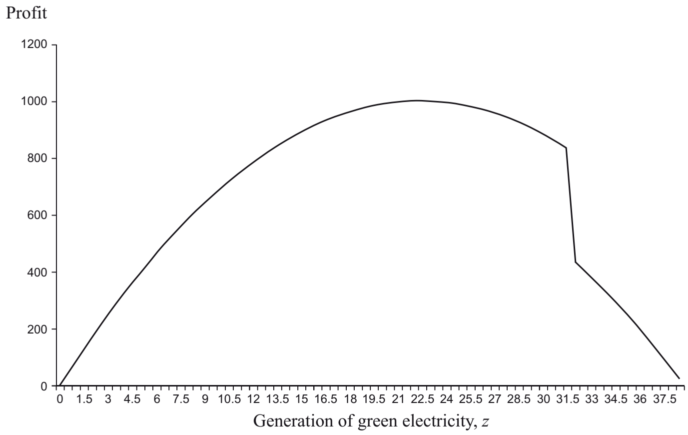
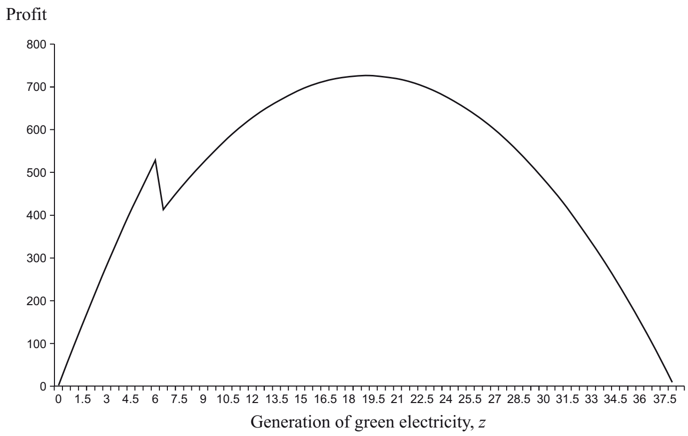



*Figure 2*
Profit of Green Electricity Generation, Stackelberg Equilibrium at Lower TGC Price




black electricity. These solutions maximize the profits of the producer of black electricity. Figure 1 shows an equilibrium at the upper TGC price bound, while Figure 2 shows an equilibrium at the lower TGC price bound.

In particular, we observe the discontinuity of the profit curves. Looking first at Figure 1, we note that the profit of the producer of green electricity drops (discontinuously) at a specific value of $z$. This is the quantity of green electricity at which total consumption of electricity is at the allowable consumption level (i.e., $z = \alpha x$). At this quantity level the TGC price jumps from the upper TGC price bound to the lower TGC price bound, resulting in a drop of profit. At lower production levels of green electricity, there is an excess demand for TGCs, i.e., the TGC price is at the upper TGC price bound. For higher production levels of green electricity, there is an excess supply of TGCs, wherefore the TGC price is at the lower bound. The case illustrates that the producer of black electricity does not necessarily induce a TGC price at the lower bound.

Figure 2 illustrates the profit curve of the green producer for increasing generation levels of green electricity when the production of black electricity is fixed at the optimal level for the producer of black electricity. This case has a lower percentage requirement, but is otherwise the same as the case of Figure 1. Again, the drop in profit takes place at a production level of green electricity for which $z = \alpha x$. At this point the TGC price drops from its upper bound to its lower bound. For higher production levels the profit starts to rise again and reaches a maximum at the level corresponding to the Stackelberg equilibrium.

## *6 Summary and Concluding Remarks*

This paper examines how an electricity market and a tradable green certificate (TGC) market function when it is recognized that such markets are strongly interlinked and the producers of electricity (green and black) take the interlinkage into account in their production decisions. The results of the paper are summarized in Propositions 1–3.

An essential element of a TGC system is that the number of TGCs issued functions as a check on total electricity consumption, in that the total amount of electricity consumed requires the possession of a number of TGCs corresponding to a given percentage of the electricity consumption. Hence, the total electricity consumption can be no larger than the number of TGCs sold divided by the percentage requirement (unless the TGC price tends to rise above a certain upper price bound as set by the regulatory authority).

The direct linkage between the two markets implies that a marginal change in the generation of electricity not only influences the price in the electricity market directly, but also indirectly through the effect on the TGC price. Recognizing this, the paper considers a setting of interactive functioning markets that goes beyond a traditional analysis of a producer operating in two markets where the producer considers the price of the other market as given, when deciding how much to produce in one of the markets. Potentially, this may give rise to problems with respect to the functioning of the markets. However, the paper shows that no such problems arise under perfect competition, as the producers take the prices as given anyway. Furthermore, and perhaps more interestingly, the same is the case under a pure monopoly, where the producer generates both green and black electricity.[18] Hence, for this case the effects of market power on market prices and quantities are shown to be as expected using standard economic models.

[18] The same would also be true for an oligopoly of identical producers, generating both green and black electricity.

However, a problem arises when there is a combination of market power and specialization of production, i.e., when some producers only produce green electricity and others only produce black electricity. The specialization implies that the producers may experience differing benefits of a high or a low TGC price. By taking the interaction between the two markets into consideration, the results are altered from the results that would emerge using standard Nash–Cournot assumptions under market power for the two markets; see Amundsen and Bergman (2012). In particular, the paper shows that market power will prevent the realization of a market-based TGC price within the specified price interval, i.e., the TGC price will be established at either the upper or the lower price bound. Thus, the TGC system will reduce to a system corresponding to direct subsidies financed through consumer/producer taxes. The paper shows this result for the case of Stackelberg leadership where the producer of black electricity acts as a leader and the producer of green electricity acts as a Nash–Cournot-playing follower. However, this result is not limited to the Stackelberg setting, but would be equally valid if black electricity were produced by a competitive fringe or by Nash–Cournot-playing oligopolists. The important feature is that a specialized producer with market power always can make the TGC price jump in a preferred direction by quantity adjustments in the electricity market.

In view of this result, the TGC price bounds will play an instrumental role in the subsidization of green electricity generation when the conditions mentioned above are fulfilled. In general, the basic rationale for adding price bounds in the TGC system is similar to the rationale for combining price bounds with a system of tradable permits in regulating the emission of a pollutant. For the latter case it has been shown that the combination of systems minimizes the expected loss due to either fixing a nonoptimal level of a Pigouvian tax or issuing a nonoptimal amount of permits in the face of missing information with respect to the true position of the marginal abatement cost curves; see, e.g., Weitzman (1974) and Roberts and Spence (1976). Hence, under perfect competition or a pure monopoly the price bounds in the TGC system are warranted. However, if the conditions are such that the TGC system reduces to a system where either the upper or the lower price bound is established, it would seem more natural to replace the system by a single unit subsidy of green electricity generation. This will also eliminate market-power exertion, as the producers cannot influence the size of the subsidy. Alternatively, regulating authorities could issue a tender for a given amount of green electricity. For the case of a fixed unit subsidy, uncertainty will pop up with respect to the quantity of green electricity generated (while the subsidy is fixed). For the case of a tender, uncertainty will pop up with respect to the size of the subsidy (while the quantity is fixed).

While the nonexistence of intermediate TGC prices in the face of market power is a clear-cut result, it will not necessarily emerge in existing (or planned) TGC markets. Several conditions must be satisfied for this to be the case: the possibility of exercising market power, producers separated into specialized production (either green or black electricity), and electricity and TGC markets that are simultaneously sensitive to price changes facing the end users of electricity. Considering the Norwegian–Swedish TGC market, the possibility of exercising market power is very small, due to the large number of producers of both kinds of electricity. Furthermore, most producers of electricity are not specialized, but have interests in the generation of both green and black electricity. Also, the retailing companies typically do not immediately pass on the TGC price to end users, but rather charge a fixed fee per unit electricity consumed to cover the purchase of TGCs at the end of the accounting period. Hence, from the point of view of end users it is as if the TGC price were fixed anyway. In this respect the result does not deviate significantly from that of a standard tax on electricity consumption.

Even though several conditions must be fulfilled to achieve the claimed result of nonexistence of intermediate TGC prices, that is not the same as saying that such conditions never will be fulfilled. Generally, many electricity companies have considerable market power in their own markets; e.g., the Danish company Dong in the Jutland price area of Nord Pool, or, indeed, the many so-called national champions like EDF. Furthermore, new large specialized producers of green electricity are entering the scene and may have an interest in keeping a high TGC price; e.g., Statoil has no fossil generation of electricity but ventures into offshore wind power generation. Hence, in considering the introduction of a TGC system, one may be well advised to reconsider the simpler system of a feed-in tariff. The TGC system may boil down to a fixed remuneration to green power generation anyway, and at a presumably much higher cost to society of running an auction and control system for TGCs. Put differently, it is unlikely that it will be cost-efficient to introduce a TGC system that ultimately functions like an ordinary subsidy scheme.

Another problematic feature of the TGC system, also shown in this paper, is that the percentage requirement in itself is not a precise policy instrument that determines the capacity level of green electricity generation (in contrast to how it is commonly perceived). An increase of the percentage requirement may, in fact, lead to a reduction of remuneration from investing in new capacity for green electricity (though it will affect the composition of black and green electricity generation in the preferred direction).[19] Along with other potential problems (e.g., compatibility with $CO_2$-emission permit systems and the strong price volatility of TGCs based on wind power), the problems revealed in this paper clearly call for caution in the design and implementation of TGC systems.

[19] Also, the TGC systems currently running or planned do not distinguish between the degrees of "greenness" or "blackness." This is contrary to what a system of permits for $CO_2$ emissions does.

## Appendix

### A.1 Proofs of the Propositions

PROOF OF PROPOSITION 1 (i) For $\underline{s} < s_c^* < \overline{s}$, inserting (3) and (4) into (1) yields the electricity price as a linear combination of marginal costs of the two groups of generation technologies in equilibrium, i.e., $p(x_c^*) = (1-\alpha)c'(y_c^*) + \alpha h'(z_c^*)$. Take the implicit derivatives of this expression with respect to $\alpha$ and arrive at

$$\frac{dz_c^*}{d\alpha} = \frac{\alpha s_c^* + x_c^*[\partial p/\partial x - (1-\alpha)c''(y_c^*)]}{D},$$
$$\frac{dy_c^*}{d\alpha} = \frac{(1-\alpha)s_c^* + x_c^*[\alpha h''(z_c^*) - \partial p/\partial x]}{D},$$

and

$$\frac{dx_c^*}{d\alpha} = \frac{s_c^* + x_c^*[\alpha h''(z_c^*) - (1-\alpha)c''(y_c^*)]}{D},$$

with $D = [\partial p/\partial x - (1-\alpha)^2 c''(y_c^*) - \alpha^2 h''(z_c^*)] < 0$. Inspection of signs verifies the claims of the proposition.

(ii) For $s_c^* = \underline{s}$ or $s_c^* = \overline{s}$, insert (4) into (3). Take the implicit derivative with respect to $\alpha$ and get $h''(z_c^*)(dz_c^*/d\alpha) = c''(y_c^*)(dy_c^*/d\alpha)$. As marginal costs are assumed increasing, it follows that $\text{sign}(dz_c^*/d\alpha) = \text{sign}(dy_c^*/d\alpha) = \text{sign}(dx_c^*/d\alpha)$. The last equality follows because $dx_c^*/d\alpha = dz_c^*/d\alpha + dy_c^*/d\alpha$. But the signs cannot be nonnegative. To see this insert (4) into (1) and take the implicit derivative with respect to $\alpha$ to obtain $(\partial p/\partial x)(dx_c^*/d\alpha) = c''(y_c^*)(dy_c^*/d\alpha) + \tilde{s}$, where $\tilde{s} = \overline{s}$ or $\underline{s}$. As $\partial p/\partial x < 0$, we must have $dx_c^*/d\alpha < 0$ for this equation to hold. Hence, $\text{sign}(dz_c^*/d\alpha) = \text{sign}(dy_c^*/d\alpha) = \text{sign}(dx_c^*/d\alpha) < 0$ for this case.

PROOF OF PROPOSITION 2 (i) To show that there may be an interior TGC price, $\underline{s} < s_M^* < \overline{s}$, it suffices to give an example. This is provided in appendix section A.2. The essential reason for the existence of such interior prices is that the monopolist is indifferent with respect to securing the high, the low, or some intermediate TGC price (and correspondingly for the wholesale price) for the case where the optimal solution satisfies $\hat{x} = \hat{y} + \hat{z} = \hat{z}/\alpha$. To see this, consider the profit function for the monopolist, $\Pi(\hat{z}, \hat{y}) = q\hat{x} + s\hat{z} - c(\hat{y}) - h(\hat{z})$. This may be rewritten $\Pi(\hat{z}, \hat{y}) = p\hat{x} + (\hat{z} - \alpha\hat{x})s - c(\hat{y}) - h(\hat{z})$. However, as $\hat{x} = \hat{z}/\alpha$, the profit function reduces to $\Pi(\hat{z}, \hat{y}) = p\hat{x} - c(\hat{y}) - h(\hat{z})$. Hence, the value of $s$ does not matter. Intuitively, a larger TGC price is exactly offset by a smaller wholesale price for this case.

(ii) To show that there may be a TGC price at either the upper or the lower price bound, it suffices to give examples satisfying the assumptions of the model. See appendix section A.2.

(iii) For $\underline{s} < s_c^* < \overline{s}$, inserting (6) and (7) into (5) yields the marginal revenue as a linear combination of marginal costs of the two groups of generation technologies in equilibrium, i.e., $p(x_M^*) + (dp/dx)x_M^* = (1-\alpha)c'(y_M^*) + \alpha h'(z_M^*)$. Take the implicit derivatives of this expression with respect to $\alpha$ and arrive at

$$dz_M^*/d\alpha = (\alpha s_M^* + x_M^*[(\partial p/\partial x) + (\partial^2 p/\partial x^2)x_M^* - (1-\alpha)c''(y_M^*)])/D,$$
$$dy_M^*/d\alpha = ((1-\alpha)s_M^* + x_M^*[\alpha h''(z_M^*) - (\partial p/\partial x) - (\partial^2 p/\partial x^2)x_M^*])/D,$$

and

$$dx_M^*/d\alpha = (s_M^* + x_M^*[\alpha h''(z_M^*) - (1-\alpha)c''(y_M^*)])/D,$$

with $D = [2(\partial p/\partial x) + (d^2 p/dx^2)x_M^* - (1-\alpha)^2 c''(y_M^*) - \alpha^2 h''(z_M^*)]$. Inspection of signs verifies the claims of the proposition. However, if $(\partial p/\partial x) + (d^2 p/dx^2)x_M^* < 0$ (which covers the case of linear demand), the signs are as in the competitive case. For $s_M^* = \underline{s}$ or $s_M^* = \overline{s}$, insert (7) into (6). Take the implicit derivative with respect to $\alpha$ and get $h''(z_M^*)(dz_M^*/d\alpha) = c''(y_M^*)(dy_M^*/d\alpha)$. As marginal costs are assumed increasing, it follows that $\operatorname{sign}(dz_M^*/d\alpha) = \operatorname{sign}(dy_M^*/d\alpha) = \operatorname{sign}(dx_M^*/d\alpha)$. The last equality follows because $dx_M^*/d\alpha = dz_M^*/d\alpha + dy_M^*/d\alpha$. To verify the signs, insert (7) into (5) and take the implicit derivative with respect to $\alpha$ to obtain $((\partial p/\partial x) + (d^2 p/dx^2)x_M^*)(dx_M^*/d\alpha) = c''(y_M^*)(dy_M^*/d\alpha) + \tilde{s}$, where $\tilde{s} = \overline{s}$ or $\underline{s}$. If $(\partial p/\partial x) + (d^2 p/dx^2)x_M^* < 0$, we must have $dx_c^M/d\alpha < 0$ for this equation to hold. Hence, $\operatorname{sign}(dz_M^*/d\alpha) = \operatorname{sign}(dy_M^*/d\alpha) = \operatorname{sign}(dx_M^*/d\alpha) < 0$ for this case (which covers the case of linear demand). If, however, $(\partial p/\partial x) + (d^2 p/dx^2)x_M^* > 0$, then the signs are generally indeterminate.

PROOF OF PROPOSITION 3 (i) Consider the quantity decision of the producer of green electricity (the follower). To show that we cannot have $\underline{s} < s_S^* < \overline{s}$, assume $\hat{z}$ is a solution satisfying the first-order conditions for the producers of green electricity and that $\overline{y} + \hat{z} = \hat{z}/\alpha$, which is a necessary condition for an intermediate TGC price. The symbol $\overline{y}$ denotes the quantity decision made by the producer of black electricity (the leader). Clearly, if $z < \hat{z}$, then $s_S^* = \overline{s}$, due to excess demand for TGCs (i.e., $z < \alpha(\overline{y} + z)$), and if $z > \hat{z}$, then $s_S^* = \underline{s}$, due to excess supply of TGCs. The total revenue function of the green producer is equal to $(q+s)z$, and the marginal revenue function is equal to $g(z,\overline{y}) = (\partial(q+s)/\partial z)z + q + s$. Observe that $g(z,\overline{y}) = (\partial p/\partial x)z + q + s$ for $z \neq \hat{z}$, as $\partial s/\partial y = 0$ for such values. Clearly, $g(z,\overline{y})$ is discontinuous at $\hat{z}$, as

$$\lim_{z\to\hat{z}^-} g(z,\overline{y}) = \frac{\partial p}{\partial x}\hat{z} + \hat{q}^- + \overline{s} \quad \text{and} \quad \lim_{z\to\hat{z}+} g(z,\overline{y}) = \frac{\partial p}{\partial x}\hat{z} + \hat{q}^+ + \underline{s},$$

where

$$\hat{q}^- = \lim_{z\to\hat{z}^-} q = p(\overline{y} + \hat{z}) - \alpha\overline{s} \quad \text{and} \quad \hat{q}^+ = \lim_{z\to\hat{z}+} q = p(\overline{y} + \hat{z}) - \alpha\underline{s}.$$

Inserting these expressions into the profit function $\Pi(\hat{z},\overline{y}) = (q+s)\hat{z} - h(\hat{z})$, we see that $(p(\overline{y}+\hat{z}) - \alpha\overline{s} + \overline{s})\hat{z} - h(\hat{z}) > (p(\overline{y}+\hat{z}) - \alpha\underline{s} + \underline{s})\hat{z} - h(\hat{z})$. Hence, profit maximization will lead the producer of green electricity to secure $\hat{q}^-$ (by an infinitesimal quantity reduction of green electricity), implying the corner solution $s_S^* = \overline{s}$.

(ii) To show that the equilibrium TGC price may be at either of the price bounds, it suffices to give examples satisfying the assumptions of the model. See appendix section A.2, and Figure 1 and Figure 2.

### *A.2 Numerical Model*

In this subsection we present a simple numerical model satisfying the assumptions of the analytical model. The model is used to give examples of the existence of some of the results referred to in the propositions in this article. It is also applied for the calculations of the numerical examples illustrated by the figures of the paper.

We assume the following functions: The inverse demand function is given by

$$p(x) = a - bx,$$

where $a$ and $b$ are strictly positive constants, $a, b > 0$. This gives $p'(x) = -b < 0$. The cost function for black electricity is

$$c(y) = 0.5y^2,$$

with $c'(y) = y > 0$ and $c''(y) = 1$. The cost function for green electricity is

$$h(z) = (k/2)z^2 + gz,$$

with $h'(z) = kz + g > 0$ and $h''(z) = k$, where $k$ and $g$ are strictly positive constants, $k, g > 0$.

Under these assumptions the optimal solutions of the various markets forms are as follows:

Perfect competition:

$$z_C^* = \frac{a - \alpha s - (1+b)(g-s)}{b + bk + k}, \quad y_C^* = \frac{k(a - \alpha s) + b(g-s)}{b + bk + k}.$$

Monopoly:

$$z_M^* = \frac{a - \alpha s - (1+2b)(g-s)}{2b + 2bk + k}, \quad y_M^* = \frac{k(a - \alpha s) + 2b(g-s)}{2b + 2bk + k}.$$

Calculations show that maximum profit is attained at a TGC price equal to $s_M^* = 76.2$ when $\alpha = 0.65$, $a = 100$, $b = 1$, $k = 2$, $g = 5$, $\overline{s} = 80$, $\underline{s} = 50$. This is seen to be an interior solution for the TGC price, and is thus in accordance with the claim of Proposition 2, that such solutions exist under monopoly. Furthermore, if $\alpha = 0.4$, $a = 100$, $b = 1$, $k = 2$, $g = 5$, $\overline{s} = 25$, $\underline{s} = 10$, maximum profit is attained at the lower bound of the TGC price, $s_M^* = \underline{s} = 10$. Also, if $\alpha = 0.6$, $a = 100$, $b = 1$, $k = 2$, $g = 5$, $\overline{s} = 25$, $\underline{s} = 100$, maximum profit is attained at the upper bound of the TGC price, $s_M^* = \overline{s} = 25$.

Stackelberg solution:

$$z_S^* = \frac{(2+b)(a-\alpha s)-(2+3b)(g-s)}{(1+b)(4b+3k)-k},$$
$$y_S^* = \frac{2[(b+k)(a-\alpha s)+b(g-s)]}{(1+b)(4b+3k)-k},$$

with $R(y) = (a-\alpha s-3(g-s)+2y)/(4b+3k)$.

Assuming $\alpha = 0.6$, $a = 100$, $b = 1$, $k = 2$, $g = 5$, $\overline{s} = 40$, $\underline{s} = 10$, calculations show that $z_S^* = 22.4$, $y_S^* = 21.4$, $x_S^* = 43.9$. Because $z_S^* < \alpha x_S^*$, we have $s_S^* = \overline{s} = 40$, i.e., the upper TGC price bound, as illustrated in Figure 1.

Assuming $\alpha = 0.2$, $a = 100$, $b = 1$, $k = 2$, $g = 5$, $\overline{s} = 40$, $\underline{s} = 10$, calculations show that $z_S^* = 25.0$, $y_S^* = 26.8$, $x_S^* = 51.8$. Because $z_S^* > \alpha x_S^*$, we have $s_S^* = \underline{s} = 10$, i.e., the lower TGC price bound, as illustrated in Figure 2.

## *References*


Amundsen, E. S., and L. Bergman (2012), "Green Certificates and Market Power on the Nordic Power Market," *The Energy Journal*, 33(2), 101–117.

— and J. B. Mortensen (2001), "The Danish Green Certificate System: Some Simple Analytical Results," *Energy Economics*, 23(5), 489–509.

— and — (2002), "Erratum to 'The Danish Green Certificate System: Some Simple Analytical Results' [Energy Economics 23 (2001) 489–509]," *Energy Economics*, 24(5), 523–524.

Böhringer, C., and K. E. Rosendahl (2010), "Green Promotes the Dirtiest: On the Interaction between Black and Green Quotas in Energy Markets," *Journal of Regulatory Economics*, 37(3), 316–325.

Butler, L., and K. Neuhoff (2008), "Comparison of Feed-In Tariff, Quota and Auction Mechanisms to Support Wind Power Development," *Renewable Energy*, 33(8), 1854–1867.

Bye, T. (2003), "On the Price and Volume Effects from Green Certificates in the Energy Market," Discussion Paper 351, Research Department, Statistics Norway, Oslo.

Chen, Y.-H., and B. Hobbs (2005), "An Oligopolistic Power Market Model with Tradable $NO_x$ Permits," *IEEE Transactions on Power Systems*, 20(1), 119–129.

Fischer, C. (2009), "Renewable Portfolio Standards: When do they Lower Energy Prices?" *The Energy Journal*, 31(1), 101–119.

Kolstad, J. T., and F. A. Wolak (2003), "Using Environmental Emissions Permit Prices to Raise Electricity Prices: Evidence from the California Electricity Market," Draft prepared for the 8th Annual POWER Conference on Energy Research and Policy, Berkeley, March 14.

Montero, J.-P. (2009), "Market Power in Pollution Permit Markets," *The Energy Journal*, 30(Special Issue 2), 115–142.

Olsen, O. J., E. S. Amundsen, and B. Donslund (2006), "How to Play the Game as the Bridge between Two European Power Markets—The Case of Western Denmark," *Energy Policy*, 34(17), 3293–3304.

Roberts, M. J., and M. Spence (1976), "Effluent Charges and Licenses under Uncertainty," *Journal of Public Economics*, 5(3–4), 193–208.

Söderholm, P., and T. Sundqvist (2007), "Empirical Challenges in the Use of Learning Curves for Assessing the Economic Prospects of Renewable Energy Technologies," *Renewable Energy*, 32(15), 2559–2578.

STEM (2005), *Översyn av elcertifikatsystemet Etapp 1* (includes English summary), ER 2005:08, Energimyndigheten (Swedish Energy Agency), Eskilstuna.

Traber, T., and C. Kemfert (2009), "Impacts of the German Support for Renewable Energy on Electricity Prices, Emissions, and Firms," *The Energy Journal*, 30(3), 155–178.

Weitzman, M. L. (1974), "Prices *vs.* Quantities," *The Review of Economic Studies*, 41(4), 477–479.

Eirik S. Amundsen
Department of Economics
University of Bergen
Box 7802
5020 Bergen
Norway
eirik.amundsen@econ.uib.no

Gjermund Nese
Norwegian Competition Authority
Box 439, sentrum
5805 Bergen
Norway
post@konkurransetilsynet.no

# Works Councils and Collective Bargaining in Germany: A Simple Theoretical Extension to Reconcile Conflicting Empirical Findings

by

Uwe Jirjahn*

Received July 9, 2015; in revised form May 11, 2016;
accepted May 20, 2016

A model by Hübler and Jirjahn (2003) suggests that redistribution activities of works councils are more limited in establishments covered by collective bargaining. The model predicts that works councils have stronger productivity effects and weaker wage effects in covered than in uncovered establishments. While empirical studies provide supporting evidence for the predicted productivity effects, the results on the wage effects are mixed. This article extends the model to reconcile the empirical findings. It takes into account that collective-bargaining coverage not only limits redistribution activities but also strengthens the effectiveness of performance-enhancing work practices negotiated between employers and works councils. (JEL: J24, J31, J51, J53)

## *1 Introduction*

German works councils provide a highly developed mechanism for establishment-level codetermination. While works councils play an important role in corporate governance in many West European countries, a unique feature of German works councils is that they have acquired quite extensive powers (Jenkins and Blyton, 2008; Rogers and Streeck (eds.), 1995). These powers have even been strengthened by the 2001 amendment of the Works Constitution Act (WCA), the law that governs the works council system. Works councils have also received attention outside Europe. In the U.S., a discussion on mandating German-style works councils has been spurred by a decline in union density and the growth of a "representation gap" (Freeman and Rogers, 1999). Furthermore, economists have shown a strong interest in works councils. This is documented by an increasing number of studies on the economic consequences of codetermination (Jirjahn, 2011).

* University of Trier, Germany.

*Journal of Institutional and Theoretical Economics* 173, 322–346 – ISSN 0932-4569
DOI: 10.1628/093245616X14739240653190 

A more comprehensive understanding of the functioning of works councils requires that other parameters of the industrial relations system be taken into account. Hübler and Jirjahn (2003) have developed a model that analyzes the interaction of works councils with collective-bargaining coverage. The model captures the idea that works councils have two faces. On the one hand, works councils play a trust-building role and thus provide a mechanism for negotiating productivity-enhancing work practices that otherwise could not be implemented. On the other hand, works councils can use their codetermination rights for redistribution activities. They push through higher wages by threatening to hinder decisions. Hübler and Jirjahn argue that coverage by a collective-bargaining agreement influences whether the generation or the redistribution of rents dominates. When distributional conflicts are moderated by unions and employers' associations outside the establishments, councils have less opportunity for redistribution, so that they are more likely to be engaged in rent-generating activities.

Hübler and Jirjahn's empirical results conform to this hypothesis. Works councils are associated with increased productivity in covered but not in uncovered establishments. By contrast, works councils have a less strong wage effect in covered than in uncovered establishments. A series of empirical follow-up studies have reexamined the interaction of works councils and collective bargaining. As to the interaction effect on productivity, most of those studies provide a remarkably clear pattern of results. They corroborate that collective-bargaining coverage fosters positive productivity effects of works councils. Yet, as to the interaction effect on wages, the findings are very mixed. While some studies confirm a weaker wage effect of works councils in covered establishments, other studies obtain the opposite result.

This paper extends Hübler and Jirjahn's model in order to reconcile the conflicting empirical findings. The aim is not to provide a sophisticated theoretical advancement or to develop theory for its own sake. The extended model rather aims at providing a better explanation of the available evidence and deriving implications for future empirical research. The basic question to be addressed is: Why have empirical studies found very mixed results with respect to the interaction effect on wages but not with respect to the interaction effect on productivity?

The extension takes into account that collective bargaining can have two moderating influences. First, as in Hübler and Jirjahn's model, collective-bargaining coverage limits the opportunities of a works council to engage in redistribution activities. Second, collective-bargaining coverage increases the effectiveness of the work practices negotiated between works council and employer. Adding the second moderating influence to that considered by Hübler and Jirjahn has crucial implications. As to the productivity effect of works councils, the two moderating influences work in the same direction. They strengthen the positive productivity effect of works councils. However, as to the wage effect of works councils the two influences work in opposite directions. Limiting the opportunities for redistribution lowers the wage effect of works councils. Improving the effectiveness of negotiated work practices increases the wage effect, as more productive work practices imply a higher rent that can be shared by the employer and the workforce. Thus, the extended theoretical model predicts an unambiguous interaction effect of works councils and collective-bargaining coverage on productivity and an ambiguous interaction effect on wages.

The model helps explain why empirical studies have produced mixed results on the wage effects of works councils in covered and uncovered establishments. The studies often differ in the industries or in the time period considered in the analysis. To the extent the relative weights of the two moderating influences of collective-bargaining coverage vary across industries or have changed over time, it makes sense that some studies find a weaker and others a stronger wage effect of works councils in covered establishments.

The extended model can also explain why empirical studies have produced mixed results only as to the wage effects and not as to the productivity effects of works councils in covered and uncovered establishments. The two moderating influences of collective-bargaining coverage strengthen the productive role of works councils. Hence, it does not matter whether the first or the second influence dominates.

The rest of the paper is organized as follows. The next section provides a background discussion. It sets the context by describing the institutional framework and discussing past research. That section also provides a rationale for the two moderating influences. The third section introduces the assumptions of the model. The fourth section derives the outcome of the negotiations between employer and works council. The fifth section compares the wage and productivity effects of works councils in covered and uncovered establishments. The sixth section concludes.

## 2 Background Discussion

### 2.1 Institutional Framework

Industrial relations in Germany are characterized by a dual structure of employee representation with both works councils and unions. Collective-bargaining agreements are usually negotiated between unions and employers' associations on a broad industrial level. They regulate wage rates and general aspects of the employment contract. Typically, establishments are covered by a collective-bargaining agreement if they are members of an employers' association. The share of establishments covered by firm-level agreements is very small.

Works councils provide a mechanism for establishment-level codetermination. Their rights are defined in the WCA. The creation of a works council depends on the initiative of the establishment's employees. Hence, councils are not present in all eligible establishments. Works councils negotiate over a bundle of interrelated establishment policies. On some issues they have the right to information and consultation, on others a veto power over management initiatives, and on still others the right to coequal participation in the design and implementation of policy. The functions of works councils are distinct from those of unions. Works councils do not have the right to strike. If council and management fail to reach an agreement, they may appeal to an internal arbitration board or to the labor court. Moreover, the WCA does not allow wage negotiations. The aim is to restrict distributional conflicts on the establishment level. Rather, works councils are designed to increase the joint establishment surplus. Council representatives are required by law to cooperate with management "in a spirit of mutual trust […] for the good of the employees and of the establishment."

### *2.2 Generation of Rents*

The possibility of employer opportunism is one explanation why works councils may play the intended role in fostering cooperative and trustful industrial relations. Employees will withhold effort and cooperation when an employer cannot credibly commit to take their interests into account. Worker representation helps protect the interests of the workforce (Askildsen, Jirjahn, and Smith, 2006; Freeman and Lazear, 1995; Kaufman and Levine, 2000; Smith, 1991, 2006). The consultation rights of the works council help reduce information asymmetries between management and workforce so that employees can better evaluate the employer's behavior. Moreover, the codetermination rights help the council prevent the employer from unilaterally taking action without considering employees' interests. Thus, a works council helps create binding commitments of the employer. This in turn increases employees' trust and fosters their cooperation with the implementation of performance-enhancing work practices.

There is a variety of situations in which works councils can potentially help avoid employer opportunism. Let us illustrate this by using performance pay as an example.[1] At their best, performance-pay schemes such as piece rates, bonuses, or profit sharing provide incentives to exert effort by aligning workers' interests with those of the employer. However, performance pay entails dysfunctional incentives

[1] Other situations involve the opportunistic use of information and the opportunistic termination of employment relationships (Jirjahn, 2009). An employer, opportunistically using information, may conceal health and safety problems from the workers or may pretend that the economic situation of the firm requires increased worker effort. Moreover, the employer may use information obtained from the workers against their interests, for example for innovations that entail job loss or intensification of the workload. Similarly, there are several forms of opportunistic termination of employment relationships. Ex ante, an employer may promise employment security in order to induce workers to accumulate firm-specific human capital. However, given that workers' marginal products are stochastic, the employer ex post may be tempted to terminate the employment contracts (Eger, 2004). This also applies to job loss resulting from a management-initiated implementation of organizational change (Frick, 2002). Moreover, the implicit promise of employment security plays a role in deferred-compensation schemes that rearrange earnings profiles by paying workers less than their marginal products early in their tenure and more than their marginal products late in their tenure. If the wages of high-tenured workers exceed their marginal products, the employer may be tempted to renege on the implicit agreements by firing these workers (Heywood and Jirjahn, 2016; Hutchens, 1986).

if workers distrust the employer (Heywood and Jirjahn, 2006). A well-known example is the ratchet effect. Workers, receiving performance pay, withhold effort when they fear that the employer will increase performance standards after a period of good performance. In addition, workers may fear that the measurement of their performance is rather arbitrary and the employer underreports their performance in order to save establishment resources. Even profit sharing will not stimulate effort if workers do not trust the accounting of profit, or fear that management does not pursue complementary investments designed to increase financial performance.

If the employer's commitment problems are not solved, inefficiencies within the establishment will result. On the employer's side, low worker effort implies lower productivity and lower innovativeness. On the workers' side, it implies that they forgo the opportunity of higher wages or better working conditions. Codetermination has the potential to overcome these inefficiencies. A works council helps ensure that any performance-pay plan is implemented as agreed upon. The codetermination rights of the council help prevent the employer from unilaterally altering the payment terms. The works council can also contribute to procedural fairness by helping set clear performance standards and making the measurement of worker performance more transparent. Moreover, the council can monitor the accounting of profit and participate in decisions that influence the financial performance of the establishment. Altogether, the ability of the works council to create binding commitments of the employer increases the trust that workers have in performance pay. Increased trust, in turn, improves the productive incentive effects of performance pay.

Of course, codetermination may not be the only solution to the employer's commitment problems. Under some circumstances, repeated games and reputation concerns can induce an employer to behave honestly (Baker, Gibbons, and Murphy, 1994; Bull, 1987; Kreps, 1990). Therefore, self-enforcing contracts might stand as an alternative in order to engender the trust that is important for workers' cooperation and effort. However, self-enforcing contracts are far from perfect. They fail if the employer overly discounts the future loss of trust and cooperation. Specifically, in situations characterized by economic distress the employer may have an incentive to behave opportunistically and to renege on implicit contracts with the workers (Jirjahn, 2009). In this situation, a works council can protect workers' interests.[2] A works council may even strengthen the functioning of implicit contracts (Hogan, 2001). The council facilitates communication and coordination among employees. To the extent coordinated actions result in more severe punishment of employer opportunism, the employer's incentive to renege on an implicit agreement is reduced.

[2] Thus, works councils appear to be more likely to be implemented in establishments with a poor economic situation (Jirjahn, 2009, 2010) or with a poor company climate (Hauser-Ditz, Hertwig, and Pries, 2013).

### *2.3 Redistribution of Rents*

However, works councils may engage not only in the generation but also in the redistribution of rents (Freeman and Lazear, 1995). Even though the WCA aims at reducing distributional conflicts, a works council may engage in informal wage negotiations with the employer. The council can use its codetermination rights to obtain employer concessions on issues where it has no legal powers. As Addison, Schnabel, and Wagner (2001, pp. 677–680) put it:

> "Although […] the law (specifically section 77(2) of the Works Constitution Act) broadly precludes works agreements on wages, it was also observed that the extensive veto powers enjoyed by works councils in non-wage areas may give them sufficient bargaining leverage to pressure management to pay higher wages […]."

If employer and works council fail to reach an agreement in the informal wage negotiations, the council can threaten to hinder decisions in areas where its consent is necessary. The council may hold up decisions on staff movements or overtime to obtain wage concessions by the employer (Müller-Jentsch, 1995).[3] Even though wage negotiations are informal, the outcome can be binding. The employer can commit to paying higher wages by placing workers in higher wage groups.[4] The council can commit to cooperate with the employer by signing work agreements on issues where it has legal powers.

### *2.4 The Moderating Role of Collective-Bargaining Coverage*

Hübler and Jirjahn (2003) have developed a bargaining model that captures both the generation and the redistribution of rents through codetermination. On the one hand, the trust-building role of a works council allows negotiating work practices that otherwise cannot be implemented. On the other hand, the works council may redistribute economic rents in favor of the employees. Hübler and Jirjahn argue that the opportunities for redistribution are more limited when distributional conflicts are moderated on a central level by unions and employers' associations. Employers' associations support the managers of establishments with expertise in case there are lawsuits. Therefore, the opportunities for a council to obtain employer concessions on issues where it has no legal powers are more restricted. Moreover,

[3] Thelen (1991, chapter 3) even characterizes the informal negotiations as a second wage round. Altogether, there appears to be a broad consensus in the literature that "there is every indication that plant level agreements have ranged well beyond those prescribed by the law" and that "works councils can informally extend their authority to issues that are nowhere covered by the statute" (Addison, 2009, p. 19). With the exception of an early examination by FitzRoy and Kraft (1985), this is supported by a series of studies showing that works councils can have an influence on both the wage level and the wage structure within establishments (for a survey see Addison, 2009, pp. 93–98).

[4] Each wage group specifies a certain pay level. Thus, a worker's remuneration depends on assignment to a particular wage group.

even unions may use their influence to prevent works councils from engaging in redistribution activities. First, negotiations between works councils and managers may undermine the unions' power and status and contribute to dispersed earnings across firms. Second, the unions' interests transcend those of the workforce in an individual establishment. Because of the centralized system of collective bargaining, unions are interested in the industry-wide employment level.

Hübler and Jirjahn's model predicts that a works council should have a more substantial influence on productivity and a less intense influence on wages if the establishment is covered by a collective-bargaining agreement. These predictions have been tested by a series of empirical studies. The studies have used two different data sets. The first one is the Hannover Firm Panel (Gerlach, Hübler, and Meyer, 2003). The data of the Hannover Firm Panel were collected in the 1990s. They cover a sample of manufacturing establishments in the West German federal state of Lower Saxony. The second one is the IAB Establishment Panel (Fischer et al., 2009). The collection of the data started in the 1990s and is still continued on a yearly basis. The IAB Establishment Panel is a sample of establishments from all sectors in the German economy.

Table 1 provides a survey of studies that have examined the productivity effects of works councils in covered and uncovered establishments. Those studies often differ in the method used. They also differ in the time period or the industries considered in the analysis. Nonetheless they show a clear pattern of results: Works councils have a stronger effect on productivity in covered than in uncovered establishments. Thus, as to the productivity effects of works councils, the empirical findings conform to the predictions of Hübler and Jirjahn's model.

As summarized in Table 2, a related pattern of results is even found when alternative indicators of establishment performance are considered. Works councils are more effective in reducing personnel turnover in covered establishments. They appear to be better able to negotiate performance pay arrangements and other HRM practices when the establishment is covered by collective bargaining. There is even evidence that works councils and collective-bargaining coverage have a positive interaction effect on the innovation success and the profitability of establishments.

However, empirical studies are inconclusive as to the wage effects of works councils in covered and uncovered establishments. Table 3 provides a summary of the findings. While some studies confirm that the wage effect is less strong in covered establishments, other studies obtain the opposite result. Gerlach and Meyer (2010) hypothesize that this reflects a decrease in the functionality of collective bargaining in Germany. However, if this were the appropriate explanation, we should also observe an attenuation of the moderating influence collective bargaining has on the productivity effect of works councils. The available studies provide no evidence of such attenuation.

Our model suggests an alternative explanation to reconcile the findings. It adds a second moderating influence to the one considered by Hübler and Jirjahn. Collective-bargaining coverage not only limits redistribution activities of works councils; it also strengthens the effectiveness of the work practices negotiated by

*Table 1*
Works Councils, Collective-Bargaining Coverage, and Productivity

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| Hübler and Jirjahn (2003) | Hannover Firm Panel. Waves 1994 and 1996. | Value added per worker. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. | Significantly positive effect of works councils on productivity in establishments covered by collective-bargaining agreements, but not in uncovered establishments. |
| Hübler (2003) | Hannover Firm Panel. Waves 1994 and 1996. | Value added per worker. | As above. Estimates for a subsample of establishments with 100–300 employees. | Significantly positive effect of works councils on productivity in covered but not in uncovered establishments. |
| Jirjahn (2003a) | Hannover Firm Panel. Waves 1994–1997. | Value added per worker. | Random-effects model. Estimates for all establishments and for a subsample of establishments with 21–100 employees. | Significantly positive effect of works councils on productivity in covered, but not in uncovered establishments. |
| Wagner et al. (2006) | 1994 wave of the Hannover Firm Panel. 2000 wave of the IAB Establishment Panel. | Logarithm of value added per worker. | OLS and quantile regression estimates. Separate estimates for manufacturing and service establishments in West and East Germany. | Significantly positive effect of works councils on productivity only in covered manufacturing establishments in West and East Germany. |
| Renaud (2008) | IAB Establishment Panel. Waves 2000–2003. | Logarithm of sales and logarithm of value added. | OLS estimates. | Significantly positive effect of works councils on productivity in covered and uncovered establishments with the effect being stronger in covered establishments. |

*Table 1*
(continued)

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| Wagner (2008) | Hannover Firm Panel. Wave 1994. | Value added per worker. | Kolmogorov–Smirnov test for first-order stochastic dominance. Analysis for a subsample of establishments with 21–100 employees. | Significantly positive effect of works councils on productivity in covered but not in uncovered establishments. |
| Brändle (2013) | IAB Establishment Panel. Waves 2005–2008. | Logarithm of value added per worker. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. | Significantly positive effect of works councils on productivity in covered and uncovered establishments, the effect being stronger in establishments covered by industry-level collective bargaining. |
| Jirjahn and Mueller (2014) | IAB Establishment Panel. Waves 2001–2006. | Logarithm of sales per worker and logarithm of value added per worker. | Random-effects model. Estimates for all establishments and for a subsample of establishments with 21–100 employees. | Positive interaction effect of works-council presence and coverage by industry-level collective bargaining. The interaction effect is statistically significant in the estimates for all establishments, but not in the estimates for establishments with 21–100 employees. |

*Notes*: The population of the Hannover Firm Panel is all manufacturing establishments with at least five employees in the federal state of Lower Saxony (Gerlach, Hübler, and Meyer, 2003). The population of the IAB Establishment Panel is all establishments with at least one employee covered by social insurance in all sectors in Germany (Fischer et al., 2009).

works councils and employers. Collective-bargaining coverage is associated with a stronger influence of unions (Klodt and Meyer, 1998). Unions usually support works councils (Müller-Jentsch, 1995; Behrens, 2009).[5] While unions have little

[5] Works councils in turn help unions recruit new members.

*Table 2*
Alternative Measures of Establishment Performance

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| *HRM Practices* | | | | |
| Heywood, Hübler, and Jirjahn (1998) | Hannover Firm Panel. Wave 1994. | Use of piece rates and use of profit sharing. | Probit. | Significantly positive effect of works councils on the use of piece rates and profit sharing in covered but not in uncovered establishments. |
| Heywood and Jirjahn (2002) | Hannover Firm Panel. Waves 1994 and 1996. | Use of various types of performance pay. | Probit. | Significantly positive effect of works councils on the use of various types of performance pay in covered but not in uncovered establishments. |
| Jirjahn (2002) | Hannover Firm Panel. Wave 1996. | Use of various types of HRM systems. | Multinomial logit. | Works councils in covered establishments exert a significantly positive influence on the use of an innovative HRM system relying on group incentives, teams, and training. |
| Heywood and Jirjahn (2009) | IAB Establishment Panel. Wave 2002. | Provision of various types of family-friendly practices by the establishment. | Probit. | Share of female employees is positively associated with the provision of family-friendly practices, specifically if there is a works council and the establishment is covered by collective bargaining. |
| Kriechel et al. (2014) | Firm-level data provided by the Institute for Vocational Education and Training (BIBB) for the year 2007. | Net investment in apprenticeship training and share of former trainees still employed. | Propensity-score matching. | Works councils have a positive influence on net investment in apprenticeship training and on the share of former trainees still employed. The effects are stronger in covered establishments. |

*Table 2*
(continued)

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| *Personnel Turnover* | | | | |
| Frick and Möller (2003) | IAB Establishment Panel. Wave 2000. | Log odds of personnel turnover. | OLS estimates. Estimates for all establishments and separate estimates for the manufacturing and the service sector in West and East Germany. | Works councils significantly reduce personnel turnover, the effect being more pronounced in covered establishments. |
| Pfeifer (2011) | IAB Establishment Panel. Wave 2003. | Logarithm of voluntary quits by employees. | Tobit. Estimates for all establishments and for a subsample of establishments with 21–100 employees. | Works councils significantly reduce quits only in establishments covered by industry-level collective bargaining. |
| Heywood, Jirjahn, and Tsertsvardze (2010) | IAB Establishment Panel. Wave 2002. | Employing older workers and hiring older workers. | Probit. Estimates for a subsample of establishments in the federal state of Lower Saxony. | Positive interaction effect of works councils and collective-bargaining coverage on employing older workers. Negative interaction effect on hiring older workers. |
| *Innovation* | | | | |
| Jirjahn (2012) | Hannover Firm Panel. Wave 1995. | Innovation success (share of the establishment's sales generated by new products). | Tobit. | Positive interaction effect of works-council incidence and collective-bargaining coverage on innovation success. |
| Addison et al. (2013) | IAB Establishment Panel. Waves 2008–2011. | Dummy variables for product and process innovations. | Probit and random-effects probit. | Positive interaction effect of works-council incidence and coverage by an industry-level agreement on incremental innovations. |

*Table 2*
(continued)

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| *Profit* | | | | |
| Hübler (2003) | Hannover Firm Panel. Waves 1994 and 1996. | Profit (value added minus wages) per employee. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. Estimates for a subsample of establishments with 100–300 employees. | Significantly positive effect of works councils on profitability in covered but not in uncovered establishments. |
| Mueller (2011) | IAB Establishment Panel. Waves 1996–2007. | Profit (value added minus wages) per employee. | Treatment-effects model to take into account the possible endogeneity of works-council presence. Estimates for a subsample of establishments with 21–300 employees. | Significantly positive effect of works councils on profitability in covered but not in uncovered establishments. |

*Notes*: The population of the Hannover Firm Panel is all manufacturing establishments with at least five employees in the federal state of Lower Saxony (Gerlach, Hübler, and Meyer, 2003). The population of the IAB Establishment Panel is all establishments with at least one employee covered by social insurance in all sectors in Germany (Fischer et al., 2009).

interest in supporting redistribution activities of a works council, they are likely to provide support to strengthen its trust-building role. The support by unions helps a works council create even stronger commitments of the employer so that the council is able to protect workers' interests to a larger degree. This in turn increases workers' trust and their willingness to cooperate and to provide effort. The increased trust and cooperativeness of the workforce imply that the practices negotiated by works council and employer are even more effective in increasing establishment performance.

The support by unions involves the provision of training. Training strengthens the competence of a works council and enables the council to understand the pro-

*Table 3*
Works Councils, Collective Bargaining, and Wages

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| Jirjahn and Klodt (1999) | Hannover Firm Panel. Waves 1994–1996. | Logarithm of wage per employee. | Random-effects estimates. | Significantly positive effect of works councils on wages in uncovered but not in covered establishments. |
| Hübler and Jirjahn (2003) | Hannover Firm Panel. Waves 1994 and 1996. | Wage per employee. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. | Significantly positive effect of works councils on wages in uncovered but not in covered establishments. |
| Hübler (2003) | Hannover Firm Panel. Waves 1994 and 1996. | Wage per employee. | As above. Estimates for a subsample of establishments with 100–300 employees. | No significant effect of works councils on wages. |
| Jirjahn (2003b) | Hannover Firm Panel. Waves 1994–1996. | Logarithm of wage per employee. | OLS. Estimates for a subsample of establishments with 21–100 employees. | Significantly positive effect of works councils on wages in covered and uncovered establishments, the effect being weaker in covered establishments. |
| Guertzgen (2009) | IAB Linked Employer–Employee Dataset. Waves 1995–2001. | Logarithm of the individual wage. | First-differenced regressions. Estimates for a subsample of employees in manufacturing establishments. | Significantly positive effect of works councils on wages only in establishments covered by industry-level collective bargaining. |

*Table 3*
(continued)

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| Addison, Teixeira, and Zwick (2010) | IAB Linked Employer–Employee Dataset. Wave 2001. | Logarithm of the individual wage. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. | Significantly positive effect of works councils on wages in covered and uncovered establishments, the effect being stronger in covered establishments. |
| Gerlach and Meyer (2010) | IAB Establishment Panel. Waves 2001 and 2005. | Wage per employee. | As above. Estimates for a subsample of establishments in the federal state of Lower Saxony. | Significantly positive effect of works councils on wages only in covered but not in uncovered establishments. |
| Guertzgen (2010) | IAB Establishment Panel. Waves 1995–2002. | Logarithm of wage per employee. | Selection approach to take into account the possible endogeneity of collective-bargaining coverage. Estimates for a subsample of establishments from the manufacturing sector. | Significantly positive effect of works councils on wages only in establishments covered by industry-level or firm-level collective bargaining. |
| Jirjahn and Kraft (2010) | Hannover Firm Panel. Wave 1997. | Intra-establishment wage inequality (difference between the wages of skilled and unskilled workers). | Least absolute deviation regression. | Works councils reduce intra-establishment wage inequality, the effect being weaker in covered establishments. |
| Blien et al. (2013) | IAB Linked Employer–Employee Dataset. Waves 1998–2006. | Logarithm of the individual wage. | OLS. Estimates for a subsample of full-time employees in West German establishments. | Significantly positive effect of works councils on wages in covered and uncovered establishments, the effect being more pronounced in establishments covered by industry-level or firm-level collective bargaining. |

*Table 3*
(continued)

| Study | Data | Dependent variable(s) | Method | Findings |
|---|---|---|---|---|
| Brändle (2013) | IAB Establishment Panel. Waves 2005–2008. | Logarithm of wage per employee. | Double-selection approach to take into account the possible endogeneity of works-council presence and collective-bargaining coverage. | Significantly positive effect of works councils on wages, the effect being similar in covered and uncovered establishments. |

*Notes*: The population of the Hannover Firm Panel is all manufacturing establishments with at least five employees in the federal state of Lower Saxony (Gerlach, Hübler, and Meyer, 2003). The population of the IAB Establishment Panel is all establishments with at least one employee covered by social insurance in all sectors in Germany (Fischer et al., 2009). The IAB Linked Employer–Employee Dataset combines worker data from the Employment Statistics Register and establishment data from the IAB Establishment Panel (Alda, Bender, and Gartner, 2005).

duction process and the economic situation of the establishment in more detail.[6] This allows the works council to more effectively reduce information asymmetries and, hence, to better evaluate the employer's behavior. For example, the training may provide know-how that improves the council's ability to monitor the measurement of worker performance and the accounting of profit. It also enables the works council to more effectively participate in a wider range of decisions and to come up with its own valuable ideas. This increases the chance that management undertakes complementary investments to increase the financial performance of the establishment. As a consequence, workers' trust in performance pay schemes such as piece rates and profit sharing is strengthened. This results in an even greater incentive effect of these schemes.

Furthermore, unions provide support and expertise in legal issues. This increases the chance of the works council to win legal disputes and thus strengthens its ability to prevent the employer from breaking promises made to the employees.[7] For example, unions' support in legal issues can increase the council's power to pre-

[6] Jirjahn, Mohrenweiser, and Backes-Gellner (2011) show that learning plays an important role in the functioning of establishment-level codetermination. The training provided by unions can reinforce the learning process.

[7] Related research on union members shows that the legal services provided by unions can be quite effective. Berger and Neugart (2011) find that union members are more likely to be successful in labor-dispute processes than nonmembers. Goerke and Pannenberg (2011) show that union members are less likely to be dismissed. Moreover, in case of a dismissal, a union member has a higher probability of receiving severance pay (Goerke and Pannenberg, 2010).

vent the employer from unilaterally altering the terms of performance pay schemes. This makes performance pay more effective, as it implies a higher degree of trust by workers in these schemes.

Altogether, collective-bargaining coverage may not only limit redistribution activities of a works council; it may also increase the effectiveness of work practices negotiated by works council and employer. As to the wages, the two moderating influences of collective-bargaining coverage work in opposite directions. The moderating influence considered by Hübler and Jirjahn lowers the wage effect of a works council, as it implies a reduction of the council's bargaining power. The additional moderating influence introduced in our theoretical extension increases the wage effect of a works council, as it leads to a higher rent that can be shared by the employer and the employees of the establishment. Thus, the interaction effect of collective bargaining and works councils is ambiguous. This can explain why empirical studies provide mixed results on the wage effects of works councils in covered and uncovered establishments. The studies often differ in the industries and in the time period considered in the analysis. To the extent that the relative weights of the two moderating influences vary across industries or have changed over time, it seems natural that studies using different data obtain mixed results. Table 3 shows that studies with the Hannover Firm Panel usually find a weaker wage effect of works councils in covered than in uncovered establishments. By contrast, studies with the IAB data often obtain a stronger wage effect in covered establishments. As these data cover more recent years and are not restricted to the manufacturing sector, one may conclude that the relative strength of the additional moderating influence considered in our extension has grown over time or is more pronounced in industries other than manufacturing.

Our extension can also explain why empirical examinations considering different time periods or industries have produced a remarkably clear pattern of results on the productivity effect of works councils. The two moderating influences of collective-bargaining coverage work in the same direction with respect to the productivity effect. Thus, it does not matter which influence dominates. As a consequence, studies using different data should indeed find the same pattern of results: The productivity effect of works councils is stronger in covered than in uncovered establishments.

## *3 The Model*

In what follows, the effects of works councils are analyzed within a Nash bargaining framework. Let us consider an establishment with a fixed number $N$ of identical workers. If no works council is present, then the establishment produces an output $F(N)$ and each worker receives a wage $\bar{w}$. We distinguish between coverage ($c$) and no coverage ($nc$) by a collective-bargaining agreement. If the establishment is covered, the wage is $\bar{w} = \bar{w}_c$. If the establishment is not covered, the wage is $\bar{w} = \bar{w}_{nc}$. One may assume $\bar{w}_c > \bar{w}_{nc}$. However, this is not crucial for analyzing the moderating role of collective bargaining.

If a works council is present, establishment-level negotiations over both work practices $e$ and wages $w$ takes place. In this case, the production function is

$$Q(e,N) = \begin{cases} (1+\beta e)F(N) & \text{if there is an agreement,} \\ \alpha F(N) & \text{if there is a conflict,} \end{cases} \tag{1}$$

where $\beta$ is the effectiveness of the work practices and $\alpha$ an inverse measure of the council's opportunities to hinder decisions. The production function captures the generation and the redistribution of rents through codetermination.

Works councils have a rent-generating face, as codetermination provides a mechanism for negotiating work practices that otherwise cannot be implemented.[8] Without a works council, workers will not cooperate with the introduction of new work practices, because they fear employer opportunism. Therefore, $e$ is equal to zero. By contrast, the presence of a works council fosters trust and cooperation, so that the introduction of productivity-enhancing work practices such as performance pay can be negotiated. In this case, $e$ is positive. However, we also allow for the case that the establishment implements work practices that decrease productivity. Employees may prefer work practices requiring lower effort. In that case, $e$ is negative.

Our extension takes into account that the effectiveness $\beta$ of the work practices depends on collective-bargaining coverage. As discussed, unions usually provide works councils with support and expertise to strengthen their position. The strengthened position enables a works council to more effectively prevent the employer from reneging on promises made to the employees. This, in turn, leads to increased trust and to a stronger willingness of the workforce to cooperate with the implementation of work practices. Thus, the effectiveness of the practices is higher in a covered establishment:

$$0 \leq \beta_{nc} < \beta_c, \tag{2}$$

where $\beta_c$ denotes the effectiveness of work practices in case of coverage and $\beta_{nc}$ their effectiveness in case of no coverage.

However, works councils can also engage in redistribution activities, as codetermination rights provide opportunities to hinder decisions if no agreement can be reached. This is captured by $\alpha$. We assume $0 \leq \alpha < 1$ to take into account that a conflict between works council and employer results in lower output. A small $\alpha$ represents a situation where the council has a strong power to disrupt production. As discussed in section 2, this power is more limited in covered establishments.

---

[8] This idea has been suggested by McCain (1980). Models of bargaining over work practices and wages have been used by Haskel (1991), Nickell, Wadhwani, and Wall (1992), and Nickell and Nicolitsas (1997). These models do not consider the case that bargaining over specific work practices is only possible when there is some form of worker representation.

Thus, we assume

$$0 \leq \alpha_{nc} < \alpha_c < 1, \tag{3}$$

where $\alpha_c$ is the inverse measure of the council's power in case of collective-bargaining coverage and $\alpha_{nc}$ the inverse measure of its power in case of no coverage. If no works council is present, we normalize $\alpha$ to be equal to one. The workforce is assumed to have no power to hinder decisions in the absence of a council.

Each worker's utility function has the Stone–Geary form:

$$u(e,w) = (\theta - e)(w - \bar{w}). \tag{4}$$

This utility function captures the idea that each worker compares his or her wage with a reference point. The worker's wage $w$ only yields positive utility if it is greater than the reference wage. We assume that the reference wage is given by $\bar{w}$, the wage the worker receives in the absence of a works council. Furthermore, utility depends on work practices. A negative value of $e$ increases utility. A positive value of $e$ decreases utility because productivity-enhancing work practices require more effort. The parameter $\theta$ is the reference level for productivity-enhancing work practices, above which the worker does not wish to work.

In case of a conflict, the establishment employs the workers by paying them $\bar{w}$. Thus, if there is a disagreement between works council and employer, then each worker has a utility $\bar{u} = 0$ and the establishment's profit is

$$\bar{\pi} = \alpha F(N) - \bar{w}N. \tag{5}$$

If there is an agreement, the profit is

$$\pi(e,w) = (1 + \beta e)F(N) - wN. \tag{6}$$

## 4 Bargaining

In the case that no works council is present, new work practices cannot be negotiated, due to the lack of trust and cooperation. Moreover, workers cannot push through higher wages, as they have no opportunity to hinder decisions, i.e., $\alpha = 1$. Hence, this situation is a no-bargaining situation, characterized by $e^* = 0$ and $w^* = \bar{w}$.

If a works council is present, we have a bargaining situation. First, the council can threaten to hinder decisions in case of a conflict. Second, council and employer can negotiate both wages and work practices. The Nash product is

$$\Omega(e,w) = [Nu(e,w) - N\bar{u}]^{0.5}[\pi(e,w) - \bar{\pi}]^{0.5}.$$

Taking equations (4), (5), and (6) into account, we obtain

$$\Omega(e,w) = [N(\theta - e)(w - \bar{w})]^{0.5}[(1 + \beta e)F(N) - wN - (\alpha F(N) - \bar{w}N)]^{0.5}.$$

The Nash product is maximized by choosing $w$ and $e$. This yields

$$
\begin{aligned}
w^{**} &= \bar{w} + \frac{1}{3}(\theta\beta + 1 - \alpha)\frac{F(N)}{N}, \\
e^{**} &= \frac{1}{3}\left[2\theta - \frac{1}{\beta}(1-\alpha)\right].
\end{aligned} \tag{7}
$$

The influence of the works council on work practices is ambiguous. If $\theta\beta \leq 0.5(1-\alpha)$, we obtain $e^{**} \leq 0$. If $\theta\beta > 0.5(1-\alpha)$, we obtain $e^{**} > 0$.

The expression $e^{**}$ for work practices is increasing in $\beta$. This implies that the work practices negotiated between works council and employer are more likely to be productivity-enhancing if the effectiveness of the practices is high. The wage $w^{**}$ is also increasing in $\beta$. Thus, the council has an effect on wages due to potentially increased establishment performance. Even if codetermination had no effect on workers' power to disrupt production, we would observe an influence on wages because the works council can help implement productivity-enhancing work practices.

However, codetermination increases workers' power to disrupt production. The expression $(1-\alpha)$ captures this effect. While $w^{**}$ is increasing in $1-\alpha$, $e^{**}$ is decreasing in $1-\alpha$. Thus, the effect of codetermination on wages is stronger and its effect on performance-enhancing work practices is less strong if the works council has more opportunities to hinder decisions.

## 5 *Works-Council Effects and Collective-Bargaining Coverage*

It is straightforward to derive the works-council effect on productivity. From equations (1) and (7) we obtain the establishment's productivity when a council is present:

$$\frac{Q^{**}}{N} = \left[1 + \frac{1}{3}(2\theta\beta - (1-\alpha))\right]\frac{F(N)}{N}.$$

In the absence of a council, the establishment's productivity is $Q^{*}/N = F(N)/N$. Thus, the works-council effect on productivity is

$$\Delta q = \frac{Q^{**}}{N} - \frac{Q^{*}}{N} = \frac{1}{3}[2\theta\beta - (1-\alpha)]\frac{F(N)}{N}.$$

As both the effectiveness of the negotiated work practices and the opportunities to hinder decisions depend on collective-bargaining coverage, we can write

$$\Delta q_c = \frac{1}{3}[2\theta\beta_c - (1-\alpha_c)]\frac{F(N)}{N}, \quad \Delta q_{nc} = \frac{1}{3}[2\theta\beta_{nc} - (1-\alpha_{nc})]\frac{F(N)}{N},$$

where $\Delta q_c$ is the effect on productivity in the presence of collective-bargaining coverage, and $\Delta q_{nc}$ the effect in the absence of collective-bargaining coverage. Taking (2) and (3) into account, it follows immediately that $\Delta q_c > \Delta q_{nc}$.

PROPOSITION 1 *The works-council effect on productivity is greater if the establishment is covered by a collective-bargaining agreement.*

For two reasons, collective-bargaining coverage exerts a positive moderating influence on the relationship between works-council presence and productivity. First, the opportunities for the works council to hinder decisions are more restricted if the establishment is covered by a collective-bargaining agreement. Thus, the works council has less power to push through work practices that require only low effort. Second, the effectiveness of the negotiated work practices is higher if the establishment is covered by a collective-bargaining agreement. As a consequence, the employer and the works council tend to negotiate a higher amount of productivity-enhancing work practices.

The works-council effect on each worker's wage is

$$\Delta w = w^{**} - w^{*} = \frac{1}{3}(\theta\beta + 1 - \alpha)\frac{F(N)}{N}.$$

The strength of this effect also depends on the collective-bargaining coverage:

$$\Delta w_c = \frac{1}{3}(\theta\beta_c + 1 - \alpha_c)\frac{F(N)}{N}, \quad \Delta w_{nc} = \frac{1}{3}(\theta\beta_{nc} + 1 - \alpha_{nc})\frac{F(N)}{N}.$$

We immediately obtain Proposition 2.

PROPOSITION 2 *If $\alpha_c - \alpha_{nc}$ is greater than (smaller than, equal to) $\theta(\beta_c - \beta_{nc})$, the works-council effect on wages is smaller (greater, the same) in a covered establishment than in an uncovered establishment.*

As to the wage effect of codetermination, the two moderating influences of collective-bargaining coverage work in opposite directions. On the one hand, collective-bargaining coverage limits the council's opportunities to engage in redistribution activities. This decreases the wage effect of the work council. On the other hand, collective-bargaining coverage improves the effectiveness of work practices negotiated between council and employer. This increases the establishment surplus shared with the workforce through higher wages. Depending on whether the first or the second influence dominates, collective-bargaining coverage weakens or strengthens the wage effect of the council.

## 6 Conclusions

This paper extends Hübler and Jirjahn's theoretical analysis by considering a second moderating influence of collective-bargaining coverage. Collective-bargaining coverage not only limits the opportunities of works councils to engage in redistribution activities. It also strengthens the trust-building role of works councils and hence enables the implementation of more effective work practices. As to the productivity effect of works councils the two moderating influences work in the same direction, implying a stronger effect in covered establishments. However, as to the wage effect of works councils they work in opposite directions, so that, depending on the relative strength of the two influences, the wage effect can be weaker or stronger in covered establishments.

The empirical evidence conforms to these predictions. While most studies find a stronger productivity effect in covered establishments, empirical research is inconclusive as to the wage effect of works councils in covered and uncovered establishments. The studies often differ in the industries or in the time period considered. It seems natural that there may be variations in the relative weights of the two moderating influences across industries or time periods. Our analysis suggests that these variations play little role in studies examining the productivity effects of works councils. As the two moderating influences of collective-bargaining coverage work in the same direction, those studies should find a stronger productivity effect in covered establishments, regardless of which influence dominates. Yet, variations in the relative weights of the two moderating influences should play a decisive role in studies examining the wage effects of works councils. If collective-bargaining coverage primarily limits redistribution activities of works councils, we should find a weaker wage effect in covered establishments. If it primarily improves the effectiveness of the work practices negotiated by works councils and employers, we should find a stronger wage effect in covered establishments.

The model has implications for future empirical research. It suggests performing separate analyses by industry. If the relative weights of the two moderating influences vary across industries, this would allow identifying industries where collective-bargaining coverage weakens the wage effects of works councils and industries where it strengthens the wage effects. Moreover, it would be interesting to perform separate analyses for different time periods. If the relative weights of the two influences have changed over time, we may identify time periods characterized by weaker and time periods characterized by stronger wage effects of works councils in covered than in uncovered establishments. This may provide important indications of changes in the functioning of centralized collective bargaining in Germany. Finally, future research could use alternative data to test the assumptions and implications of the extended model in more detail. Specifically, the NIFA panel provides detailed information on HRM practices. This would allow testing whether collective-bargaining coverage indeed contributes to increased effectiveness of practices negotiated by works council and employer.

## *References*

Addison, J. T. (2009), *The Economics of Codetermination: Lessons from the German Experience*, Palgrave Macmillan, New York.

—, C. Schnabel, and J. Wagner (2001), "Works Councils in Germany: Their Effects on Establishment Performance," *Oxford Economic Papers*, 53(4), 659–694.

—, P. Teixeira, K. Evers, and L. Bellmann (2013), "Collective Bargaining and Innovation in Germany: Cooperative Industrial Relations?" Discussion Paper 7871, IZA, Bonn.

—, —, and T. Zwick (2010) "German Works Councils and the Anatomy of Wages," *Industrial & Labor Relations Review*, 63(2), 247–270.

Alda, H., S. Bender, and H. Gartner (2005), "The Linked Employer–Employee Dataset Created from the IAB Establishment Panel and the Process-Produced Data of the IAB (LIAB)," *Schmollers Jahrbuch*, 125(2), 327–336.

Askildsen, J. E., U. Jirjahn, and S. C. Smith (2006), "Works Councils and Environmental Investment: Theory and Evidence from German Panel Data," *Journal of Economic Behavior & Organization*, 60(3), 346–372.

Baker, G., R. Gibbons, and K. J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, 109(4), 1125–1156.

Behrens, M. (2009), "Still Married after All these Years? Union Organizing and the Role of Works Councils in German Industrial Relations," *Industrial & Labor Relations Review*, 62(3), 275–293.

Berger, H., and M. Neugart (2011), "How German Labor Courts Decide: An Econometric Case Study," *German Economic Review*, 13(1), 56–70.

Blien, U., W. Dauth, T. Schank, and C. Schnabel (2013), "The Institutional Context of an 'Empirical Law': The Wage Curve under Different Regimes of Collective Bargaining," *British Journal of Industrial Relations (BJIR)*, 51(1), 59–79.

Brändle, T. (2013), "Flexible Collective Bargaining Agreements: Still a Moderating Effect on Works Council Behaviour?" Discussion Paper 96, IAW, Tübingen.

Bull, C. (1987), "The Existence of Self-Enforcing Implicit Contracts," *The Quarterly Journal of Economics*, 102(1), 147–159.

Eger, T. (2004), "Opportunistic Termination of Employment Contracts and Legal Protection against Dismissal in Germany and the USA," *International Review of Law and Economics*, 23(4), 381–403.

Fischer, G., F. Janik, D. Müller, and A. Schmucker (2009) "The IAB Establishment Panel – Things Users Should Know," *Schmollers Jahrbuch*, 129(1), 133–148.

FitzRoy, F. R., and K. Kraft (1985), "Unionization, Wages and Efficiency: Theories and Evidence from the US and West Germany," *Kyklos*, 38(4), 537–554.

Freeman, R. B., and E. P. Lazear (1995), "An Economic Analysis of Works Councils," in: J. Rogers and W. Streeck (eds.), *Works Councils: Consultation, Representation, and Cooperation in Industrial Relations*, The University of Chicago Press, Chicago (IL), pp. 27–52.

— and J. Rogers (1999), *What Workers Want*, ILR Press, Ithaca (NY).

Frick, B. (2002), " 'High Performance Work Practices' und betriebliche Mitbestimmung: komplementär oder substitutiv? Empirische Befunde für den deutschen Maschinenbau," *Industrielle Beziehungen*, 9(1), 79–102.

— and I. Möller (2003), "Mandated Works Councils and Firm Performance: Labor Productivity and Personnel Turnover in German Establishments," *Schmollers Jahrbuch*, 123(3), 423–454.

Gerlach, K., O. Hübler, and W. Meyer (2003), "The Hannover Firm Panel (HFP)," *Schmollers Jahrbuch*, 123, 463–470.

— and W. Meyer (2010), "Wage Effects of Works Councils and Collective Agreements in Germany," in: C. Brown, B. Eichengreen, and M. Reich (eds.), *Labor in the Era of Globalization*, Cambridge University Press, New York, pp. 295–316.

Goerke, L., and M. Pannenberg (2010), "An Economic Analysis of Dismissal Legislation: Determinants of Severance Pay in West Germany," *International Review of Law and Economics*, 30(1), 71–85.

— and — (2011), "Trade Union Membership and Dismissals," *Labour Economics*, 18(6), 810–821.

Guertzgen, N. (2009), "Rent-Sharing and Collective Bargaining Coverage: Evidence from Linked Employer–Employee Data," *The Scandinavian Journal of Economics*, 111(2), 323–349.

— (2010), "Rent-Sharing and Collective Wage Contracts – Evidence from German Establishment-Level Data," *Applied Economics*, 42(22), 2835–2854.

Haskel, J. (1991), "Imperfect Competition, Work Practices and Productivity Growth," *Oxford Bulletin of Economics and Statistics*, 53(3), 265–279.

Hauser-Ditz, A., M. Hertwig, and L. Pries (2013), "Collective Employee Representation in German Companies: Coverage and Contextual Conditions," *Employee Relations*, 35(5), 509–526.

Heywood, J. S., O. Hübler, and U. Jirjahn (1998), "Variable Payment Schemes and Industrial Relations: Evidence from Germany," *Kyklos*, 51(2), 237–257.

— and U. Jirjahn (2002), "Payment Schemes and Gender in Germany," *Industrial & Labor Relations Review*, 56(1), 44–64.

— and — (2006), "Performance Pay: Determinants and Consequences," in: D. Lewin (ed.), *Contemporary Issues in Employment Relations*, Labor and Employment Relations Association, Champaign (IL), pp. 149–188.

— and — (2009), "Family-Friendly Practices and Worker Representation in Germany," *Industrial Relations*, 48(1), 121–145.

— and — (2016), "The Hiring and Employment of Older Workers in Germany: A Comparative Perspective," *Journal for Labour Market Research*, published Online First December 10, 2015, DOI: 10.1007/s12651-015-0195-4.

—, —, and G. Tsertsvardze (2010), "Hiring Older Workers and Employing Older Workers: German Evidence," *Journal of Population Economics*, 23(2), 595–615.

Hogan, C. (2001), "Enforcement of Implicit Contracts through Unionization," *Journal of Labor Economics*, 19(1), 171–195.

Hübler, O. (2003), "Zum Einfluss des Betriebsrates in mittelgroßen Unternehmen auf Investitionen, Löhne, Produktivität und Renten – Empirische Befunde," in: N. Goldschmidt (ed.), *Wunderbare Wirtschaftswelt – Die New Economy und ihre Herausforderungen: 4. Freiburger Wirtschaftssymposium*, Nomos Verlagsgesellschaft, Baden-Baden, pp. 77–94.

— and U. Jirjahn (2003), "Works Councils and Collective Bargaining in Germany: The Impact on Productivity and Wages," *Scottish Journal of Political Economy*, 50(4), 471–491.

Hutchens, R. (1986), "Delayed Payment Contracts and a Firm's Propensity to Hire Older Workers," *Journal of Labor Economics*, 4(4), 439–457.

Jenkins, J., and P. Blyton (2008), "Works Councils," in: P. Blyton, N. Bacon, J. Fiorito, and E. Heery (eds.), *The SAGE Handbook of Industrial Relations*, SAGE Publications, London, pp. 346–357.

Jirjahn, U. (2002), "The German Experience with Performance Pay," in: M. Brown and J. S. Heywood (eds.), *Paying for Performance: An International Comparison*, M.E. Sharpe, New York and London, pp. 148–178.

— (2003a), "Produktivitätswirkungen betrieblicher Mitbestimmung – Welchen Einfluss haben Betriebsgröße und Tarifbindung?" *Zeitschrift für Betriebswirtschaft*, 73(Ergänzungsheft 4), 63–85.

— (2003b), "Betriebsräte, Tarifverträge und betriebliches Lohnniveau," *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 36(4), 649–660.

— (2009), "The Introduction of Works Councils in German Establishments – Rent Seeking or Rent Protection?" *British Journal of Industrial Relations (BJIR)*, 47(3), 521–545.

— (2010), "Works Councils and Employment Growth in German Establishments," *Cambridge Journal of Economics*, 34(3), 475–500.

— (2011), "Ökonomische Wirkungen der Mitbestimmung in Deutschland: Ein Update," *Schmollers Jahrbuch*, 131(1), 3–57.

— (2012), "Industrielle Beziehungen und Innovationserfolg," in: K. Gerlach, O. Hübler, and S. L. Thomsen (eds.), *Arbeitsmarkt und Arbeitsmarktpolitik in Niedersachsen*, NIW Vortragsreihe, Band 18, Hannover, pp. 43–53.

— and T. Klodt (1999), “Lohnhöhe, industrielle Beziehungen und Produktmärkte,” in: L. Bellmann, S. Kohaut, and M. Lahner (eds.), *Zur Entwicklung von Lohn und Beschäftigung auf der Basis von Betriebs- und Unternehmensdaten*, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nuremberg, pp. 27–54.
— and K. Kraft (2010), “Teamwork and Intra-Firm Wage Dispersion among Blue-Collar Workers,” *Scottish Journal of Political Economy*, 57(4), 404–429.
—, J. Mohrenweiser, and U. Backes-Gellner (2011), “Works Councils and Learning: On the Dynamic Dimension of Codetermination,” *Kyklos*, 64(3), 427–447.
— and S. Mueller (2014), “Non-Union Worker Representation, Foreign Owners and the Performance of Establishments,” *Oxford Economic Papers*, 66(1), 140–163.
Kaufman, B. E., and D. I. Levine (2000), “An Economic Analysis of Employee Representation,” in: B. E. Kaufman and D. G. Taras (eds.), *Nonunion Employee Representation: History, Contemporary Practice, and Policy*, M.E. Sharpe, New York, pp. 149–175.
Klodt, T., and W. Meyer (1998), “Empirical Analysis of Inter-Firm Differences in Trade Union Density,” Discussion Paper 13, Forschungsstelle Firmenpanel, Universität Hannover.
Kreps, D. M. (1990), “Corporate Culture and Economic Theory,” in: J. E. Alt and K. A. Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge University Press, Cambridge, pp. 90–143.
Kriechel, B., S. Muehlemann, H. Pfeifer, and M. Schütte (2014), “Works Councils, Collective Bargaining, and Apprenticeship Training – Evidence from German Firms,” *Industrial Relations*, 53(2), 199–222.
McCain, R. A. (1980), “A Theory of Codetermination,” *Zeitschrift für Nationalökonomie / Journal of Economics*, 40(1/2), 65–90.
Mueller, S. (2011), “Works Councils and Firm Profits Revisited,” *British Journal of Industrial Relations (BJIR)*, 49(S1), s27–s43.
Müller-Jentsch, W. (1995), “Germany: From Collective Voice to Co-Management,” in: Rogers and Streeck (eds.) (1995), pp. 53–78.
Nickell, S., and D. Nicolitsas (1997), “Wages, Restrictive Practices and Productivity,” *Labour Economics*, 4(3), 201–221.
—, S. Wadhwani, and M. Wall (1992), “Productivity Growth in U.K. Companies, 1975–1986,” *European Economic Review*, 36(5), 1055–1091.
Pfeifer, C. (2011), “Works Councils, Union Bargaining and Quits in German Firms,” *Economic and Industrial Democracy*, 32(2), 243–260.
Renaud, S. (2008), *Arbeitnehmermitbestimmung im Strukturwandel*, Metropolis-Verlag, Marburg.
Rogers, J., and W. Streeck (eds.) (1995), *Works Councils: Consultation, Representation, and Cooperation in Industrial Relations*, The University of Chicago Press, Chicago (IL).
Smith, S. C. (1991), “On the Economic Rationale for Codetermination Law,” *Journal of Economic Behavior & Organization*, 16(3), 261–281.
— (2006), “Employee Participation Rights in Corporate Governance: An Economic Rationale, a Test of a Leading Theory, and Some Initial Policy Proposals,” in: P. Kalmi and M. Klinedinst (eds.), *Advances in the Economic Analysis of Participatory and Labor-Managed Firms, Volume 9: Participation in the Age of Globalization and Information*, Emerald Group Publishing, Bingley, pp. 105–146.
Thelen, K. A. (1991), *Union of Parts: Labor Politics in Postwar Germany*, Cornell University Press, Ithaca (NY).
Wagner, J. (2008), “German Works Councils and Productivity: First Evidence from a Nonparametric Test,” *Applied Economics Letters*, 15(9), 727–730.
—, T. Schank, C. Schnabel, and J. T. Addison (2006), “Works Councils, Labor Productivity and Plant Heterogeneity: First Evidence from Quantile Regressions,” *Jahrbücher für Nationalökonomie und Statistik*, 226(5), 505–518.

Uwe Jirjahn
Lehrstuhl für Arbeitsmarktökonomik
Fachbereich IV
Universität Trier
Universitätsring 15
54286 Trier
Germany
jirjahn@uni-trier.de

# Optimal Partial Privatization with Asymmetric Demand Information

by

John S. Heywood, Xiangting Hu, and Guangliang Ye*

Received June 18, 2015; in revised form March 14, 2016;
accepted May 25, 2016

We study a mixed duopoly in which only the private firm directly knows product demand and examine a pooling equilibrium in which a welfare-maximizing government may partially privatize the public firm. We show that the optimal extent of privatization differs, often dramatically, from that without asymmetric information. Indeed, we identify circumstances in which the optimal extent of privatization is zero, in sharp contrast with previous work. In addition, the consumer surplus in the pooling equilibrium routinely exceeds that without asymmetric information, and the social welfare exceeds that without asymmetric information when the cost convexity is small. (JEL: L1, L3)

## 1 Introduction

Public firms are often partially privatized in order to improve efficiency through monitoring by shareholders and the resulting "yardstick competition." At the same time public firms may also suffer from being less aware and responsive to customer demand. This lack of responsiveness has been noted by researchers but has never been incorporated into an analysis of the optimal extent of privatization. We provide such an analysis, showing that the optimal extent of privatization is dra-

* John S. Heywood: Department of Economics, University of Wisconsin–Milwaukee, Milwaukee (WI), U.S.A.; Xiangting Hu: Hanqing Advanced Institute of Economics and Finance, Renmin University of China, Beijing, P.R. China; Guangliang Ye (corresponding author): Antitrust and Competition Policy Center, Hanqing Institute, National Academy of Development and Strategy at Renmin University of China, Beijing, P.R. China. The authors thank Toshihiro Matsumura, Leonard Wang and participants at the seminar at University of Tokyo and Pacific Rim Conference of Western Economic Association International for helpful comments. Financial support is acknowledged from the Chinese National Natural Science Foundation (No. 71273270 and No. 71133006), the Fok Ying Tung Education Foundation (No. 141082), the Ten Thousand Talent Program, the Fundamental Research Fund for Central Universities, and the Research Fund of Renmin University (No. 14XNI006).

*Journal of Institutional and Theoretical Economics* 173, 347–375 – ISSN 0932-4569
DOI: 10.1628/093245616X14700337830936 

matically changed when a public firm faces asymmetric information relative to its private-sector competitor. Moreover, the optimal response of the government may actually be to retain a larger public share, not privatize more, as our analysis will show.

At the center of the mixed-oligopoly literature stands the issue of whether or not the government can improve social welfare by taking an ownership position, a possibility that Merill and Schneider (1966) originally labeled as "regulation by participation." The issue is not trivial, as the government, by assumption, cannot fully replace the private sector and mimic the results of perfect competition. Thus, despite the welfare-maximizing objective of the government, it faces constraints that leave in doubt the welfare consequences of its participation. Previous research views these constraints as exclusively on the cost or supply side of the market. If costs are linear, government firms are assumed to have higher costs than private firms (see, for example, White, 2002), or if costs are identical but convex, the government is limited to having a single firm and so its increased production gives it higher costs than private firms (De Fraja and Delbono, 1989). The suggestion that public firms have higher costs reflects, in part, the notion that political considerations, or bureaucracies themselves, reduce technical efficiency. We suggest that these same factors may also make the public firm less attuned to, and aware of, the exact nature of consumer demand. We examine the ability of the government to regulate by participation when assuming asymmetric information about product demand.

De Fraja and Delbono (1989) show that the presence of a public firm increases welfare when there are only a few Cournot rivals but decreases welfare when the number of rivals is large. The welfare effects remain ambiguous when strategic delegation is allowed (Barros, 1995; Du, Heywood, and Ye, 2013), and empirical studies present conflicting effects of privatization on firm and market performance (see Megginson and Netter, 2001, for an early survey). Yet, in contrast to such examinations of "pure" public firms, many government equity positions are in partially privatized firms. Jones et al. (1999) show that of share-issue privatizations from 59 countries fully 90 % of firms were only partially privatized. Maw (2002) and Fan, Wong, and Zhang (2007) illustrate that privatization in many transition economies has been mainly partial, while D'Souza and Megginson (1999) and Bortolotti, Fantini, and Siniscalco (2003) emphasize the importance of partially privatized companies in both developed and transition economies.[1] Matsumura's (1998) seminal paper reflects the importance of such mixed-ownership firms by showing that the optimal position for a government to take in a single firm in a duopoly is never complete ownership. He shows that an intermediate equity position, partial

[1] A variety of alternative rationales have been presented for partial privatization. Some scholars argue that it reflects government budget constraints and legal institutions (Dewenter and Malatesta, 2001; Bortolotti, Fantini, and Siniscalco, 2003), while others emphasize that it could be politically motivated to attract median voters (Biais and Perotti, 2002). There may also be minority stakes that are sufficient for financial markets to discipline otherwise fully public firms (Gupta, 2005).

privatization, allows the optimal combination of beneficial output expansion and harmful cost increases. We return to this issue in our model of asymmetric information by identifying the optimal degree of partial privatization when a government regulates by participation.

A substantial literature isolates how incomplete or private information influences the behavior of private firms. Milgrom and Roberts (1982) show that incomplete information can deter entry when an established firm has private, payoff-relevant information such as costs.[2] More germane, Riordan (1985) studies firms that do not directly observe the demand curve or the previous quantity decisions of rivals. We follow this tradition by presenting a dynamic model in which a private firm and a public firm simultaneously choose quantities in each period, but only the private firm observes the true demand level. The public firm observes only the resulting market price. Again, we emphasize the notion that the public firm fulfills political obligations and has a resulting bureaucracy that makes it less sensitive to variations in market demand. As an illustration, the top management teams in (partially privatized) state-owned firms in China remain appointed by government officials. Li and Zhou (2005) argue that the decision-makers in these firms are more attuned to political incentives than to market conditions. Such concerns also exist in Western economies, where a substantial literature has emphasized the role that political considerations play in the pricing decisions of public firms (see Klien, 2014, for a recent review). We exploit the possibility that public firms may be less attuned to market demand to identify a pooling equilibrium in which the private firm produces a high level of output even when demand is high in order to present a low price and so hide the true information from the public firm. This is done in order to induce the public firm to produce less. Yet, the consequence is that total equilibrium output and so consumer surplus are greater than without information asymmetry. Indeed, for lower cost convexity social welfare is also greater.

We demonstrate that the optimal extent of privatization in a pooling equilibrium differs, often dramatically, from that without asymmetric information. When cost convexity is small, the optimal degree of privatization exceeds that without asymmetric information. This follows because the private firm in the pooling equilibrium produces a high enough level of output that the marginal social benefit from the public firm increasing output is less than without asymmetric information. Thus, to maximize social welfare, the government privatizes the public firm more to induce a lower output level and reduce costs. When cost convexity is high, this logic is reversed and the optimal extent of privatization is shown to be less than that without asymmetric information. Indeed, for sufficient cost convexity, there is actually no incentive to partially privatize a fully public firm, a sharp departure from the typical finding under an otherwise similar model without asymmetric information (Matsumura, 1998). This happens because the private firm actually increases production sufficiently that it is above that of the public firm during high demand. The

[2] Also see Matthews and Mirman (1983), Gaskins (1971), and Seamans (2013), among others.

resulting production asymmetry increases costs that can be alleviated (rather than made worse) by additional public production.

We recognize that the quality of the information might vary with the extent of privatization. Thus, with sufficient privatization, the government creates a management that is just as attuned to variations in demand as is the management of the private firm. In an extension, we modify our model to imagine such a threshold level of privatization and identify the conditions under which the pooling equilibrium may still exist.

The remainder of the paper is organized as follows. Section 2 sets up the model under the assumption of quadratic costs. Section 3 solves the equilibrium in the benchmark case with symmetric information, and section 4 derives the pooling equilibrium with information asymmetry. In section 5, we make comparisons and highlight that the optimal extent of privatization in the pooling equilibrium can be either higher or lower than in the benchmark and can include no privatization. Section 6 relaxes the assumption of quadratic costs and shows that many, but not all, of the critical results can still be derived. In general we show that the pooling equilibrium is associated with higher consumer surplus and can be associated with greater social welfare. Section 7 discusses the case in which the ability of the public firm to detect market information depends on the extent of privatization, and section 8 concludes.

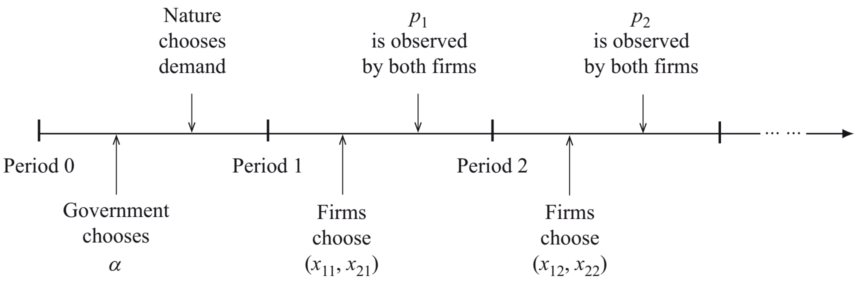
## 2 *Model Setup*

Consider a public firm and a private firm producing a homogeneous good in a single market. Let firm 1 be the public firm and firm 2 be the private firm. Suppose that the game has an infinite horizon from period 0, 1, 2, ..., $t$, ...to infinity, with discount factor $\delta$. The price in period $t$ is denoted as $p_t$, and the quantities of firm $i$ $(i = 1,2)$ in period $t$ are denoted as $x_{it}$. Firms face two types of demand: $p^r(x) = d^r - x$, $r = h,l$, where $h$ denotes the high demand and $l$ denotes the low demand, and we let $d^h > d^l$. Nature randomly decides whether demand is high ($h$) or low ($l$) at the beginning of the game; and demand is $h$ with probability $\mu_0$ and $l$ with probability $1-\mu_0$. Following our discussion in the previous section, we assume that only the private firm observes the true state of demand.[3]

The timing of the game is illustrated in Figure 1. At the beginning of the game (period 0), the government first chooses the optimal privatization ratio to maximize the ex ante social welfare; then, Nature randomly decides whether demand is high

[3] However, we will allow the partially privatized firm to gain access to the demand information in section 7.

*Figure 1*
Timing of the Game




or low. After that (periods 1 to $\infty$), the firms play a simultaneous quantity game in each period, and at the end of each period, both firms observe the price.[4]

In our initial presentation the cost functions are quadratic, $c_i(x) = cx_{it}^2$, an assumption we relax later. The private firm maximizes its profit:

$$\pi_{2t}(x_{2t}) = p_t x_{2t} - cx_{2t}^2.$$

Following Matsumura (1998), the public firm (potentially partially privatized) maximizes the expected value of a convex combination of social welfare and its own profit:

$$Eu_{1t}(x_{1t}) = \alpha E w_t + (1-\alpha)E\pi_{1t},$$

where $\pi_{1t}$ denotes the profit in period $t$, and $w_t$ denotes the social welfare (sum of all profits and consumer surplus) in period $t$. The parameter $\alpha \in [0,1]$ is the nationalization ratio (thus, $1-\alpha$ is the privatization ratio). When $\alpha = 0$, the firm is fully private and maximizes profit; and when $\alpha = 1$, the firm is fully public and maximizes welfare. The government chooses $\alpha$ at the beginning of the game with the objective of maximizing social welfare (Matsumura, 1998). At the beginning of the game, the public firm knows that the demand is $h$ with probability $\mu_0$.

In each period, the firms simultaneously choose quantities. If the public firm holds the belief that the demand is $h$ with probability $\mu_t$ at the beginning of period $t$, its expected profit in period $t$ is

$$E\pi_{1t}(x_{1t}, x_{2t}^h, x_{2t}^l) = \mu_t \cdot \left[p_t^h(x_t^h) \cdot x_{1t} - cx_{1t}^2\right] + (1-\mu_t) \cdot \left[p_t^l(x_t^l) \cdot x_{1t} - cx_{1t}^2\right],$$

where $x_{2t}^h$ and $x_{2t}^l$ denote the quantities firm 2 chooses in period $t$ when the demand is $h$ or $l$, respectively. To simplify notation we let $x_t^r = x_{1t} + x_{2t}^r$ for $r = h, l$. The

---

4 We adopt an infinite repeated-game setting because it represents a natural setting for the signaling process. Thus, we focus on a steady state rather than a single period. In addition, compared to a dynamic game with a finite horizon (a two-period model, for example), the infinite horizon minimizes the effect of any discount factor on equilibrium outcomes and payoffs, as we will see later.

expected social welfare in period $t$ is

$$
\begin{aligned}
Ew_t(x_{1t},x_{2t}^h,x_{2t}^l) = \mu_t \cdot \left[\int_0^{x_t^h} p^h(q)dq - cx_{1t}^2 - c(x_{2t}^h)^2\right] \\
+ (1-\mu_t)\cdot\left[\int_0^{x_t^l} p^l(q)dq - cx_{1t}^2 - c(x_{2t}^l)^2\right].
\end{aligned}
\tag{1}
$$

Next, we consider the benchmark case in which both firms observe the realized demand level. We present this so as to compare it with the asymmetric information case outlined and identify the consequences of the information structure.

## 3 Benchmark Case with Symmetric Information

In this section, both firms observe the realized demand. For repeated games with an infinite horizon, there would exist infinite equilibria. In this paper, we restrict our equilibrium set by requiring behavior to have a Markovian property – i.e., firms' behavioral strategies in period $t$ depend only on $\mu_t$. When information is symmetric, the equilibrium is that the two firms play one-shot Nash in each period.

In each period, payoff maximization generates the first-order conditions (FOCs) for two firms when demand is $r$ $(r = h,l)$:

$$
\begin{aligned}
\frac{\partial Eu_{1t}}{\partial x_{1t}^r} &= -(1-\alpha)\cdot x_{1t}^r + d^r - (x_{1t}^r + x_{2t}^r) - 2cx_{1t}^r = 0,\\
\frac{\partial \pi_{2t}}{\partial x_{2t}^r} &= -x_{2t}^r + d^r - (x_{1t}^r + x_{2t}^r) - 2cx_{2t}^r = 0.
\end{aligned}
$$

Simultaneously solving the above equations generates the equilibrium quantities $(x_{1t}^{r,B}, x_{2t}^{r,B})$,

$$
\begin{aligned}
x_{1t}^{h,B} &= \frac{(1+2c)d^h}{3-2\alpha+8c-2c\alpha+4c^2}, & x_{2t}^{h,B} &= \frac{(1-\alpha+2c)d^h}{3-2\alpha+8c-2c\alpha+4c^2},\\
x_{1t}^{l,B} &= \frac{(1+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2}, & x_{2t}^{l,B} &= \frac{(1-\alpha+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2},
\end{aligned}
$$

where the superscript $B$ denotes the *benchmark* case without asymmetric information. These quantities make clear that for any degree of nationalization $\alpha > 0$ the output of the private firm is below that of its public rival. This difference in outputs increases the total production cost (relative to dividing the same total output between the two firms) and so, even as increasing nationalization increases total output and so consumer surplus, it inefficiently increases production costs. This is the essence of the point by Matsumura (1998) that an intermediate level of nationalization maximizes welfare by balancing the increase in consumer surplus with the increase in costs associated with the difference in outputs.

Since firm behavior is identical in every period, we omit the subscript $t$, and the expected welfare becomes[5]

$$
\begin{aligned}
Ew^B(\alpha) = \mu_0 \cdot & \left[ \int_0^{x_1^{h,B}+x_2^{h,B}} [d^h - q]dq - c(x_1^{h,B})^2 - c(x_2^{h,B})^2 \right] \\
& + (1-\mu_0) \cdot \left[ \int_0^{x_1^{l,B}+x_2^{l,B}} [d^l - q]dq - c(x_1^{h,B})^2 - c(x_2^{h,B})^2 \right].
\end{aligned} \tag{2}
$$

Finally, maximization of (2) with respect to $\alpha$ generates the optimal benchmark ratio, $\alpha_B$, the extent of nationalization:

$$
\alpha_B = \frac{7c\mu_0 d^l - 7c\mu_0 d^h + \mu_0 (d^h)^2 - \mu_0 (d^l)^2 + 10c^2 - 7cd^l + (d^l)^2}{6c\mu_0 d^l - 6c\mu_0 d^h + \mu_0 (d^h)^2 - \mu_0 (d^l)^2 + 8c^2 - 6cd^l + (d^l)^2}. \tag{3}
$$

As we will see, the addition of asymmetric information can reverse the relative size of the outputs of the public and private firm and so dramatically change the optimal extent of nationalization.

## 4 *Asymmetric Information: Pooling Equilibrium*

### *4.1 Equilibrium*

We now imagine that the public firm does not observe the realized demand. In this case, there are two possible equilibria: a pooling equilibrium or a separating equilibrium. In a separating equilibrium, the information is fully revealed, and thus the equilibrium remains identical to that in the symmetric information case solved above. As a consequence, we focus on the pooling equilibrium in which the public firm holds belief $\mu_0$ in all periods.

We detail the pooling equilibrium in this section and then compare it with the benchmark case in the next section. If a pooling equilibrium exists, the price remains unchanged by whether the true demand is high or low. In such an equilibrium, firm 2 has no incentive to mimic high demand if demand is low. If the demand is low, but firm 1 (the public firm) believes that it is high, firm 1 increases its production, thus lowering the price and squeezing firm 2's production. This reduces firm 2's profit. Therefore, when demand is low, given that firm 1's production is $x_{1t}$, firm 2 chooses the optimal quantity satisfying

$$
\frac{\partial \pi_{2t}}{\partial x_{2t}^{l,P}} = -x_{2t}^{l,P} + d^l - (x_{1t} + x_{2t}^{l,P}) - 2cx_{2t}^{l,P} = 0,
$$

where the superscript $P$ indicates *pooling*. The above expression generates firm 2's best response $x_{2t}^{l,P}(x_{1t})$. Therefore, when demand is low, given its own quantity $x_{1t}$,

[5] More generally, we let $W = (1-\delta)E_0[\sum_{t=0}^{\infty} w_t]$ be the discounted sum of welfare. But note that the firms have the same actions in all periods, so $W = E_0 w_t$.

firm 1 will observe the price that satisfies

$$p^l(x_{1t}) = d^l - x_{1t} - x_{2t}^{l,P}(x_{1t}).$$

Based on these observations, we imagine that the off-equilibrium belief is that firm 1 believes that $\mu_t = 1$ in the next period if $p_t \neq p^l(x_{1t})$.

When demand is low, firm 2 has no incentive to deviate. When demand is high, in a pooling equilibrium, firm 2 hides the true demand level. Given firm 1's strategy and belief, the one-shot deviation that provides the best payoff for firm 2 is to produce

$$\tilde{x}_{2t} \in \arg\max_{x_{2t}} \pi_{2t}^h(x_{1t}^P, x_{2t}),$$

where $\pi_2^h(x_{1t}^P, x_{2t}) = (d^h - x_{1t}^P - x_{2t})x_{2t} - c(x_{2t})^2$ in the current period. The FOC implies

$$\tilde{p}_t - \tilde{x}_{2t} - 2c\tilde{x}_{2t} = 0,$$

where $\tilde{p}_t = d^h - x_{1t}^P - \tilde{x}_{2t}$. From the next period, firm 1 will hold belief $\mu_t = 1$ and the two firms have a one-shot Nash equilibrium in all the following periods – i.e., firm 1 produces $x_1^{h,B}$ and firm 2 produces $x_2^{h,B}$, resulting in price $p^{h,B} = d^h - x_1^{h,B} - x_2^{h,B}$. Specifically, the incentive compatibility (IC) condition is

$$x_{2t}^{h,P} \cdot p_t^P - c(x_{2t}^{h,P})^2 \geq (1-\delta)\big[\tilde{x}_{2t} \cdot \tilde{p}_t - c\tilde{x}_{2t}^2)\big] + \delta\big[x_2^{h,B} \cdot p^{h,B} - c(x_2^{h,B})^2\big], \tag{4}$$

where $\tilde{p}_t = d^h - x_{1t}^P - \tilde{x}_{2t}$. For firm 2, the left-hand side of (4) is the profit from pooling, and the right-hand side is the profit after taking the one-shot deviation. When (4) is satisfied, there will be no profitable one-shot deviation from the pooling equilibrium when demand is high.

In addition to the IC condition, a pooling equilibrium also requires

$$\left.\frac{\partial E u_{1t}}{\partial x_{1t}}\right|_{\left(x_{1t}^P, x_{2t}^{h,P}, x_{2t}^{l,P}, p_t^P\right)} = -(1-\alpha)x_{1t}^P + p_t^P - 2cx_{1t}^P = 0, \tag{5}$$

$$\left.\frac{\partial \pi_{2t}^l}{\partial x_{2t}}\right|_{\left(x_{1t}^P, x_{2t}^{h,P}, x_{2t}^{l,P}, p_t^P\right)} = -x_{2t}^{l,P} + (d^l - x_{1t}^P - x_{2t}^{l,P}) - 2cx_{2t}^{l,P} = 0, \tag{6}$$

and

$$p_t^P = d^h - x_{1t}^P - x_{2t}^{h,P} = d^l - x_{1t}^P - x_{2t}^{l,P}, \tag{7}$$

where equation (5) is from firm 1's profit-maximization problem; (6) is from firm 2's maximization with low demand; and (7) guarantees an identical price at different demand levels. In sum, a pooling equilibrium exists if

$\{x_{1t}^P, x_{2t}^{h,P}, x_{2t}^{l,P}, p_t^P\}_{t=0,\ldots,\infty}$ such that (4), (5), (6), and (7) are satisfied.[6] Since, in a pooling equilibrium, firms take the same actions in all periods, we omit the subscript $t$ in the following equilibrium expressions:

$$x_1^P = \frac{(1+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2}, \tag{8}$$

$$x_2^{l,P} = \frac{(1-\alpha+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2}, \tag{9}$$

$$x_2^{h,P} = (d^h-d^l)+\frac{(1-\alpha+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2}, \tag{10}$$

$$p^P = \frac{(1+2c)(1-\alpha+2c)d^l}{3-2\alpha+8c-2c\alpha+4c^2}. \tag{11}$$

A pooling equilibrium exists when $(x_1^P, x_2^{l,P}, x_2^{h,P}, p^P)$ satisfies the IC condition (4). For low demand and for any degree of nationalization $\alpha > 0$, the output of the private firm remains below that of its public rival. However, for high demand, the private firm increases its output so as to mislead the public firm. When there is a large difference between $d^h$ and $d^l$, or when there are high degrees of convexity, this increase in output can be so large that the private firm produces even more than public firm. To see this, note that from (8) and (11), $x_2^{h,P} > x_1^P$ if

$$d^h > \left(1+\frac{\alpha}{3-2\alpha+8c-2c\alpha+4c^2}\right)d^l.$$

This condition will hold either when $d^h$ is sufficiently higher than $d^l$, or when $c$ is sufficiently large (such that $\alpha/(3-2\alpha+8c-2c\alpha+4c^2)$ is small). When $x_2^{h,P} > x_1^P$, i.e., when the private firm produces more than the public firm, an increase in nationalization will increase the public firm's output, reduce the private firm's output, and potentially lower total costs. This can reverse the traditional logic for partial privatization of the public firm.

### *4.2 The Optimal Privatization Level*

In this subsection we present and describe the optimal privatization ratio in the pooling equilibrium. We demonstrate that full privatization is not optimal in terms of social welfare. Critically, and in contrast to Matsumura (1998), full nationalization can be desirable for society under certain conditions.

---

[6] Note that in our settings, the FOCs are also sufficient conditions to solve the maximization problems.

Now, let us consider the optimal nationalization level. Plug equilibrium quantities satisfying (8)–(11) into (1), and we have the ex ante welfare

$$\begin{aligned} Ew^P(\alpha) = \mu_0 \cdot \left[ \int_0^{x_1^P + x_2^{h,P}} (d^h - q)dq - c_1(x_1^P) - c_2(x_2^{h,P}) \right] \\ + (1-\mu_0) \cdot \left[ \int_0^{x_1^P + x_2^{l,P}} (d^l - q)dq - c_1(x_1^P) - c_2(x_2^{l,P}) \right]. \end{aligned} \tag{12}$$

The welfare maximization with respect to $\alpha$ generates the optimal nationalization ratio in a pooling equilibrium, denoted as $\alpha_P$.

Proposition 1 $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=0} > 0$.

Proof

$$\left.\frac{\partial Ew^P(\alpha)}{\partial\alpha}\right|_{\alpha=0} = \frac{2c\mu_0 d^l(2c+3)(d^h-d^l)+(2c+1)(d^l)^2}{(2c+3)^3(1+2c)} > 0.$$

*Q.E.D.*

Proposition 1 demonstrates that full privatization is never optimal. When the two firms share an identical cost structure, it remains optimal to have at least some extent of public ownership in order to increase output and so welfare.

Proposition 2

$$\left.\frac{\partial Ew^P(\alpha)}{\partial\alpha}\right|_{\alpha=1} \begin{cases} > 0 & \textit{if } 1+6c+4c^2 > \dfrac{d^l}{\mu_0(d^h-d^l)}, \\ < 0 & \textit{otherwise}. \end{cases}$$

Proof

$$\left.\frac{\partial Ew^P(\alpha)}{\partial\alpha}\right|_{\alpha=1} = \frac{2c(1+2c)d^l((4\mu_0c^2+6c\mu_0+\mu_0)(d^h-d^l)-d^l)}{(1+6c+4c^2)^3}.$$

It is straightforward to see that $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=1} > 0$ if $1+6c+4c^2 > d^l/[\mu_0(d^h-d^l)]$, and vice versa. *Q.E.D.*

In contrast to Matsumura (1998), Proposition 2 indicates that full nationalization may be optimal for society. In particular, we have the optimal nationalization ratio at the pooling equilibrium as follows:

$$\alpha_P = \begin{cases} \dfrac{8c^3\mu_0d^h-8c^3\mu_0d^l+16c^2\mu_0d^h+4c^2d^l-16c^2\mu_0d^l-6c\mu_0d^l+6c\mu_0d^h+4cd^l+d^l}{-4c^2\mu_0d^l+4c^2\mu_0d^h+4c^2d^l+6cd^l-4c\mu_0d^l+4c\mu_0d^h+d^l} & \text{if } 1+6c+4c^2 \le \dfrac{d^l}{\mu_0(d^h-d^l)}, \\ 1 & \text{if } 1+6c+4c^2 > \dfrac{d^l}{\mu_0(d^h-d^l)}. \end{cases} \tag{13}$$

With a high convexity in costs, the output of the public firm is low, but the private firm still produces large quantities in order to hide information when demand is high. When the public firm is fully nationalized, output moves from the private firm to the public firm, and is more equally distributed between the two firms. Thus, instead of gaining consumer surplus at high total costs due to less-equally distributed production, the gain in consumer surplus comes with lower production costs.

PROPOSITION 3 $\partial\alpha_P/\partial\mu_0 > 0$ *if* $1+6c+4c^2 < d^l/[\mu_0(d^h-d^l)]$.

PROOF

$$\frac{\partial\alpha_P}{\partial\mu_0} = \frac{2c(1+2c)(8c^3+20c^2+14c-1)d^l(d^h-d^l)}{(-4c^2\mu_0 d^l+4c^2\mu_0 d^h+4c^2 d^l+6cd^l-4c\mu_0 d^l+4c\mu_0 d^h+d^l)^2} > 0.$$

*Q.E.D.*

Proposition 3 shows that a larger initial probability of high demand is associated with a smaller extent of privatization. When demand is high, firm 2 produces more output than when demand is low, but firm 1 produces the same output in both scenarios. Therefore, when demand is high, the production costs are distributed less evenly, in the sense that firm 2 produces too much, but firm 1 produces too little compared with the optimal production arrangement. As a result, when demand is high, a greater $\alpha$ (smaller privatization) is required for optimality. When $\mu_0$ increases, more weight is put on welfare with high demand. As a consequence, the optimal $\alpha$ increases in $\mu_0$.

## 5 *Comparison with the Benchmark*

In this section, we compare the pooling equilibrium in the asymmetric-information case with that in the benchmark. The comparison shows that the optimal privatization ratio in the asymmetric-information case may be either higher or lower than that in the benchmark.

When cost convexity is low, the optimal extent of privatization is larger under the asymmetric-information case. When cost convexity is high, the opposite is the case.

PROPOSITION 4 $\alpha_B > \alpha_P$ *when* $c$ *is sufficiently small, but* $\alpha_B < \alpha_P$ *when* $c$ *is sufficiently large.*

PROOF When $c$ is sufficiently small, $\partial Ew^B(\alpha)/\partial\alpha|_{\alpha=\alpha_P} > 0$ and the concavity of the function $Ew^B(\alpha)$ directly implies that $\alpha_B > \alpha_P$. Proposition 2 shows that $\alpha_P = 1$ when $c$ is large enough and full nationalization is never optimal in the benchmark case (Matsumura, 1998); thus $\alpha_B < \alpha_P$ for a high enough $c$. *Q.E.D.*

Comparing (3) with (13), Proposition 4 shows that $\alpha_P < \alpha_B$ when $c$ is sufficiently small. The intuition follows from recognizing that for any given $\alpha$, the total output in the pooling equilibrium is less than the socially optimal output level, but strictly greater than that in the benchmark case. This reduces the deadweight loss resulting from oligopolistic power, and thus increases welfare. In addition, in contrast to the symmetric-information case, firm 1 produces less, whereas firm 2 produces much more, in the pooling equilibrium, which is associated with high production cost. When the cost convexity is low, that gain in consumer surplus outweighs the increase of the total production cost. When convexity is high, the increase in total production cost outweighs the increase in consumer surplus. The comparison of the extent of privatization follows.

*Figure 2*
The Optimal $\alpha$ ($d^h = 10$, $d^l = 9$, $\mu_0 = 0.6$)

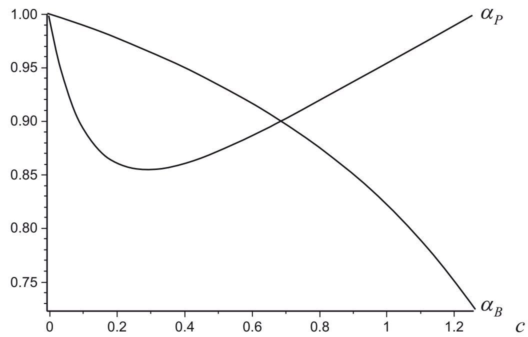



Figure 2 illustrates the comparison of $\alpha_B$ and $\alpha_P$ for a specific set of demand and probability values. It clearly portrays Proposition 4, with the nationalization parameter for the asymmetric information case starting below that of the benchmark but eventually crossing above. It also shows the nationalization parameter for the asymmetric case reaching 1.0 for sufficient cost convexity. At and beyond this point, there is no incentive for the government to partially privatize, as that would increase the output of the private firm, making output more asymmetric and costly.

## 6 The Model with General Cost Functions

In this section, we replace the quadratic costs in the previous sections with general functions $c_1(x)$ and $c_2(x)$ to show that similar results can hold. In addition to the assumption that both firms produce positive output, we assume $c_1(x)$ and $c_2(x)$ satisfy the following assumption.

ASSUMPTION $c_1(x)$ *and* $c_2(x)$ *are twice differentiable for any* $x > 0$, *with* $c_1''(x) > 0$ *and* $c_2''(x) > 0$ *for any* $x \in \mathbb{R}^+$.

We will keep this assumption in the rest of the paper.

### 6.1 Equilibrium Conditions

Similarly to the discussion with quadratic costs, with symmetric information, the two firms play one-shot Nash in each period. The equilibrium quantities and prices are determined by the following first-order conditions (FOCs) for two firms when demand is $r$ $(r = h, l)$:

$$\frac{\partial E u_1}{\partial x_1^r} = -(1-\alpha)\cdot x_1^r + [d^r - (x_1^r + x_2^r)] - c_1'(x_1^r) = 0, \tag{14}$$

$$\frac{\partial \pi_2}{\partial x_2^r} = -x_2^r + [d^r - (x_1^r + x_2^r)] - c_2'(x_2^r) = 0. \tag{15}$$

Simultaneously solving the above equations generates the equilibrium quantities $(x_1^{r,B}, x_2^{r,B})$ and equilibrium prices $p^{r,B}$, with $p^{r,B} = d^r - (x_1^{r,B} + x_2^{r,B})$.

With asymmetric information, as discussed in section 4, if a pooling equilibrium exists, price remains unchanged whether the true demand is high or low. In such an equilibrium, firm 2 has no incentive to mimic high demand if demand is low. If the demand is low, but firm 1 (the public firm) believes that demand is high, firm 1 increases its production, thus lowering the price and squeezing firm 2's production. This reduces firm 2's profit. Therefore, when demand is low, given that firm 1's production is $x_1$, firm 2 chooses the optimal quantity satisfying

$$\frac{\partial \pi_2}{\partial x_2^{l,P}} = -x_2^{l,P} + d^l - (x_1 + x_2^{l,P}) - c_2'(x_2^{l,P}) = 0,$$

where the superscript $P$ indicates *pooling*. The above expression generates firm 2's best response $x_2^{l,P}(x_1)$. Therefore, when demand is low, given its own quantity $x_1$, firm 1 will observe the price that satisfies

$$p^l(x_1) = d^l - x_1 - x_2^{l,P}(x_1).$$

Based on these observations, we imagine that the off-equilibrium belief is that firm 1 believes that $\mu_t = 1$ in the next period if $p_t \neq p^l(x_1)$.

When demand is low, firm 2 has no incentive to deviate. When demand is high, in a pooling equilibrium, firm 2 hides the true demand level. Given firm 1's strategy and belief, the one-shot deviation that provides the best payoff for firm 2 is to produce

$$\tilde{x}_2 \in \arg\max_{x_2} \pi_2^h(x_1^P, x_2),$$

where $\pi_2^h(x_1^P, x_2) = (d^h - x_1^P - x_2)x_2 - c_2(x_2)$, in the current period. The FOC implies

$$\tilde{p} - \tilde{x}_2 - c_2'(\tilde{x}_2) = 0,$$

where $\tilde{p} = d^h - x_1^P - \tilde{x}_2$. From the next period, firm 1 will hold belief $\mu_t = 1$, the two firms have a one-shot Nash equilibrium in all the following periods – i.e., firm 1 produces $x_1^{h,B}$ and firm 2 produces $x_2^{h,B}$, resulting in price $p^{h,B} = d^h - x_1^{h,B} - x_2^{h,B}$. Hence, the IC condition is

$$x_2^{h,P} \cdot p^P - c_2(x_2^{h,P}) \geq (1-\delta)[\tilde{x}_2 \cdot \tilde{p} - c_2(\tilde{x}_2)] + \delta[x_2^{h,B} \cdot p^{h,B} - c_2(x_2^{h,B})], \tag{16}$$

where $\tilde{p} = d^h - x_1^P - \tilde{x}_2$. For firm 2, the left-hand side of (16) is the profit from pooling, and the right-hand side is the profit after taking the one-shot deviation. When (16) is satisfied, there will be no profitable one-shot deviation from the pooling equilibrium when demand is high.

In addition to the IC condition, a pooling equilibrium also requires

$$\left.\frac{\partial Eu_1}{\partial x_1}\right|_{\left(x_1^P, x_2^{h,P}, x_2^{l,P}, p^P\right)} = -(1-\alpha)x_1^P + p^P - c_1'(x_1^P) = 0, \tag{17}$$

$$\left.\frac{\partial \pi_2^l}{\partial x_2}\right|_{\left(x_1^P, x_2^{h,P}, x_2^{l,P}, p^P\right)} = -x_2^{l,P} + (d^l - x_1^P - x_2^{l,P}) - c_2'(x_2^{l,P}) = 0, \tag{18}$$

and

$$p^P = d^h - x_1^P - x_2^{h,P} = d^l - x_1^P - x_2^{l,P}. \tag{19}$$

Any $(x_1^P, x_2^{h,P}, x_2^{l,P}, p^P)$ satisfying (16) to (19) constitutes a pooling equilibrium.

### *6.2 Existence of Pooling Equilibrium*

The following proposition gives a sufficient condition for the existence of a pooling equilibrium. We provide all proofs of lemmas and propositions for this section in the appendix.

PROPOSITION 5 *Suppose that the Assumption is satisfied. Then, given any* $\delta < 1$*, there exists* $\bar{d}(\delta) > d^l$ *such that a pooling equilibrium exists for any* $\alpha \in [0,1]$ *if* $d^h \in (d^l, \bar{d}(\delta))$.

The intuition of Proposition 5 is as follows. To successfully hide information, firm 2 produces more than the Cournot quantity. When $d^h$ is far away from $d^l$, the opportunity cost of hiding the true demand grows. As a consequence, $d^h$ should not be too large relative to $d^l$ to sustain the pooling equilibrium.

Note that Proposition 5 implies that, with *any cost functions* satisfying the Assumption, we can always find $\bar{d}(\delta) > d^l$ such that a pooling equilibrium exists for any $\alpha \in [0,1]$ if $d^h \in (d^l, \bar{d}(\delta))$. The existence of the pooling equilibrium lies in the fact that, with *any cost function*, we can find a set of $d^h$ such that the profit of the private firm in the pooling equilibrium – i.e., $x_2^{h,P} \cdot p^P - c_2(x_2^{h,P})$ – exceeds that in a Cournot equilibrium – i.e., $x_2^{h,B} \cdot p^{h,B} - c_2(x_2^{h,B})$. This further implies that it is better for the private firm to hide the demand information. The detailed proof is shown in the appendix.

### *6.3 The Optimal Privatization Level*

The next two propositions summarize the optimal privatization ratio in a pooling equilibrium. We can again demonstrate that full privatization is not optimal in terms of social welfare. We also show that when the slope of the marginal cost is sufficiently small, partial privatization is more desirable for society.

We first derive Lemma 1, which is useful for understanding most of the results in this paper. Lemma 1 states that when $\alpha$ increases, firm 1 produces more, but firm 2 produces less. As firm 1 cares about welfare, it increases its quantity, which induces its rival (firm 2) to produce less due to strategic substitution.

LEMMA 1 *In a pooling equilibrium,*

$$\frac{\partial x_1^P}{\partial \alpha} > 0, \quad \frac{\partial x_2^{l,P}}{\partial \alpha} = \frac{\partial x_2^{h,P}}{\partial \alpha} < 0, \quad \frac{\partial (x_1^P + x_2^{l,P})}{\partial \alpha} > 0, \quad \textit{and} \quad \frac{\partial (x_1^P + x_2^{h,P})}{\partial \alpha} > 0.$$

Now, let us consider the optimal nationalization level. Plug equilibrium quantities satisfying (4)–(7) into (1), and we have the ex ante welfare:

$$\begin{aligned} Ew^P(\alpha) = {} & \mu_0 \cdot \left[ \int_0^{x_1^P + x_2^{h,P}} (d^h - q)dq - c_1(x_1^P) - c_2(x_2^{h,P}) \right] \\ & + (1-\mu_0) \cdot \left[ \int_0^{x_1^P + x_2^{l,P}} (d^l - q)dq - c_1(x_1^P) - c_2(x_2^{l,P}) \right]. \end{aligned} \tag{20}$$

The welfare maximization with respect to $\alpha$ generates the optimal nationalization ratio in a pooling equilibrium, denoted as $\alpha_P$.

PROPOSITION 6 $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=0} > 0$ *if* $c_1(\cdot) = c_2(\cdot)$.

Proposition 6 predicts that full privatization is not optimal when two firms share an identical cost structure. In general, this result holds true if firm 1 is strictly more efficient than firm 2, but does not necessarily hold when firm 1 is less efficient than firm 2.

PROPOSITION 7 $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=1} < 0$ *if* $c_2''$ *is sufficiently small;* $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=1} > 0$ *if* $c_2''$ *is sufficiently large and* $\mu_0$ *is sufficiently close to* 1.

Proposition 7 indicates that full nationalization also is not optimal for society when cost is not too convex. In a pooling equilibrium, first note that when $\alpha = 1$, firm 1's FOC requires that $p^P = c_1'(x_1^P)$, and firm 2's FOC requires that $p^P - c_2'(x_2^{l,P}) = x_2^{l,P} > 0$. When $c_2''$ is sufficiently small, $c_2'(x_2^{h,P})$ is close enough to $c_2'(x_2^{l,P})$ that $p^P > c_2'(x_2^{h,P})$. Since $p^P = c_1'(x_1^P)$ – i.e., $\partial Ew^P/\partial x_1 = 0$ – a slight reduction in $x_1$ does not reduce $w^P$. But since $p^P > c_2'(x_2^{h,P})$ and $p^P > c_2'(x_2^{l,P})$, a slight increase in $x_2$ improves $w^P$. From Lemma 1, we know $\partial x_1^P/\partial\alpha > 0$ and $\partial x_2^{r,P}/\partial\alpha < 0$, so a reduction in $\alpha$ will increase welfare. This mimics Matsumura (1998). However, when $c_2''$ is so large that $p^P < c_2'(x_2^{h,P})$, we may have $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=1} > 0$.[7]

[7] However, a large $c_2''$ must guarantee the existence of the pooling equilibrium, and such a large $c_2''$ indeed exists, as shown in the quadratic example.

In other words, firm 2 overproduces at an extremely inefficient level in order to hide the true information about demand, and a slight decrease in $x_2$ improves $w^P$. Following Lemma 1, an increase in $\alpha$ to the full-nationalization value in order to reduce $x_2$ may generate optimal social welfare.

Proposition 8 $\partial\alpha_P/\partial\mu_0 > 0$ *if* $\alpha_P \in \mathrm{int}(\Lambda)$ *and* $\partial^2 Ew^P(\alpha)/\partial\alpha^2|_{\alpha_P} \neq 0$*, where* $\Lambda \subset [0,1]$ *is the set of* $\alpha$ *for which there exists a pooling equilibrium.*

Proposition 8 again shows that when the initial probability of high demand is large, the extent of the privatization is small. The intuition is as follows. When demand is high, firm 2 produces more output than when demand is low, but firm 1 produces the same output in both scenarios. Therefore, when demand is high, the production costs are distributed less evenly, in the sense that firm 2 produces too much, but firm 1 produces too little, than in the optimal production arrangement. As a result, when demand is high, a greater $\alpha$ is required for optimality. When $\mu_0$ increases, more weight is put on welfare with high demand. As a consequence, the optimal $\alpha$ increases in $\mu_0$.

### *6.4 Comparison of the Optimal Privatization Levels*

In this subsection, we compare the optimal nationalization level $\alpha$ in the asymmetric-information case with that in the benchmark. The optimal nationalization $\alpha$ maximizes welfare, which can also be generally written as

$$w(x_1,x_2) = \left[\int_0^{x_1+x_2}[d-q]dq - C(x_1+x_2)\right] - [c_1(x_1)+c_2(x_2)-C(x_1+x_2)],$$

where the function $C(x)$ is defined by $C(x) = \min_{x_1}[c_1(x_1)+c_2(x-x_1)]$. We view $C(x)$ as the least cost to produce a total output level $x$. With this form, it is clear that the welfare consists of two parts: The first term in the expression – i.e., $\Phi_1 = \int_0^{x_1+x_2}[d-q]dq - C(x_1+x_2)$ – is the total surplus if the output were produced in the most efficient way, which depends only on the total output $x_1+x_2$. We call this term, $\Phi_1$, the potential social surplus. The second term – i.e., $\Phi_2 = c_1(x_1)+c_2(x_2) - C(x_1+x_2)$ – is the inefficiency caused when the total output $x$ is not allocated efficiently between the two firms. We call this term, $\Phi_2$, the quantity-allocation inefficiency.

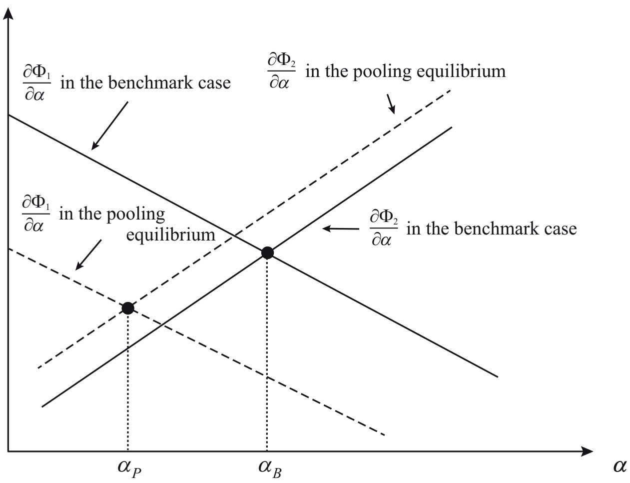
As $\alpha$ increases, according Lemma 1, $x_1$ rises while $x_2$ falls, and the total output $x = x_1+x_2$ rises. Hence, $\alpha$ affects welfare in two ways. First, the total output increases as $\alpha$ increases, and thus the potential social surplus $\Phi_1$ is changed. Second, an increase in $\alpha$ shifts production from the private firm to the public firm, and thus influences the quantity-allocation inefficiency $\Phi_2$. The balance between the marginal gain in social surplus and the marginal loss in quantity-allocation inefficiency determines the optimal $\alpha$.

In order to compare the optimal $\alpha$ in these two cases, we first consider the total outputs in equilibrium, and then the effects on the potential social surplus. The next lemma indicates that, in a pooling equilibrium, when demand is low, firms produce the same quantities as in the symmetric-information case, but when demand is high, the total output level is greater than that in the symmetric-information case. This is because a pooling equilibrium requires the price under high demand to be the same as that under low demand. As a result, when demand is high, the price in a pooling equilibrium is lower than that in the symmetric-information case, leading to larger total quantities.

LEMMA 2 *Given the existence of a pooling equilibrium, we have: (1)* $x_1^P + x_2^{l,P} = x_1^{l,B} + x_2^{l,B}$*; (2)* $x_1^P + x_2^{h,P} > x_1^{h,B} + x_2^{h,B}$.

From Lemma 2, the total output is greater in the pooling equilibrium. In addition, note that by concavity, the marginal gain in potential social surplus decreases as total output increases. This implies that in the pooling equilibrium the marginal gain in potential social surplus is less. When costs are not too convex, the marginal loss due to quantity-allocation inefficiency is bounded. Therefore, the optimal extent of nationalization under the symmetric-information case exceeds that under the asymmetric-information case when marginal costs are sufficiently flat. This can be intuitively shown by Figure 3, and is stated formally by the following proposition, with the proof shown in the appendix.

*Figure 3*
Comparison of the Optimal Privatization Levels when $c_i''$ is Not Too Large




PROPOSITION 9 $\partial Ew^B(\alpha)/\partial\alpha|_{\alpha_P} > 0$ *when* $c_i''$, $i = 1,2$, *is sufficiently small if* $\alpha_P \in \text{int}(\Lambda)$.

When $Ew^B(\alpha)$ is concave, Proposition 9 directly implies that $\alpha_B > \alpha_P$ when $c_i''$ is small. In order to maximize social welfare, the government nationalizes more to increase the output level when information is transparent. As a result, the total effect is that $\alpha_B > \alpha_P$.

When the costs are more convex, however, it is not possible to get general results, because one is not able to see the change of $\Phi_2$ without detailed functional forms of costs. We have not been able to isolate more general conditions under which $\alpha_B < \alpha_P$, but, using quadratic costs as an example, we have shown in section 5 that $\alpha_B > \alpha_P$ when costs are sufficiently flat, and $\alpha_B < \alpha_P$ when costs are sufficiently convex (Proposition 4 and Figure 2).

### 6.5 *Welfare*

We next compare the welfare in a pooling equilibrium with that in the benchmark. At issue is how the information asymmetry affects welfare, and whether the information rent paid by the private firm can possibly enhance social welfare. The following Proposition 10 indicates that social welfare in a pooling equilibrium exceeds that under symmetric information when marginal cost functions are sufficiently flat.

PROPOSITION 10 *For any $\alpha \in [0,1]$, there exists $\varepsilon(\alpha) > 0$ such that if $c_i'' < \varepsilon(\alpha)$ $(i = 1,2)$, then $Ew^P(\alpha) \geq Ew^B(\alpha)$.*

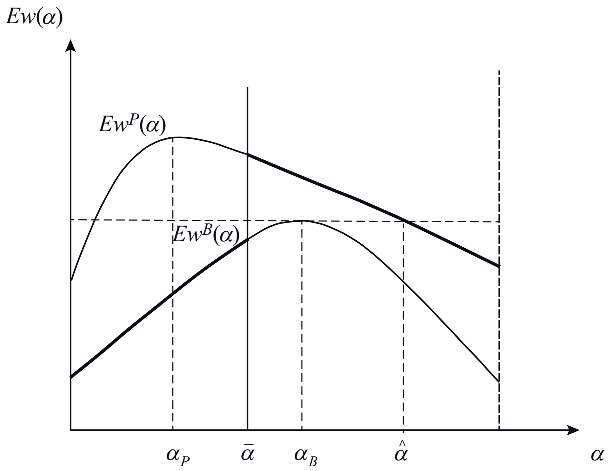
Intuitively, information asymmetry influences welfare by affecting firms' output levels and, further, through its effects on the potential social surplus $\Phi_1$ and the quantity-allocation inefficiency $\Phi_2$. The balance of these two effects determines whether information asymmetry increases or decreases social welfare. Consider first the effects on the potential social surplus. We show that the dead-weight loss is smaller in the asymmetric-information case when costs are not too convex. To see this, first denote the social optimal output level as $\bar{x}^r$ when demand is $r$, $r = h,l$ – i.e., let $\bar{x}^r$ be the $x$ satisfying $d^r - x = C'(x)$, where $C(x) = \min_{x_1}[c_1(x_1) + c_2(x - x_1)]$. Apparently, there is a welfare loss when $x_1 + x_2 \neq \bar{x}^r$; and the more the difference between $x_1 + x_2$ and $\bar{x}^r$, the more dead-weight loss is incurred. When demand is low, there is no difference in the dead-weight loss between these two cases, since $x_1^P + x_2^{l,P} = x_1^{l,B} + x_2^{l,B}$ (by Lemma 2). When demand is high, by Lemma 2, $x_1^{h,B} + x_2^{h,B} < x_1^P + x_2^{h,P}$; and, further, $x_1^{h,B} + x_2^{h,B} < x_1^P + x_2^{h,P} < \bar{x}^h$ when costs are not too convex. This implies that the dead-weight loss caused by the oligopolistic power is less, as firms produce more in the pooling equilibrium.

In addition, there is quantity-allocation inefficiency when production is not optimally distributed between two firms – that is, when $c_1(x_1) + c_2(x_2) > C(x_1 + x_2)$. When $c_i''$ is small, the difference of welfare loss from quantity-allocation inefficiency between the symmetric-information case and the asymmetric-information case is bounded. As a result, the quantity-allocation inefficiency is limited and therefore dominated by the effect of restriction on market power. As a consequence, the welfare in the pooling equilibrium is greater than that with symmetric information.

## 7 Information Structure and Partial Privatization

In the model we have assumed that only the private firm observes the demand. Yet, in choosing the extent of privatization, the government may also be setting the sensitivity of the public firm in detecting market demand. Thus, one advantage of increased privatization could be a management that is more informed about demand. We now consider a case in which with sufficient privatization the information asymmetry disappears. We imagine that there exists a threshold $\bar{\alpha}$, such that the public firm has equal access to the information about market demand when $\alpha < \bar{\alpha}$. Therefore, if the government chooses any partial privatization smaller than $\bar{\alpha}$, firms play a game with symmetric information and achieve the associated welfare. In the range of $\alpha \geq \bar{\alpha}$, we suppose that the pooling equilibrium appears. We have adopted a scenario with low convexity such that the optimal extent of nationalization under the pooling equilibrium is less than that in the benchmark, and $Ew^P(\alpha) \geq Ew^B(\alpha)$ holds for all $\alpha \in [0,1]$. This scenario is shown in Figure 4.

*Figure 4*
Welfare with a Threshold $\bar{\alpha}$

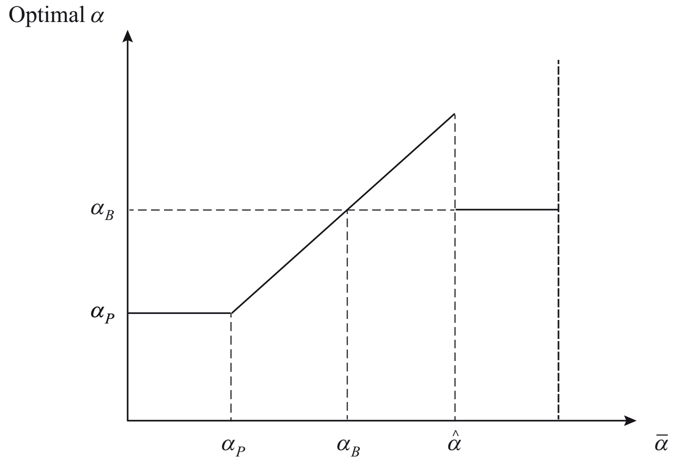



Suppose that both $Ew^B(\alpha)$ and $Ew^P(\alpha)$ are concave. From Figure 4, for any given $\bar{\alpha} \in [0,1]$, the welfare-maximizing government's optimal partial nationalization choice will be

$$\alpha^* = \begin{cases} \alpha^P & \text{if } \bar{\alpha} < \alpha^P, \\ \bar{\alpha} & \text{if } \alpha^P < \bar{\alpha} < \hat{\alpha}, \\ \alpha^B & \text{if } \hat{\alpha} < \bar{\alpha}, \end{cases}$$

where $\hat{\alpha} := \{\alpha \in [0,1] | Ew^B(\alpha) = Ew^P(\alpha)\}$. When the threshold for the public firm to access the information is high (deep privatization – i.e., small $\bar{\alpha}$), the government will choose an optimal partial privatization to obtain pooling equilibrium. When such a threshold becomes sufficiently low (deep nationalization – i.e., a large $\bar{\alpha}$ such that $\hat{\alpha} < \bar{\alpha}$), the government will choose a low nationalization level such that the public firm can get full information. See Figure 5, which illustrates the different degrees of privatization as the threshold varies.

*Figure 5*
Optimal $\alpha$ with a Threshold $\bar{\alpha}$




The government builds the realization of the symmetric information into its optimal choice of partial privatization, but the potential importance of the pooling equilibrium remains. In other words, there is no expectation that the government will always simply adopt the extent of privatization associated with symmetric information.

## 8 *Conclusion*

We have investigated a duopoly with a partially privatized public firm under asymmetric information and have identified the conditions for a pooling equilibrium. We confirm the logic that as the government firm places less emphasis on welfare (becomes increasingly privatized), it produces less and the private firm produces more. Yet, the standard logic that some privatization will reduce production costs does not always apply. Under quadratic costs the pooling equilibrium can result in very different partial privatization choices by the government. For low convexity, the fact that the government firm has much greater production is sufficient to encourage greater privatization. For high convexity, the production of the private firm can be greater than that of the government firm, and the extent of the optimal privatization is smaller under asymmetric information. Indeed, allowing no privatization can be optimal, as it serves to equalize the production between the two firms even as it increases output. Many, but not all, of these demonstrations carry over to the case of generalized cost functions. Importantly, we show that consumer surplus in a pooling equilibrium is greater than that without information asymmetry, and a pooling equilibrium is often associated with higher social welfare.

There remain a variety of directions for further research. Among the most obvious would be to generalize the number of private firms. Our efforts at this generalization created a degree of complexity that defied easy analytical solution, yet we recognize it as an important robustness check. If one assumes that all private firms have complete information, a key difference from the duopoly case would be the nature of coordination that might generate a pooling equilibrium. When there is only one private firm, it decides if the benefit of hiding the market information is worth the cost of producing more output and reducing the price. How such costs might be borne among multiple private firms deserves careful thought. Yet, it seems likely that given a pooling equilibrium, additional and even full privatization could remain sensible. The additional production taken on by the private firms continues to create an incentive for the government to have the mixed firm increase its production by having a larger public share. This logic would not change. Beyond the generalization to more private firms, it would be interesting to examine the effect of the public firm's learning process – that is, when the public firm can observe other signals reflecting the true demand level – on the optimal privatization ratio. Moreover, quantities are sometimes more easily observed than prices in certain markets, so future research might use quantity rather than price as the information signal. We leave all three of these extensions to future work.

## *Appendix: Proofs for the Lemmas and Propositions with General Cost Functions*

### *A.1 Proof of Proposition 5*

A pooling equilibrium exists if we can find $(x_1^P,\ x_2^{h,P},\ x_2^{l,P},\ p^P)$ satisfying the conditions (16)–(19). Given a set of primitives $(d^h, d^l, c_1(\cdot), c_2(\cdot))$, it is straightforward to solve (17)–(19) to get the values of $(x_1^P,\ x_2^{h,P},\ x_2^{l,P},\ p^P)$. The remaining problem is whether those values of $(x_1^P, x_2^{h,P}, x_2^{l,P},\ p^P)$ satisfy the IC condition (16).

Note that the IC condition can be written as

$$\begin{aligned} &\left[x_2^{h,P}\cdot p^P - c_2(x_2^{h,P})\right] - \left[x_2^{h,B}\cdot p^{h,B} - c_2(x_2^{h,B})\right] \\ &\geq (1-\delta)\left\{\left[\tilde{x}_2\cdot\tilde{p} - c_2(\tilde{x}_2)\right] - \left[x_2^{h,B}\cdot p^{h,B} - c_2(x_2^{h,B})\right]\right\}. \end{aligned} \tag{A1}$$

We denote the left-hand side of the inequality as $\Delta$, i.e.,

$$\Delta = \left[x_2^{h,P}\cdot p^P - c_2(x_2^{h,P})\right] - \left[x_2^{h,B}\cdot p^{h,B} - c_2(x_2^{h,B})\right].$$

Since $(x_2^{h,P},\ p^P,\ p^{h,B}, x_2^{h,B})$ depend on $d^h$, we regard $\Delta$ as a function of $d^h$ and write $\Delta$ as $\Delta(d^h)$. Note that $\Delta(d^h) = 0$ when $d^h = d^l$, since $x_1^{l,B} = x_1^P = x_1^{h,B}$, and $x_2^{h,B} = x_2^{h,P} = x_2^{l,B} = x_2^{l,P}$ when $d^h = d^l$.

Similarly, we denote the right-hand side of (A1) as $(1-\delta)\Psi(d^h)$, i.e.,

$$\Psi(d^h) = \left[\tilde{x}_2 \cdot \tilde{p} - c_2(\tilde{x}_2)\right] - \left[x_2^{h,B} \cdot p^{h,B} - c_2(x_2^{h,B})\right].$$

Also note that $\Psi(d^h) = 0$ when $d^h = d^l$, and thus $\Delta(d^h) - (1-\delta)\Psi(d^h) = 0$ when $d^h = d^l$. The IC condition then can be written as $\Delta(d^h) - (1-\delta)\Psi(d^h) > 0$.

We next show that (i) there exists $\eta > 0$ such that $\partial\Delta/\partial d^h|_{d^h=d^l} > \eta$ uniformly for all $\alpha \in [0,1]$; and (ii) $\partial\Psi/\partial d^h|_{d^h=d^l} = \partial\Delta/\partial d^h|_{d^h=d^l}$ for all $\alpha \in [0,1]$. If both (i) and (ii) are true, then given any $\delta < 1$, we can find $\zeta(\delta) = \eta\delta > 0$ such that

$$\left.\frac{\partial(\Delta - (1-\delta)\Psi)}{\partial d^h}\right|_{d^h=d^l} = \delta \cdot \left.\frac{\partial\Delta}{\partial d^h}\right|_{d^h=d^l} > \zeta(\delta)$$

uniformly for all $\alpha \in [0,1]$. This implies that given any $\delta < 1$, there exists $\bar{d}(\delta) > d^l$ such that when $d^h < \bar{d}(\delta)$, one has $\Delta - (1-\delta)\Psi > 0$ uniformly for any $\alpha \in [0,1]$, that is, the IC condition (A1) is satisfied when $d^h < \bar{d}(\delta)$.

Take the derivative of $\Delta$ with respect to $d^h$. Since $\partial x_2^{h,p}/\partial d^h = 1$ and $\partial p^P/\partial d^h = 0$, we can get

$$\begin{aligned}
\frac{\partial\Delta}{\partial d^h} &= p^P - c_2'(x_2^{h,P}) - \left[1 - \frac{\partial x_1^{h,B}}{\partial d^h} - \frac{\partial x_2^{h,B}}{\partial d^h}\right] \cdot x_2^{h,B} - p^{h,B} \cdot \frac{\partial x_2^{h,B}}{\partial d^h} + c_2'(x_2^{h,B}) \cdot \frac{\partial x_2^{h,B}}{\partial d^h} \\
&= p^P - c_2'(x_2^{h,P}) - \left[1 - \frac{\partial x_1^{h,B}}{\partial d^h}\right] \cdot x_2^{h,B} - \frac{\partial x_2^{h,B}}{\partial d^h} \cdot \left[p^{h,B} - c_2'(x_2^{h,B}) - x_2^{h,B}\right].
\end{aligned}$$

From the FOC of firm 2, $p^{h,B} - x_2^{h,B} - c_2'(x_2^{h,B}) = 0$. Hence,

$$\frac{\partial\Delta}{\partial d^h} = p^P - c_2'(x_2^{h,P}) - \left[1 - \frac{\partial x_1^{h,B}}{\partial d^h}\right] \cdot x_2^{h,B}.$$

When $d^h = d^l$, we have $x_1^{l,B} = x_1^P = x_1^{h,B}$, $x_2^{h,B} = x_2^{h,P} = x_2^{l,B} = x_2^{l,P}$, and $p^P = p^{l,B} = p^{h,B}$. Thus,

$$\left.\frac{\partial\Delta}{\partial d^h}\right|_{d^h=d^l} = \left[p^{l,B} - x_2^{l,B} - c_2'(x_2^{l,B})\right] + \left.\frac{\partial x_1^{h,B}}{\partial d^h}\right|_{d^h=d^l} \cdot x_2^{l,B}.$$

From the FOC of firm 2, $p^{l,B} - x_2^{l,B} - c_2'(x_2^{l,B}) = 0$; therefore,

$$\left.\frac{\partial\Delta}{\partial d^h}\right|_{d^h=d^l} = \left.\frac{\partial x_1^{h,B}}{\partial d^h}\right|_{d^h=d^l} \cdot x_2^{l,B}.$$

Now, consider $\partial x_1^{h,B}/\partial d^h|_{d^h=d^l}$. From the FOCs of the two firms' maximization problems, we can get

$$\begin{bmatrix} \frac{\partial x_1^{h,B}}{\partial d^h} \\ \frac{\partial x_2^{h,B}}{\partial d^h} \end{bmatrix} = \frac{1}{\left(2-\alpha+c_1''(x_1^{h,B})\right)\left(2+c_2''(x_2^{h,B})\right)-1} \cdot \begin{bmatrix} 1+c_2''(x_2^{h,B}) \\ 1-\alpha+c_1''(x_1^{h,B}) \end{bmatrix}.$$

Note that $c_i''$ is bounded from above, since $x_i \in [0, d^h]$ in any equilibrium. Hence, there exists $\varepsilon > 0$ such that $\partial x_1^{h,B}/\partial d^h|_{d^h=d^l} > \varepsilon$ with any $\alpha \in [0,1]$.

Next, consider $x_2^{l,B}$. Note that $x_2^{l,B}$ decreases in $\alpha$. When $\alpha = 1$, from the FOCs of the two firms' maximization problems, $p^{l,B} = c_1'(x_1^{l,B})$ and $x_2^{l,B} = p^{l,B} - c_2'(x_2^{l,B})$. Note that $x_2^{l,B} > 0$. Let $\xi$ be the value of $x_2^{l,B}$ when $\alpha = 1$, and let $\eta = \varepsilon\xi$; as a result, we have $\partial\Delta/\partial d^h|_{d^h=d^l} > \eta > 0$ for any $\alpha \in [0,1]$.

We complete the proof by showing that $\partial\Psi/\partial d^h|_{d^h=d^l} = \partial\Delta/\partial d^h|_{d^h=d^l}$. A straightforward calculation shows that

$$\frac{\partial\Psi}{\partial d^h} = \left\{\left[1 - \frac{\partial x_1^P}{\partial d^h}\right]\cdot\tilde{x}_2 + \frac{\partial\tilde{x}_2}{\partial d^h}\cdot[\tilde{p} - c_2'(\tilde{x}_2) - \tilde{x}_2]\right\}$$
$$- \left\{\left[1 - \frac{\partial x_1^{h,B}}{\partial d^h}\right]\cdot x_2^{h,B} + \frac{\partial x_2^{h,B}}{\partial d^h}\cdot\left[p^{h,B} - c_2'(x_2^{h,B}) - x_2^{h,B}\right]\right\}.$$

Since $\tilde{p} - c_2'(\tilde{x}_2) - \tilde{x}_2 = 0$, $p^{h,B} - c_2'(x_2^{h,B}) - x_2^{h,B} = 0$, and $\partial x_1^P/\partial d^h = 0$,

$$\frac{\partial\Psi}{\partial d^h} = \tilde{x}_2 - \left[1 - \frac{\partial x_1^{h,B}}{\partial d^h}\right]\cdot x_2^{h,B}.$$

When $d^h = d^l$, we have $x_1^{l,B} = x_1^P = x_1^{h,B}$, $x_2^{h,B} = x_2^{h,P} = x_2^{l,B} = x_2^{l,P} = \tilde{x}_2$, and $p^P = p^{l,B} = p^{h,B}$. Thus,

$$\frac{\partial\Psi}{\partial d^h} = \left.\frac{\partial x_1^{h,B}}{\partial d^h}\right|_{d^h=d^l}\cdot x_2^{l,B} = \left.\frac{\partial\Delta}{\partial d^h}\right|_{d^h=d^l}.$$

*Q.E.D.*

*A.2 Proof of Lemma 1*

Take derivatives of equilibrium quantities with respect to $\alpha$, we have

(A2) $$\frac{\partial x_1^P}{\partial\alpha} = \frac{x_1^P\left(2 + c_2''(x_2^{l,P})\right)}{\left(2 - \alpha + c_1''(x_1^P)\right)\left(2 + c_2''(x_2^{l,P})\right) - 1} > 0,$$

(A3) $$\frac{\partial x_2^{l,P}}{\partial\alpha} = \frac{\partial x_2^{h,P}}{\partial\alpha} = -\frac{x_1^P}{\left(2 - \alpha + c_1''(x_1^P)\right)\left(2 + c_2''(x_2^{l,P})\right) - 1} < 0,$$

and

$$\frac{\partial(x_1^P + x_2^{l,P})}{\partial\alpha} = \frac{\partial(x_1^P + x_2^{h,P})}{\partial\alpha} = \frac{x_1^P(1 + c_2''(x_2^{l,P}))}{\left(2 - \alpha + c_1''(x_1^P)\right)\left(2 + c_2''(x_2^{l,P})\right) - 1} > 0.$$

*Q.E.D.*

*A.3 Proof of Proposition 6 and Proposition 7*

Note that

$$\frac{\partial Ew^P(\alpha)}{\partial\alpha} = \mu_0\left[\left[p^P - c_1'(x_1^P)\right]\frac{\partial x_1^P}{\partial\alpha} + \left[p^P - c_2'(x_2^{h,P})\right]\frac{\partial x_2^{h,P}}{\partial\alpha}\right] + (1-\mu_0)\left[\left[p^P - c_1'(x_1^P)\right]\frac{\partial x_1^P}{\partial\alpha} + \left[p^P - c_2'(x_2^{l,P})\right]\frac{\partial x_2^{l,P}}{\partial\alpha}\right].$$

From (A2)–(A3), we know that

$$-\frac{\partial x_1^P/\partial\alpha}{\partial x_2^{l,P}/\partial\alpha} = 2 + c_2''(x_2^{l,P}).$$

Using the fact that $\partial x_2^{l,P}/\partial\alpha = \partial x_2^{h,P}/\partial\alpha$, we find that

$$\text{(A4)}\quad \frac{\partial Ew^P(\alpha)}{\partial\alpha} = -\frac{\partial x_2^{l,P}}{\partial\alpha}\cdot\Big\{\mu_0\Big[\left[2+c_2''(x_2^{l,P})\right]\left[p^P - c_1'(x_1^P)\right] - \left[p^P - c_2'(x_2^{h,P})\right]\Big] + (1-\mu_0)\Big[\left[2+c_2''(x_2^{l,P})\right]\left[p^P - c_1'(x_1^P)\right] - \left[p^P - c_2'(x_2^{l,P})\right]\Big]\Big\}.$$

When $\alpha = 0$, if $c_1(\cdot) = c_2(\cdot)$, then $x_1^P = x_2^{l,P}$, since the firms are symmetric in this case. In addition, note that $p^P - c_2'(x_2^{l,P}) = x_2^{l,P} > 0$. Thus, $[2+c_2''(x_2^{l,P})][p^P - c_1'(x_1^P)] - [p^P - c_2'(x_2^{l,P})] > 0$. Since $c_2'(x_2^{h,P}) > c_2'(x_2^{l,P})$, the inequality $[2+c_2''(x_2^{l,P})][p^P - c_1'(x_1^P)] - [p^P - c_2'(x_2^{h,P})] > 0$ also holds. Therefore, $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=0} > 0$.

When $\alpha = 1$, note that $p^P - c_1'(x_1^P) = 0$ and $p - c_2'(x_2^{l,P}) = x_2^{l,P}$; hence,

$$\left.\frac{\partial Ew^P(\alpha)}{\partial\alpha}\right|_{\alpha=1} = -\frac{\partial x_2^{l,P}}{\partial\alpha}\cdot\Big[-x_2^{l,P} + \mu_0\left[c_2'(x_2^{h,P}) - c_2'(x_2^{l,P})\right]\Big],$$

which is less than 0 if $c_2'(x_2^{h,P}) - c_2'(x_2^{l,P})$ is sufficiently small – i.e., if $c_2''$ is sufficiently close to zero. However, when $c_2''$ is so large that $p^P < c_2'(x_2^{h,P})$, then taking into account firm 2's FOC $p^P - c_2'(x_2^{l,P}) = x_2^{l,P}$, we have $[c_2'(x_2^{h,P}) - c_2'(x_2^{l,P})] - x_2^{l,P} > 0$. Therefore, there always exists a $\mu_0 \in [0,1]$ such that $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=1} > 0$. *Q.E.D.*

*A.4 Proof of Proposition 8*

If $\alpha_P \in \text{int}(\Lambda)$, then $\partial Ew^P(\alpha)/\partial\alpha|_{\alpha=\alpha_P} = 0$, and $\partial^2 Ew^P(\alpha)/\partial\alpha^2|_{\alpha=\alpha_P} \leq 0$. If $\partial^2 Ew^P(\alpha)/\ \partial\alpha^2|_{\alpha=\alpha_P} \neq 0$, then

$$\frac{\partial\alpha_P}{\partial\mu_0} = -\frac{\partial^2 Ew^P(\alpha)/(\partial\alpha\partial\mu_0)}{\partial^2 Ew^P(\alpha)/\partial\alpha^2}.$$

It is straightforward to show that $\partial^2 Ew^P(\alpha)/(\partial\alpha\partial\mu_0) > 0$ according to (A4), since $c_2'(x_2^{h,P}) > c_2'(x_2^{l,P})$. Therefore, $\partial\alpha_P/\partial\mu_0 > 0$. *Q.E.D.*

### *A.5 Proof of Lemma 2*

From (14)–(15) and (17)–(19), it is clear that $x_1^P = x_1^{l,B}$ and $x_2^{l,P} = x_2^{l,B}$, so that $x_1^P + x_2^{l,P} = x_1^{l,B} + x_2^{l,B}$.

To compare $x_1^P + x_2^{h,P}$ with $x_1^{h,B} + x_2^{h,B}$, note that

$$\begin{aligned} & x_1^P + x_2^{h,P} - \left[x_1^{h,B} + x_2^{h,B}\right] \\ &= x_1^P + x_2^{h,P} - (x_1^P + x_2^{l,P}) + \left[(x_1^{h,B} + x_2^{h,B}) - (x_1^{l,B} + x_2^{l,B})\right] \\ &= d^h - d^l + \left[(x_1^{h,B} + x_2^{h,B}) - (x_1^{l,B} + x_2^{l,B})\right] \\ &= \left[d^h - (x_1^{h,B} + x_2^{h,B})\right] - \left[d^l - (x_1^{l,B} + x_2^{l,B})\right]. \end{aligned}$$

From (15), we know

$$\left[d^h - (x_1^{h,B} + x_2^{h,B})\right] - \left[d^l - (x_1^{l,B} + x_2^{l,B})\right] = (x_1^{h,B} - x_1^{l,B}) + c(x_1^{h,B}) - c(x_1^{l,B}),$$

which is strictly greater than 0 for any $\alpha$. Therefore, $x_1^P + x_2^{h,P} > x_1^{h,B} + x_2^{h,B}$ holds for any $\alpha$. *Q.E.D.*

### *A.6 Proof of Proposition 9*

Take the derivative of $Ew^P(\alpha)$,

$$\begin{aligned} \frac{\partial Ew^P(\alpha)}{\partial \alpha} = -\frac{\partial x_2^{l,P}}{\partial \alpha} \cdot \Big\{ & \mu_0 \Big[\left[1 + c_2''(x_2^{l,P})\right]\left[p^P - c_1'(x_1^P)\right] + \left[c_2'(x_2^{h,P}) - c_1'(x_1^P)\right]\Big] \\ & + (1-\mu_0)\Big[\left[1 + c_2''(x_2^{l,P})\right]\left[p^P - c_1'(x_1^P)\right] + \left[c_2'(x_2^{l,P}) - c_1'(x_1^P)\right]\Big]\Big\}. \end{aligned}$$

For notational simplicity, let $G^{l,P} = [1 + c_2''(x_2^{l,P})][p^P - c_1'(x_1^P)] + [c_2'(x_2^{l,P}) - c_1'(x_1^P)]$ and $G^{h,p} = [1 + c_2''(x_2^{l,P})][p^P - c_1'(x_1^P)] + [c_2'(x_2^{h,P}) - c_1'(x_1^P)]$. Thus,

$$\text{(A5)} \qquad \frac{\partial Ew^P(\alpha)}{\partial \alpha} = -\frac{\partial x_2^{l,P}}{\partial \alpha} \cdot [\mu_0 G^{h,P} + (1-\mu_0) G^{l,P}].$$

Then, consider $\partial Ew^B(\alpha)/\partial\alpha$. It is straightforward to get

$$\begin{aligned} & \frac{\partial Ew^B(\alpha)}{\partial \alpha} \\ &= \left(-\frac{\partial x_2^{h,B}}{\partial \alpha}\right)\mu_0\Big[\left[1 + c_2''(x_2^{h,B})\right]\left[p^{h,B} - c_1'(x_1^{h,B})\right] + \left[c_2'(x_2^{h,B}) - c_1'(x_1^{h,B})\right]\Big] \\ & \quad + \left(-\frac{\partial x_2^{l,B}}{\partial \alpha}\right)(1-\mu_0)\Big[\left[1 + c_2''(x_2^{l,B})\right]\left[p^{l,B} - c_1'(x_1^{l,B})\right] + \left[c_2'(x_2^{l,B}) - c_1'(x_1^{l,B})\right]\Big]. \end{aligned}$$

Denote $G^{h,B} = [1 + c_2''(x_2^{h,B})][p^{h,B} - c_1'(x_1^{h,B})] + [c_2'(x_2^{h,B}) - c_1'(x_1^{h,B})]$ and $G^{l,B} = [1 + c_2''(x_2^{l,B})][p^{l,B} - c_1'(x_1^{l,B})] + [c_2'(x_2^{l,B}) - c_1'(x_1^{l,B})]$. Hence,

$$\text{(A6)} \qquad \frac{\partial w^B(\alpha)}{\partial \alpha} = \left(-\frac{\partial x_2^{h,B}}{\partial \alpha}\right) \cdot \mu_0 G^{h,B} + \left(-\frac{\partial x_2^{l,B}}{\partial \alpha}\right)(1-\mu_0) G^{l,B}.$$

Now, let us compare (A5) with (A6). Since $x_1^{l,B} = x_1^{l,P}$ and $x_2^{l,B} = x_2^{l,P}$, and thus $-(\partial x_2^{l,P}/\partial\alpha)(1-\mu_0)G^{l,P} = -(\partial x_2^{l,B}/\partial\alpha)(1-\mu_0)G^{l,B}$, we only need to compare $-(\partial x_2^{l,P}/\partial\alpha)\mu_0 G^{h,P}$ with $-(\partial x_2^{h,B}/\partial\alpha)\mu_0 G^{h,B}$.

We first show that $0 < -\partial x_2^{l,P}/\partial\alpha < -\partial x_2^{h,B}/\partial\alpha$. By the FOCs of firms' problems, we can get

$$-\frac{\partial x_2^{l,B}}{\partial\alpha} = \frac{x_1^{l,B}}{\big(2-\alpha+c_1''(x_1^{l,B})\big)\big(2+c_2''(x_2^{l,B})\big)-1}$$

and

$$-\frac{\partial x_2^{h,B}}{\partial\alpha} = \frac{x_1^{h,B}}{\big(2-\alpha+c_1''(x_1^{h,B})\big)\big(2+c_2''(x_2^{h,B})\big)-1}.$$

To compare the numerators, note that from (14) and (15),

$$\frac{\partial x_1^{r,B}}{\partial d^r} = \frac{1+c_2''(x_2^{r,B})}{\big(1+c_2''(x_2^{r,B})\big)\big(2-\alpha_P+c_1''(x_1^{r,B})\big)-1} > 0;$$

hence, $x_1^{h,B} > x_1^{l,B}$ for any cost function. From the above expressions of $-\partial x_2^{l,P}/\partial\alpha$ and $-\partial x_2^{h,B}/\partial\alpha$, when $c_2'' = 0$, obviously, we have $-\partial x_2^{l,B}/\partial\alpha < -\partial x_2^{h,B}/\partial\alpha$. By the continuity of (14) and (15), when $c_2''$ is sufficiently close to 0, $-\partial x_2^{l,B}/\partial\alpha < -\partial x_2^{h,B}/\partial\alpha$ also holds.

Next, we show that $G^{h,P} < G^{h,B}$ when $c_1''$ and $c_2''$ are sufficiently close to 0. By firm 1's maximization, we know that $p^{h,B} - c_1'(x_1^{h,B}) = (1-\alpha)x_1^{h,B}$ and $p^{l,B} - c_1'(x_1^{l,B}) = (1-\alpha)x_1^{l,B}$. Therefore,

$$\begin{aligned}
G^{h,B} &= \big[1+c_2''(x_2^{h,B})\big](1-\alpha)x_1^{h,B} + \big[c_2'(x_2^{h,B}) - c_1'(x_1^{h,B})\big],\\
G^{h,P} &= \big[1+c_2''(x_2^{h,P})\big](1-\alpha)x_1^{P} + \big[c_2'(x_2^{h,P}) - c_1'(x_1^{P})\big].
\end{aligned}$$

When $c_1'' = 0$ and $c_2'' = 0$, since $x_1^{h,B} > x_1^{l,B} = x_1^P$, we have $G^{h,P} < G^{h,B}$. Therefore, by the continuity of (14) and (15), when $c_1''$ and $c_2''$ are sufficiently close to 0, $G^{h,P} < G^{h,B}$ still holds.

Therefore, $\partial E w^P(\alpha)/\partial\alpha < \partial E w^B(\alpha)/\partial\alpha$. Since $\partial w^P(\alpha)/\partial\alpha = 0$ at $\alpha = \alpha_P$, we have $\partial w^B(\alpha)/\partial\alpha|_{\alpha=\alpha_P} > 0$. *Q.E.D.*

### *A.7 Proof of Proposition 10*

Note that $x_1^P = x_1^{l,B}$, $x_2^{l,P} = x_2^{l,B}$. Hence, given $\alpha$,

$$\begin{aligned}
\frac{1}{\mu_0}\big[Ew^P(\alpha) - Ew^B(\alpha)\big] = &\left[\int_0^{x_1^P + x_2^{h,P}} \big[d^h - q\big]dq - c_1(x_1^P) - c_2(x_2^{h,P})\right]\\
&- \left[\int_0^{x_1^{h,B} + x_2^{h,B}} \big[d^h - q\big]dq - c_1(x_1^{h,B}) - c_2(x_2^{h,B})\right].
\end{aligned}$$

Define $C(x) = \min_{x_1}\left[c_1(x_1)+c_2(x-x_1)\right]$. We need to show that $w^P(\alpha)-w^B(\alpha)\geq 0$, which is equivalent to

(A7)
$$\int_0^{x_1^P+x_2^{h,P}}\left[d^h-q\right]dq - C(x_1^P+x_2^{h,P}) - \left[c_1(x_1^P)+c_2(x_2^{h,P})-C(x_1^P+x_2^{h,P})\right]$$
$$\geq \int_0^{x_1^{h,B}+x_2^{h,B}}\left[d^h-q\right]dq - C(x_1^{h,B}+x_2^{h,B}) - \left[c_1(x_1^{h,B})+c_2(x_2^{h,B})-C(x_1^{h,B}+x_2^{h,B})\right].$$

Define $g^r(x)=\int_0^x(d^r-q)dq-C(x)$ for $r=h,l$. Taking the derivative with respect to $x$, we know that $g^r(x)$ is increasing in $x$ when $d^r-x\geq C'(x)$. Let $\bar{x}^r$ be the $x$ satisfying

$$d^r - x = C'(x) \tag{A8}$$

for $r=h,l$. From Lemma 2, we know that $x_1^P+x_2^{h,P}>x_1^{h,B}+x_2^{h,B}$, so if $x_1^P+x_2^{h,P}<\bar{x}^h$, then

$$\int_0^{x_1^P+x_2^{h,P}}\left[d^h-q\right]dq - C(x_1^P+x_2^{h,P}) > \int_0^{x_1^{h,B}+x_2^{h,B}}\left[d^h-q\right]dq - C(x_1^{h,B}+x_2^{h,B}) \tag{A9}$$

is true.

Next, we show that $x_1^P+x_2^{h,P}<\bar{x}^h$ is true. First, from the FOCs, when $\alpha<1$, $p>c_1'(x_1^P)$ and $p>c_2'(x_2^{l,P})$, so $p>\max\{c_1'(x_1^P),c_2'(x_2^{l,P})\}$. In addition, for $x=x_1^P+x_2^{l,P}$ we have $C'(x)=c_1'(x_1^*)=c_2'(x_2^*)$, where $x_1^*\in\operatorname{argmin}_{x_1}[c_1(x_1)+c_2(x-x_1)]$ and $x_2^*=x-x_1^*$. Note that either $x_1^*\leq x_1^P$ or $x_2^*\leq x_2^{l,P}$; so either $c_1'(x_1^*)\leq c_1'(x_1^P)$ or $c_2'(x_2^*)\leq c_2'(x_2^{l,P})$. This implies that $p>\max\{c_1'(x_1^P),c_2'(x_2^{l,P})\}\geq C'(x_1^P+x_2^{l,P})$, i.e., $x_1^P+x_2^{l,P}<\bar{x}^l$. Hence, $x_1^P+x_2^{h,P}<\bar{x}^l+(d^h-d^l)$. By (A8), $\bar{x}^h=\bar{x}^l+(d^h-d^l)-[C'(\bar{x}^h)-C'(\bar{x}^l)]$. Since $x_1^P+x_2^{h,P}<\bar{x}^l+(d^h-d^l)$ holds with strictly inequality, $x_1^P+x_2^{h,P}<\bar{x}^h$ will hold when $C'(\bar{x}^h)-C'(\bar{x}^l)$ is sufficiently small, which is true when $c_i''$ $(i=1,2)$ is sufficiently small.

Hence, (A9) holds with strict inequality when $c_i''$ is sufficiently small. Now, let $\eta>0$ be a value satisfying

$$0<\eta<\left[\int_0^{x_1^P+x_2^{h,P}}\left[d^h-q\right]dq - C(x_1^P+x_2^{h,P})\right]-\left[\int_0^{x_1^{h,B}+x_2^{h,B}}\left[d^h-q\right]dq - C(x_1^{h,B}+x_2^{h,B})\right].$$

Note that for any $\eta>0$, when $c_i''$ $(i=1,2)$ is sufficiently close to zero, we have

$$\left[c(x_1^P)+c(x_2^{h,P})-C(x_1^P+x_2^{h,P})\right]-\left[c(x_1^{h,B})+c(x_2^{h,B})-C(x_1^{h,B}+x_2^{h,B})\right]<\eta.$$

Therefore, (A7) is true and $Ew^P(\alpha)\geq Ew^B(\alpha)$ when $c_i''(i=1,2)$ is sufficiently small. *Q.E.D.*

*References*

Barros, F. (1995), “Incentive Schemes as Strategic Variables: An Application to a Mixed Duopoly,” *International Journal of Industrial Organization*, 13(3), 373–386.

Biais, B., and E. Perotti (2002), “Machiavellian Privatization,” *The American Economic Review*, 92(1), 240–258.

Bortolotti, B., M. Fantini, and D. Siniscalco (2003), “Privatisation around the World: Evidence from Panel Data,” *Journal of Public Economics*, 88(1–2), 305–332.

De Fraja, G., and F. Delbono (1989), “Alternative Strategies of a Public Enterprise in Oligopoly,” *Oxford Economic Papers*, 41(2), 302–311.

Dewenter, K. L., and P. H. Malatesta (2001), “State-Owned and Privately Owned Firms: An Empirical Analysis of Profitability, Leverage, and Labor Intensity,” *The American Economic Review*, 91(1), 320–334.

D’Souza, J., and W. L. Megginson (1999), “The Financial and Operating Performance of Privatized Firms during the 1990s,” *The Journal of Finance*, 54(4), 1397–1438.

Du, N., J. S. Heywood, and G. Ye (2013), “Strategic Delegation in an Experimental Mixed Duopoly,” *Journal of Economic Behavior & Organization*, 87, 91–100.

Fan, J. P. H., T. J. Wong, and T. Zhang (2007), “Politically Connected CEOs, Corporate Governance, and Post-IPO Performance of China’s Newly Partially Privatized Firms,” *Journal of Financial Economics*, 84(2), 330–357.

Gaskins, D. W., Jr. (1971), “Dynamic Limit Pricing: Optimal Pricing under Threat of Entry,” *Journal of Economic Theory*, 3(3), 306–322.

Gupta, N. (2005), “Partial Privatization and Firm Performance,” *The Journal of Finance*, 60(2), 987–1015.

Jones, S. L., W. L. Megginson, R. C. Nash, and J. M. Netter (1999), “Share Issue Privatizations as Financial Means to Political and Economic Ends,” *Journal of Financial Economics*, 53(2), 217–253.

Klien, M. (2014), “Corporatization and the Behavior of Public Firms: How Shifting Control Rights Affects Political Interference in Water Prices,” *Review of Industrial Organization*, 44(4), 393–422.

Li, H., and L.-A. Zhou (2005), “Political Turnover and Economic Performance: The Incentive Role of Personnel Control in China,” *Journal of Public Economics*, 89(9–10), 1743–1762.

Matsumura, T. (1998), “Partial Privatization in Mixed Duopoly,” *Journal of Public Economics*, 70(3), 473–483.

Matthews, S. A., and L. J. Mirman (1983), “Equilibrium Limit Pricing: The Effects of Private Information and Stochastic Demand,” *Econometrica*, 51(4), 981–996.

Maw, J. (2002), “Partial Privatization in Transition Economies,” *Economic Systems*, 26(3), 271–282.

Megginson, W. L., and J. M. Netter (2001), “From State to Market: A Survey of Empirical Studies on Privatization,” *Journal of Economic Literature*, 39(2), 321–389.

Merrill, W. C., and N. Schneider (1966), “Government Firms in Oligopoly Industries: A Short-Run Analysis,” *The Quarterly Journal of Economics*, 80(3), 400–412.

Milgrom, P., and J. Roberts (1982), “Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis,” *Econometrica*, 50(2), 443–459.

Riordan, M. H. (1985), “Imperfect Information and Dynamic Conjectural Variations,” *The RAND Journal of Economics*, 16(1), 41–50.

Seamans, R. C. (2013), “Threat of Entry, Asymmetric Information, and Pricing,” *Strategic Management Journal*, 34(4), 426–444.

White, M. D. (2002), “Political Manipulation of a Public Firm’s Objective Function,” *Journal of Economic Behavior & Organization*, 49(4), 487–499.

John S. Heywood
Department of Economics
University of Wisconsin–Milwaukee
P.O. Box 413
Milwaukee, WI 53201
U.S.A.
heywood@uwm.edu

Xiangting Hu
Hanqing Advanced Institute of
Economics and Finance
Renmin University of China
59 Zhongguancun Street
100872 Beijing
P.R. China
xiangthu@ruc.edu.cn

Guangliang Ye
Antitrust and Competition Policy Center
Hanqing Institute
National Academy of Development and
Strategy
Renmin University of China
59 Zhongguancun Street
100872 Beijing
P.R. China
gye@ruc.edu.cn

# All Deceptions Are Not Alike: Bayesian Mechanism Design with a Social Norm against Lying

by

Ville Korpela*

Received January 10, 2015; in revised form June 2, 2016;
accepted June 10, 2016

We say that a society has a weak norm against lying if, all other things being equal, agents try to avoid getting caught lying. We show that if this is the case, and it usually is, then Bayesian monotonicity is no longer a constraint in implementation and all incentive-compatible social-choice functions are Bayesian implementable. In contrast to the previous literature, our result does not rely on any kind of intrinsic lying aversion, on which the experimental evidence is mixed. (JEL: B41, C72, D78, D82)

## 1 Introduction

From the very beginning, mechanism design has relied heavily on one of its foundation stones – the revelation principle. This principle says, roughly speaking, that any social-choice function (SCF) that can be realized as a Bayes–Nash equilibrium of some mechanism can also be realized as a truthful equilibrium of a direct mechanism where each agent is simply asked to announce his or her type. It is hard to say who should be credited for discovering this obvious yet powerful principle. According to Myerson (2008) it is a theorem that was found independently by several authors (e.g., Dasgupta, Hammond, and Maskin, 1979; Harris and Townsend, 1981; Holmström, 1977; Myerson, 1979; Rosenthal, 1978), all building on the previous ideas of Gibbard (1973) and Aumann (1974).[1]

* Turku School of Economics, Turku, Finland. I thank Johannes Hörner, Biung-Ghi Ju, Juuso Välimäki, Hannu Vartiainen, Hannu Salonen, Hannu Nurmi, Nadine Chlaß, three anonymous referees, and participants of SING10 (Krakow), HECER Applied Microeconomics Seminar (Helsinki), and UECE Lisbon Meetings 2014: Game Theory and Applications (Lisbon) for useful comments and suggestions. Financial support from the Academy of Finland is gratefully acknowledged.

1 The reason why it is so hard to give priority in this matter was probably best stated by Rosenthal when he saw the result restated some years after his own discovery: "Why

*Journal of Institutional and Theoretical Economics* 173, 376–393 – ISSN 0932-4569
DOI: 10.1628/093245616X14689075188007 

However, as old as the principle itself is the observation that the associated direct mechanism may have other equilibria beside the truthful one – equilibria where multiple agents are lying simultaneously and that do not implement the SCF. Moreover, sometimes these other equilibria are not present in all indirect mechanisms (e.g., Palfrey, 1992; Palfrey and Srivastava, 1993; Feldman and Serrano, 2006, chapter 16). The problem, now commonly known as *the multiple-equilibrium problem*, is that for some reason or another, a reason outside the mechanism and beyond the control of the mechanism designer, one of these other equilibria may become focal and end up being played. This is an unfortunate shortcoming of the revelation principle, and there is no consensus on how severe the problem really is.

One possible way to approach this issue is to look what the extensive experimental literature on dishonesty and deceitful behavior can tell us. All behavioral experiments suggest that a certain fraction of people behave in a predicted way while the rest do not, the fit being far from perfect (e.g., Fischbacher and Föllmi-Heusi, 2013; Gneezy, 2005; Mazar, Amir, and Ariely, 2008; Hurkens and Kartik, 2009; Greene and Paxton, 2009). So what should this fraction be – 70 %, 80 %, 90 %, or perhaps a utopian 99.99 %? The fact is that we do not want the mechanism to implement the SCF only some of the time; rather we want the mechanism to always implement it – or at least to a very large degree, the failures being an extremely rare and exceptional cases. In other words, instead of relying on experimental results, it would be better to find something that is a truly invariant feature of human behavior, something that even the pure logic of the situation suggests is necessary.

This is where social norms, the customs and conventions that govern behavior in societies, can be helpful. No one would deny that there is such thing as a social norm against lying.[2] This does not imply that people do not lie, and there is certainly ample evidence to the contrary (Hao and Houser, 2011); rather it only implies that, other things being equal, people prefer not to get caught doing so. Surprisingly, as weak as this regularity may seem to be (since it does not rely on intrinsic preference for honesty but only on outside pressure), it turns out that a social norm against lying is sufficient to guarantee that there is nothing operationally wrong in relying on the revelation principle: The search for an optimal mechanism can be directed to the set of all incentive-compatible SCFs without a need to worry about the multiple-equilibrium problem, since all incentive-compatible SCFs can be uniquely implemented.

Our argument goes as follows: First we show all Bayes–Nash equilibria of any direct mechanism can be divided into two disjoint sets – those in which everybody believes that some agent is certainly lying and those in which everybody only be-

---

on earth would anybody want to write that down?" (Radner and Ray, 2003). Myerson himself has shown how powerful this principle is in economic analysis (Myerson, 1981).

[2] Psychologists have observed that lying is cognitively more expensive than truth-telling (Greene and Paxton, 2009). It is likely that this is due to the social norm against lying, since it implies that one has to be more careful when lying so that one does not get caught.

lieves that this agent may be lying or is in fact truthful. Then, we give a refinement of the standard Bayes–Nash equilibrium that requires, roughly speaking, that if an agent has to select from two strategies that give him exactly the same material outcome, he will always select the one in which everybody believes that he may be lying rather than the one in which everybody believes he is certainly lying.[3] Finally, to complete the argument, we show that with this behavior the set of incentive-compatible SCFs coincides exactly with the set of Bayesian-implementable SCFs. Therefore, there is nothing wrong in relying on the revelation principle from an operational point of view. Incentive compatibility is a necessary condition for both implementation and realization, while in the presence of a social norm against lying it is sufficient as well, and thus it is completely legitimate to restrict the search for a best SCF to the set of incentive-compatible direct mechanisms.

The rest of this paper is organized as follows. In section 2 we formulate the Bayesian mechanism design problem in a quasilinear environment and then give an exact statement of the revelation principle. In section 3 we explain why some strategies are preferred to others in the presence of a weak social norm against lying. Based on this observation, we give a refinement of the standard Bayes–Nash equilibrium and prove that all incentive-compatible SCFs now become implementable.[4] The actual proof, which is based on a canonical mechanism that can implement any incentive-compatible SCF in the presence of a weak social norm against lying, is relegated to the appendix. Section 4 gives a concrete example of the ideas used in the proof. Section 5 presents some further connections with the literature, and section 6 concludes with a brief discussion.

## 2 *A General Mechanism Design Setting with a Numéraire*

There is a finite group of agents that interact to make a joint decision. We denote the set of *agents* by $N = \{1,2,\ldots,n\}$, with a generic element represented by $i$, $j$, or $k$, and the set of *decisions* by $D$, with a generic element represented by $d$ or $d'$. All agents hold private information, and the information of agent $i$ is represented by a *type* $\theta_i$ that lies in a finite set $\Theta_i$. A *state* is any profile of types $\theta = (\theta_1,\ldots,\theta_n) \in \Theta = \times_{i\in N} \Theta_i$. Let

$$\Theta_{-i} = \times_{j\neq i} \Theta_j \quad \text{and} \quad \theta_{-i} = (\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_n),$$

as usual. We assume that there is a common prior belief $p(\cdot)$ over the set of states $\Theta$. At the interim stage, after the type $\theta_i$ of agent $i$ has been realized, beliefs are up-

[3] Dutta and Sen (2012) has done this under complete information. Their concept is known as *partial honesty*.

[4] See d'Aspremont and Gérard-Varet (1979, 1982) for more on incentive compatibility.

dated using *Bayes's rule*:

$$p(\theta_{-i} \mid \theta_i) = \frac{p(\theta_{-i},\theta_i)}{\sum_{\theta'_{-i} \in \Theta_{-i}} p(\theta'_{-i},\theta_i)}.$$

For notational simplicity we assume that $\text{supp}(\Theta) = \{\theta \in \Theta \mid p(\theta) > 0\} = \Theta$. Otherwise we would need to qualify everything by saying that it holds on the support of $\Theta$.[5]

A *decision rule* (or an *allocation rule*) is a mapping $d : \Theta \to D$ that selects a decision $d(\theta) \in D$ as a function of the state $\theta \in \Theta$. In order to provide incentives it is possible for the mechanism designer to tax or subsidize agents. This is represented by a *transfer function* $t : \Theta \to \mathbb{R}^n$, where $t_i(\theta)$ is the *expected payment* that agent $i$ receives or makes (if negative) when the state is $\theta \in \Theta$. A *social-choice function* (SCF) $f : \Theta \to D \times \mathbb{R}^n$, or the *goal*, is any mapping $f = (d(\cdot),t(\cdot))$ defined by a decision rule $d(\cdot)$ together with a transfer function $t(\cdot)$. Each agent has a preference relation over decisions and money representable by a utility function $u_i : D \times \mathbb{R}^n \times \Theta \to \mathbb{R}$ that is assumed to be quasilinear in money. Thus, we can write $u_i(d,t;\theta) = v_i(d;\theta) + t_i$, and $u_i(d,t;\theta) > u_i(d',t';\theta)$ indicates that agent $i$ prefers $(d,t)$ to $(d',t')$ when the state is $\theta \in \Theta$.

To complete the model we need to define how agents select when they face an uncertain prospect. To this end, let $F$ be the set of all possible SCFs. A generic element of this set will be denoted by $f$, $g$, or $h$. The utility function $u_i$ together with the prior belief $p(\cdot)$ determines an (interim) expected utility of the SCF $f = (d(\cdot),t(\cdot)) \in F$ for agent $i$ with type $\theta_i \in \Theta_i$ as

$$U_i(f;\theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} [v_i(d(\theta_i,\theta_{-i});(\theta_{-i},\theta_i)) - t_i(\theta_i,\theta_{-i})]\, p(\theta_{-i} \mid \theta_i).$$

We call $E = (N, F, \Theta, p(\cdot), \{U_i\})$ an *environment*.

A *mechanism* is a pair $\Gamma = (M,\mu)$, where $M = M_1 \times \cdots \times M_n$ is the *message space* and $\mu : M \to D \times \mathbb{R}^n$ is the *outcome function*. Thus, for any profile of messages $m = (m_1,\ldots,m_n) \in M$, $\mu(m) = (\mu_0(m),\mu_1(m),\ldots,\mu_n(m))$ is the resulting decision $\mu_0(m) \in D$ together with the expected transfers $\mu_1(m),\ldots,\mu_n(m) \in \mathbb{R}$. A *strategy* of agent $i$ is a function $\sigma_i : \Theta_i \to M_i$. We write $\Sigma_i$ for the set of all strategies of agent $i$, and $\Sigma = \times_{i=1}^n \Sigma_i$ for the set of all strategy profiles. A strategy profile $\sigma = (\sigma_1,\ldots,\sigma_n) \in \Sigma$ is a *Bayes–Nash equilibrium* of $\Gamma$ if, and only if, $U_i(\mu(\sigma_i,\sigma_{-i});\theta_i) \geq U_i(\mu(m_i,\sigma_{-i});\theta_i)$ for all $i \in N$, $\theta_i \in \Theta_i$, and $m_i \in M_i$.[6] Let us denote the set of all Bayes–Nash equilibria in $\Gamma$ by $\text{BNE}(\Gamma)$.

We say that mechanism $\Gamma$ *realizes the SCF $f \in F$ in Bayes–Nash equilibrium* if there exists $\sigma \in \text{BNE}(\Gamma)$ such that $\mu(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$. Sometimes, however, the mechanism designer may have a stronger objective in mind.

[5] See Jackson (1991) for a definition of incentive compatibility in the case $\text{supp}(\Theta) \neq \Theta$.

[6] See Harsanyi (1967, 1968a,b), for an in-depth analysis of this equilibrium concept. Notice that $\mu(m_i,\sigma_{-i})$ is constant with respect to $\theta_i$.

We say that mechanism $\Gamma$ *implements the SCF* $f \in F$ *in Bayes–Nash equilibrium* if $\mathrm{BNE}(\Gamma) \neq \emptyset$ and for each $\sigma \in \mathrm{BNE}(\Gamma)$ we have $\mu(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$. As usual, we say that $f$ is *Bayesian implementable* if there exists a mechanism $\Gamma$ that implements $f$ in Bayes–Nash equilibrium.[7]

A *direct mechanism* is a mechanism $\Gamma^f = (\Theta, f)$ where the message space is the set of states $\Theta$ and the outcome function is an SCF $f$. In other words, direct mechanisms simply ask agents to announce their types and then select whatever outcome is recommended by the SCF. A strategy of agent $i$ in $\Gamma^f$ is a type announcement function $\alpha_i : \Theta_i \to \Theta_i$. The identity function is called the *revelation strategy* and denoted by $\widehat{\alpha}_i$ for agent $i$. All other type announcement functions are called *deceptions*. An SCF $f$ is called *incentive-compatible* (IC) if the revelation strategy profile $\widehat{\alpha} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_n)$ is a Bayes–Nash equilibrium of the associated direct mechanism $\Gamma^f$, that is, if

$$U_i(f(\widehat{\alpha}_i, \widehat{\alpha}_{-i}); \theta_i) \geq U_i(f(\theta_i', \widehat{\alpha}_{-i}); \theta_i) \quad \text{for all } i \in N \text{ and all } \theta_i, \theta_i' \in \Theta_i.$$

This says that when all agents except one are telling the truth, the remaining agent does not have an incentive to lie either. We are finally ready to state the celebrated revelation principle.[8]

The Revelation Principle for Bayes–Nash Equilibrium *If a mechanism* $\Gamma = (M, \mu)$ *realizes (or implements) the SCF* $f \in F$ *in Bayes–Nash equilibrium, then the revelation strategy profile* $\widehat{\alpha}$ *must be an equilibrium of the associated direct mechanism* $\Gamma^f$*. In other words, the social-choice function* $f$ *must be IC.*

This result is famous for its ability to make hard problems tractable. It tells us that instead of looking at all possible mechanisms we can restrict attention to revelation strategies of incentive-compatible direct mechanisms. Unfortunately, it has been shown many times by many different authors that an incentive-compatible direct mechanism can have other equilibria beside the truthful one. This means that sometimes there exists another Bayes–Nash equilibrium in which multiple agents are lying simultaneously, and it may not even be possible to implement an SCF that can be realized using an incentive-compatible direct mechanism (Jackson, 1991; Duggan, 1995; Bergin, 1995).[9] That is to say, any mechanism, direct or indirect, has another equilibrium that does not coincide with the SCF. Therefore, the revelation principle relies heavily on the revelation strategy profile being somehow focal, or the mechanism designer at least being able to make it so.[10]

[7] Mookherjee and Reichelstein (1990), Jackson (1991), Palfrey and Srivastava (1993), and Duggan (1995) all provide characterizations of Bayesian-implementable SCFs and more generally of social-choice correspondences.

[8] See Dasgupta, Hammond, and Maskin (1979) for a proof.

[9] Since IC is not sufficient for implementation in the standard sense.

[10] There are nice stories in Myerson (2009) that suggest nearly any equilibrium can become focal. See also Schelling (1960).

### 3 Bayesian Implementation with a Weak Social Norm against Lying

Recently, in Korpela (2014), it was shown that if all agents are partially honest, which means, roughly speaking, that agents do not lie unless it affects their material payoff, the set of incentive-compatible SCFs coincides exactly with the set of implementable SCFs in any standard resource allocation problem. In other word, if by focality of truth-telling we mean that all agents are partially honest and therefore do not lie just for the sake of the act itself, then the revelation principle works fine as a practical tool.

There are, however, situations where intrinsic preference for honesty does not have a bite, while a social norm against lying still governs behavior. The intuition for this is that intrinsic preference for honesty applies only when truth-telling is chosen instead of lying, while a social norm against lying applies also when one way of lying is chosen instead of another. In line with this intuition, we acknowledge that people lie but prefer to do it in a concealed rather than obvious way. To define this distinction exactly, let us consider *augmented direct mechanisms* where other things in addition to type are announced.[11] Suppose that the message space of agent $i$ is $M_i = \Theta_i \times Q_i$, where $Q_i$ is an abstract set with no special meaning attached to it. Moreover, let us divide the strategy $\sigma_i : \Theta_i \to M_i$ of agent $i$ into two parts $\sigma_i = (\alpha_i, s_i)$, where $\alpha_i : \Theta_i \to \Theta_i$ is the *type announcement* and $s_i : \Theta_i \to Q_i$ is the *auxiliary strategy*. Two kinds of type announcements can be a part of Bayes–Nash equilibrium: those in which everybody believes that an agent is certainly lying, and those in which everybody believes that an agent may be lying or may in fact be truthful.

Formally, if agent 1 with type $\theta_1$ lies that he is of type $\theta_1'$ instead (that is, $\alpha_1(\theta_1) = \theta_1'$), and at the same time with type $\theta_1'$ lies that he is of type $\theta_1''$ instead (that is, $\alpha_1(\theta_1') = \theta_1''$), then everybody believes that agent 1 is lying when the true type is $\theta_1$, as they do not expect to see $\theta_1'$ played under $\theta_1'$. On the other hand, if under another deception $\alpha_1'$ agent 1 does not lie when his type is $\theta_1'$ (that is, $\alpha_1'(\theta_1') = \theta_1'$), then everybody believes only that he may be lying under $\theta_1$, as the true type could really be $\theta_1'$. That is, others cannot tell whether he is lying or not, or at least they cannot accuse him of such behavior.

*Figure 1*
Obvious and Concealed Deception
(no leaving arrow means that the type does not lie)

$$\theta_1 \xrightarrow[\alpha_1]{} \theta_1' \xrightarrow[\alpha_1]{} \theta_1'' \qquad \theta_1 \xrightarrow[\alpha_1']{} \theta_1' \qquad \theta_1''$$

[11] This concept was defined by Mookherjee and Reichelstein (1990). Although the distinction between concealed and obvious deception can be made using a *pure direct mechanism*, where only type is announced, our result cannot be derived without an augmented direct mechanism (see Example 2).

EXAMPLE 1. There are two agents $N = \{1,2\}$ with two possible types $\Theta_1 = \Theta_2 = \{L(\text{azy}), P(\text{roductive})\}$. The set of decisions is $D = \{d, d'\}$, where $d$ is interpreted as work together and $d'$ is interpreted as work separately, with payoffs $v_1(d;L) = v_1(d;P) = v_2(d;L) = v_2(d;P) = 2$ and $v_1(d';L) = v_1(d';P) = v_2(d';L) = v_2(d';P) = 1$. In other words, no matter what the types are, it is always better for the agents to work together. The mechanism designer, however, wants agents to work together only if they both have the same type. Therefore, she wants to implement the SCF that the direct mechanism given in Figure 2 realizes under truthful behavior.

One possible interpretation of this SCF is the following: If agents have the same type, then the product that they make together will be superior, while if agents have different types, then the product that they make separately will be superior.

*Figure 2*
The Goal of the Mechanism Designer

| | | Agent 2: $L$ | Agent 2: $P$ |
|---|---|---|---|
| Agent 1 | $L$ | $d$ | $d'$ |
| | $P$ | $d'$ | $d$ |

Now suppose that beliefs are such that $p(\cdot) = p_1(\cdot)p_2(\cdot)$ and $p_1(L) = p_1(P) = p_2(L) = p_2(P) = 1/2$. The deception $\alpha = (\alpha_1, \alpha_2)$, where $\alpha_1(L) = \alpha_1(P) = P$ and $\alpha_2(L) = \alpha_2(P) = P$, is a Bayes–Nash equilibrium of the direct mechanism in Figure 2 under these beliefs. In this equilibrium both agents are using a concealed deception. Another Bayes–Nash equilibrium of this mechanism is the deception $\alpha = (\alpha_1, \alpha_2)$, where $\alpha_1(L) = \alpha_2(L) = P$ and $\alpha_1(P) = \alpha_2(P) = L$. In this equilibrium, however, agents are using obvious deceptions (each believes that the other is always lying).

Mathematically the difference between an obvious and a concealed deception is simply whether the deception is an idempotent function or not.

DEFINITION 1 *A deception $\alpha_i$ is idempotent if and only if $\alpha_i \circ \alpha_i \equiv \alpha_i$. In all other cases it is called nonidempotent, and there will then exist a type $\theta_i \in \Theta_i$ such that $\alpha_i(\alpha_i(\theta_i)) \neq \alpha_i(\theta_i)$. An idempotent deception is called concealed, in line with the lying interpretation, and a nonidempotent deception is called obvious.*

In words, if the deception of agent $i$ is obvious, then some type of this agent will mimic another type that would not be truthful either. With this definition in mind, we can express the idea that some strategies are preferred to others in the presence of a weak social norm against lying.

DEFINITION 2 *Let $\Gamma = (M, \mu)$ be a mechanism with $M_i = \Theta_i \times Q_i$ for all $i \in N$. We say that agent $i$ can avoid the obvious deceptions of the strategy $\sigma_i = (\alpha_i, s_i)$*

*at $\sigma = (\sigma_i, \sigma_{-i})$ if there exists another strategy $\sigma_i' = (\alpha_i', s_i')$ such that $\alpha_i'$ is either a concealed deception or the revelation strategy and the expected utility at $\sigma' = (\sigma_i', \sigma_{-i})$ is at least as large as the expected utility at $\sigma$.*

DEFINITION 3 *Let $\Gamma = (M, \mu)$ be a mechanism with $M_i = \Theta_i \times Q_i$ for all $i \in N$. We say that $\sigma \in \mathrm{BNE}(\Gamma)$ is a Bayes–Nash equilibrium under the weak social norm against lying if those agents who are using an obvious deception as part of their strategy cannot avoid it.*

This can be seen as a refinement of the standard Bayes–Nash equilibrium. It says that agents have a strict aversion towards obvious deceptions provided their expected payoffs are not affected. Notice that it is an implicit assumption in Definition 3 that agents cannot break the norm against lying by themselves – even if all other agents use obvious deceptions, one still prefers a concealed deception. One possible interpretation of this is that the norm originates outside the mechanism.[12] Thus, since agents know that they cannot affect the norm, it is better to retain an honest appearance.[13]

EXAMPLE 2. Let us take another look at the SCF studied in Example 1. The Bayes–Nash equilibrium $\alpha = (\alpha_1, \alpha_2)$, where agents play obvious deceptions such that $\alpha_1(L) = \alpha_2(L) = P$ and $\alpha_1(P) = \alpha_2(P) = L$, is no longer an equilibrium under the weak social norm against lying. Both agents can deviate to the revelation strategy without reducing the expected payoff. The other Bayes–Nash equilibrium, where agents play concealed deceptions such that $\alpha_1(L) = \alpha_1(P) = L$ and $\alpha_2(L) = \alpha_2(P) = L$, is a Bayes–Nash equilibrium also under the weak social norm against lying, since any deviation from this equilibrium would reduce the expected payoff. We shall soon see, however, that an augmented direct mechanism will allow us to weed out this equilibrium too.

The ideas of realization and implementation generalize in a straightforward way. Let us denote the set of all Bayes–Nash equilibria with a weak social norm against lying by $\mathrm{BNE}^-(\Gamma)$. We say that the mechanism $\Gamma$ *realizes the SCF $f \in F$ in Bayes–Nash equilibrium with a weak social norm against lying* if there exists at least one $\sigma \in \mathrm{BNE}^-(\Gamma)$ such that $\mu(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$. Similarly, we say that the mechanism $\Gamma$ *implements the SCF $f \in F$ in Bayes–Nash equilibrium with a weak social norm against lying* if for every equilibrium $\sigma \in \mathrm{BNE}^-(\Gamma)$, we have $\mu(\sigma(\theta)) = f(\theta)$ for all $\theta \in \Theta$, and moreover $\mathrm{BNE}^-(\Gamma) \neq \emptyset$.

It is important to understand that IC is a necessary condition for realization and implementation also under the weak norm against lying. On the other hand, it is no longer sensible to speak about the revelation principle, since our solution concept requires that strategies include a type report. In some cases, however, the mechanism designer may be able to use an abstract strategy space $M_i = Q_i$ and enforce a rule of conduct $s_i(\theta_i) = m_i$. If agent $i$ feels that playing against $s_i$ is not acceptable,

12 In this case $N$ must be a small subpopulation and not the entire society.
13 See Diekmann, Przepiorka, and Rauhut (2011), however.

or perhaps regards it as lying, our Theorem (given below) will continue to hold, and we can speak about the revelation principle too. In addition, it might then be possible to use simpler mechanisms.

Now we are finally ready to state our result. A complete proof is given in the appendix.

THEOREM *All incentive-compatible SCFs are Bayes–Nash implementable under the weak norm against lying.*

Suppose that, all other things equal, a dishonest person would always prefer to lie in a concealed way. Under this assumption, if a mechanism can realize an SCF, then it must be IC. On the other hand, by the above Theorem we know that any incentive-compatible SCF can be implemented in Bayes–Nash equilibrium under the weak social norm against lying. Thus, we can reach the following conclusion: In a methodological sense there is nothing wrong in relying on the revelation principle. Excluding implausible equilibria, the set of incentive-compatible SCFs coincides exactly with the set of implementable SCFs, and therefore it is legitimate to search for an optimal mechanism from this set, despite the multiple-equilibrium problem. It is just that a direct revelation mechanism may not be able to implement the SCF and a more complex construction is needed.

EXAMPLE 3. The SCF that was studied in Examples 1 and 2 is IC (check). Therefore, our Theorem claims that this SCF can be implemented under the weak social norm against lying, although not with a direct mechanism, by Example 2. Unfortunately, it is difficult to present the mechanism that we used in the proof graphically, since it has an infinite message space. Nevertheless, for specific problems like this one, there exists a finite mechanism with essentially the same working principle. One possible mechanism to implement this SCF is given in Figure 3.

*Figure 3*
A Mechanism that Implements the SCF
(here $\epsilon$, $\xi$, and $\lambda$ must be suitably chosen)

| | | Agent 2 | | | | | |
|---|---|---|---|---|---|---|---|
| | | $(L,0)$ | $(P,0)$ | $(L,L)$ | $(L,P)$ | $(L,1)$ | $(P,1)$ |
| Agent 1 | $(L,0)$ | $d,0,0$ | $d',0,0$ | $d,0,-\xi$ | $d',0,\epsilon$ | $d',0,0$ | $d,0,0$ |
| | $(P,0)$ | $d',0,0$ | $d,0,0$ | $d,0,\epsilon$ | $d,0,-\xi$ | $d,0,0$ | $d',0,0$ |
| | $(L,L)$ | $d,-\xi,0$ | $d',\epsilon,0$ | $d,0,\lambda$ | $d,0,\lambda$ | $d,0,\lambda$ | $d,\lambda,0$ |
| | $(L,P)$ | $d',\epsilon,0$ | $d,-\xi,0$ | $d,0,\lambda$ | $d,0,\lambda$ | $d,\lambda,0$ | $d,0,\lambda$ |
| | $(L,1)$ | $d',0,0$ | $d,0,0$ | $d,0,\lambda$ | $d,\lambda,0$ | $d,0,\lambda$ | $d,0,\lambda$ |
| | $(P,1)$ | $d,0,0$ | $d',0,0$ | $d,\lambda,0$ | $d,0,\lambda$ | $d,0,\lambda$ | $d,0,\lambda$ |

The following observations verify this:

(1) There does not exist an equilibrium where agent 1 is playing strategies $\{(L,L),(L,P),(L,1),(P,1)\}$ and/or agent 2 is playing strategies $\{(L,L),(L,P),(L,1),(P,1)\}$. This, however, requires that $\lambda$ be sufficiently large in comparison with $\epsilon$ and $\xi$.[14]

(2) This means that in all equilibria (if any) agent 1 is playing strategies $\{(L,0),(P,0)\}$ and agent 2 is playing strategies $\{(L,0),(P,0)\}$. Furthermore, by IC the strategy $\sigma = (\sigma_1,\sigma_2)$, where $\sigma_i(L) = (L,0)$ and $\sigma_i(P) = (P,0)$, is an equilibrium that coincides with the SCF as long as $\xi$ is sufficiently large.

(3) If agent 1 is playing a concealed deception $\sigma_1(L) = \sigma_1(P) = P$ or $\sigma_1(L) = \sigma_1(P) = L$, then agent 2 will deviate under type $L$ to $(L,L)$ or $(L,P)$ accordingly. These strategies can be interpreted as type $L$ lying and type $P$ lying. The same is true if agent 2 is playing a concealed deception.

(4) Agent 1 does not want to play an obvious deception ($\sigma_1(L) = (P,0)$ and $\sigma_1(P) = (L,0)$, for example) either, since he can deviate to strategies $\{(L,1),(D,1)\}$ and get the same outcome while being honest. This is where we need the weak social norm against lying and is the only place we use it. Again, the same is true for agent 2.

This example shows how concealed deceptions of agent 1 are controlled by agent 2, while obvious deceptions are controlled by agent 1 himself. This mechanism is complicated because it highlights the working principles behind our canonical mechanism. More straightforward mechanisms can, and probably do, exist.

It must be emphasized that the traditional *economic environment* assumption (Jackson, 1991; Palfrey and Srivastava, 1993) is not sufficient for our mechanism to work properly. Roughly speaking, an economic environment means that the same decision cannot be the best alternative for any $n-1$ agents at the same time, just as in any standard resource allocation problem where the best alternative for each agent is to keep all resources to himself. This, however, does not guarantee that we can find all the out-of-equilibrium outcomes needed in our mechanism. Whether there exists some mechanism that works in an economic environment is an open question.

## *4 Intrinsic Preference for Honesty versus Social Norm against Lying: A Comparison*

Which is a stronger requirement – intrinsic preference for honesty or a weak social norm against lying? In turns out that they are simply different and neither one is stronger than the other.

[14] It must be profitable to get $\lambda$ under one type even when one consequently gets $-\xi$ under the other.

Let us assume that $\Theta = \{\theta_1, \theta_1', \theta_1''\} \times \{\theta_2, \theta_2'\}$ and $D = \{d, d', d'', o, v\}$. Preferences in different states are given in the table. If types are independent and uniformly distributed, so that $p(\cdot) = p_1(\cdot)p_2(\cdot)$, where $p_1(\theta_1) = p_1(\theta_1') = p_1(\theta_1'') = 1/3$ and $p_2(\theta_2) = p_2(\theta_2') = 1/2$, then the direct mechanism defined in Figure 4 is IC (check).[15]

*Table*
Preferences in Different States

| | $v_1(x;\theta_1)$ | $v_1(x;\theta_1')$ | $v_1(x;\theta_2'')$ | $v_2(x;\theta_2)$ | $v_2(x;\theta_2')$ |
|---|---|---|---|---|---|
| $d$ | 0 | 1 | 1 | 2 | 1 |
| $d'$ | 1 | 1 | 1 | 1 | 2 |
| $d''$ | 2 | 0 | 0 | 0 | 0 |
| $o$ | −5 | −5 | 5 | −5 | 5 |
| $v$ | 0 | 0 | 0 | 0 | 0 |

*Figure 4*
A Direct Revelation Mechanism

| | | Agent 2 | |
|---|---|---|---|
| | | $\theta_2$ | $\theta_2'$ |
| Agent 1 | $\theta_1$ | $d$ | $d''$ |
| | $\theta_1'$ | $d'$ | $d'$ |
| | $\theta_1''$ | $v$ | $o$ |

Now consider the following pair of deceptions:

$$\alpha_1(\theta_1) = \theta_1', \quad \alpha_1(\theta_1') = \theta_1', \quad \text{and} \quad \alpha_1(\theta_1'') = \theta_1;$$
$$\alpha_2(\theta_2) = \theta_2 \quad \text{and} \quad \alpha_2(\theta_2') = \theta_2.$$

It is easy to see that $\alpha = (\alpha_1, \alpha_2)$ is a Bayes–Nash equilibrium of the direct revelation mechanism given in Figure 4 (check). Furthermore, it is an equilibrium even if both agents have an intrinsic preference for honesty. This is because all types that are lying get a higher expected utility than by telling the truth. However, $\alpha$ is not an equilibrium under the weak social norm against lying. The deception of agent 1 is obvious, while the concealed deception $\alpha_1'(\theta_1) = \alpha_1'(\theta_1') = \alpha_1'(\theta_1'') = \theta_1'$ would give him exactly the same expected utility.

To see the reverse, that an equilibrium under the weak social norm against lying is not necessarily an equilibrium when agents have an intrinsic preference for honesty, is much simpler. Assume that all types are completely indifferent between all alternatives, and look at the direct revelation mechanism in Figure 4. If both agents

[15] The transfer part $t$ is identically zero.

have an intrinsic preference for honesty, this mechanism has exactly one equilibrium, where both agents are always truthful. Under the weak social norm against lying, however, it has many equilibria where both agents use concealed deceptions. By definition these could only be broken by material gains, of which there are none.

Our example reveals an interesting and not so obvious fact. Namely, under incomplete information it may not be logically coherent to assume that agents have an intrinsic preference for honesty, unless one also assumes that they would select a concealed deception rather than an obvious one. It feels almost contradictory that someone who has an intrinsic preference for honesty does not care if others believe that he is lying. Our example shows that this can happen if one does not explicitly require that agents prefer concealed deception. On the other hand, there does not seem to be a similar flaw in assuming that agents prefer concealed deception, and yet do not have an intrinsic preference for honesty. Only in this sense is a weak social norm against lying a weaker assumption than intrinsic preference for honesty.

## *5 Further Connections with the Literature*

The implementation literature that studies the effect of intrinsic preference for honesty is growing fast (see Lombardi and Yoshihara, 2011; Dutta and Sen, 2012; Doghmi and Ziad, 2013; Kartik, Tercieux, and Holden, 2014; Saporiti, 2014; Lombardi and Yoshihara, 2013; Ortner, 2015; Doghmi and Ziad, 2015; to name a few). The focus, however, is almost entirely on the complete-information case, whereas we have shown here that the incomplete-information case has its own peculiar features. Our result applies only in the simplest case of one principal and multiple agents. On the other hand, Saran (2011) has recently shown that in this case the revelation principle is robust to deviations from the rational framework, which gives scope for possible generalizations of our result, but if there are many competing principals, then the revelation principle itself may no longer hold (Epstein and Peters, 1999; Peters, 2001), and therefore it is not likely that our result will generalize either.

In addition to intrinsic preference for honesty, the burgeoning literature on behavioral economics is full of results that have potential applications in mechanism design (see DellaVigna, 2009; Diamond and Vartiainen, 2007). Two that are important here are those obtained in the literature on lying costs (Kartik, 2009; Abeler, Becker, and Falk, 2014) and contagiousness of norm violations (Diekmann, Przepiorka, and Rauhut, 2011; Houser, Vetter, and Winter, 2012). The literature on lying costs assumes that lying has a real cost in utility terms. This, however, does not imply that IC is not needed, unless a lower bound greater than zero for the cost is known for sure.[16] Whether this lower bound can be known or not has no essential effect on our result. We can still use exactly the same mechanism to implement

[16] There are also other reasons why IC is not always necessary for implementation; see Glazer and Rubinstein (2014) and Kartik, Tercieux, and Holden (2014) for example.

any SCF that is IC net of this minimal lying cost. By this we mean that, assuming others are truthful, the utility from lying minus the minimal cost of lying cannot exceed the utility from truth-telling.

On the other hand, if norm violations are contagious in the sense that agents do not care whether they are lying or not if others do so, or perhaps even favor lying in that case, then our mechanism does not work anymore. This is a direct consequence of the example studied in section 4. Here the only reason that the strategy profile where both agents are lying under both types is not an equilibrium (in the revelation part) is that agents rather want to tell the truth without changing the outcome. Neither one is willing to do this, since the other is lying if norm violations are contagious. This is why the assumption that the norm originates outside the mechanism is crucial.

Another obvious possibility would be to consider models of bounded rationality (see Rubinstein, 1998). One notable paper in this respect is de Clippel, Saran, and Serrano (2014), who get a similar result to the one we do here by assuming bounded depth of reasoning.[17] Our result, however, derives from exactly opposite reasoning. We assume that agents are rational in the standard sense and, in addition, take the prevailing social norms against lying into account when making decisions. Thus, agents in our model could be characterized as superrational rather than boundedly rational or behavioral.

The paper that is closest to ours in spirit is Matsushima (1993). He shows that if a condition called *no consistent deceptions* (NCD) is satisfied, then IC is a full characterization of Bayesian-implementable SCFs in any economic environment. In contrast to our paper, and despite the name, NCD is in fact a restriction on the information structure $p(\cdot)$ rather than on the deceptions. Furthermore, it is not entirely obvious why some information structures should be ruled out a priori.

## 6 Conclusion

We have argued that IC is a full characterization of Bayesian-implementable SCFs under the weak social norm against lying. By weak norm we mean that agents lie but do not want to get caught doing so. A strong norm, in contrast, could be defined as something that ultimately generates an intrinsic preference for honesty. This is what all other papers on mechanism design and honesty seem to assume (e.g., Matsushima, 2008a,b; Dutta and Sen, 2012; Lombardi and Yoshihara, 2011; Doghmi and Ziad, 2013; Kartik, Tercieux, and Holden, 2014; Korpela, 2014). Our result holds also in the case of two players, which is usually very different from the general case, and moreover, it implies that there is nothing operationally wrong in relying on the revelation principle. An interesting open question is whether this result holds also for social-choice correspondences.[18]

[17] See also de Clippel (2014) for a great introduction to behavioral implementation.

[18] See Jackson (1991) for a definition of social-choice correspondences in a Bayesian mechanism design setting.

One nice feature of the mechanism that we use in the proof is that it has exactly one equilibrium. Thus implementation cannot fail simply because agents try to coordinate in different equilibria. This is good because the possibility that agents do not coordinate into the same equilibrium grows dramatically as the number of agents increases. Still, we are using the standard integer-games construction, so the result can be considered only suggestive, although it is hard to avoid integer games when deriving general characterization.

## *Appendix: Proof of the Theorem*

Suppose that $f$ is IC. We construct a mechanism that implements $f$ in Bayes–Nash equilibrium with a social norm against lying. Let $\Delta_i = \{\delta_i \mid \delta_i : \Theta_i \to \Theta_i\}$, and assume that agents are placed on a circle in such a way that 1 is between $n$ and 2. Thus, the predecessor of agent 1 is agent $n$, and we denote $0 = n$ accordingly. The message space of agent $i$ is $M_i = \Theta_i \times \{\Theta_{i-1} \cup \{T\}\} \times \Delta_i \times \{0,1,2\} \times \mathbb{N}$,[19] with a typical message denoted by $m^i = (\theta_i, \theta_{i-1}, \delta_i, k_i, n_i)$, while the outcome function $\mu$ is defined via the following rules:[20]

(1) If $m_4^i = 0$ for all $i \in N \setminus \{j\}$, and $m_4^j = 0$ or $[m_4^j = 1$ and $m_2^j = T]$, then

$$\mu(m) = f(m_1^1, \ldots, m_1^i, \ldots, m_1^n).$$

(2) If there exists $j \in N$ such that $m_4^i = 0$ for all $i \in N \setminus \{j\}$, but $m_4^j = 1$ and $m_2^j \neq T$, then the outcome is determined as in rule (1) except that

$$\mu(m)_j = \begin{cases} f(m_1^1, \ldots, m_1^i, \ldots, m_1^n)_j + \epsilon & \text{if } m_2^j = \theta_{j-1} \neq m_1^{j-1}, \\ f(m_1^1, \ldots, m_1^i, \ldots, m_1^n)_j - A & \text{if } m_2^j = \theta_{j-1} = m_1^{j-1}. \end{cases}$$

(Here $A$ must be large enough so that agent $j$ will never want to deviate from rule (1) if there is a possibility of incurring the loss.)[21]

(3) If there exists $j \in N$ such that $m_4^i = 0$ for all $i \in N \setminus \{j\}$, but $m_4^j = 2$ and $m_3^j = \delta_j$, then

$$\mu(m) = f(m_1^1, \ldots, \delta_j(m_1^j), \ldots, m_1^n).$$

(4) In all other cases, denote $k = \max\{i \mid m_5^i \geq m_5^j \text{ for all } j = 1, \ldots, n\}$, and let the outcome be as in rule (1) except that[22]

$$\mu(m)_k = f(m_1^1, \ldots, m_1^n)_k + \lambda.$$

---

[19] For comparison with our previous notation, here $Q_i = \{\Theta_{i-1} \cup \{T\}\} \times \Delta_i \times \{0,1,2\} \times \mathbb{N}$. This set can be replaced with any abstract set as long as the correspondence between messages and outcomes is preserved.

[20] Notice that $f(m_1^1, \ldots, m_1^i, \ldots, m_1^n)$ is an $(n+1)$-tuple. We denote the first component of this by $f(m_1^1, \ldots, m_1^i, \ldots, m_1^n)_0$ (the decision), so that $f(m_1^1, \ldots, m_1^i, \ldots, m_1^n)_j$ is the transfer made by agent $j$.

[21] Denote $p' = \min\{p(\theta) \mid \theta \in \Theta\}$. One possibility is to select $A$ and $\epsilon$ in such a way that $p'A > \epsilon$.

[22] Ties can be broken, for example, in favor of the largest number.

*Remark.* Type announcements can be divided into two disjoint sets – permutations and the rest. In our mechanism agent $i$ can use the set $\Theta_{i-1} \cup \{T\}$ to indicate that he believes agent $i-1$ is truthful in general ($T$) or lying under some type ($\Theta_{i-1}$).[23] If agent $i-1$ is using any type announcement that is not a permutation under rule (1), then it is profitable for agent $i$ to deviate to rule (2), since some type of agent $i$ is never part of any message. On the other hand, agents do not want to use deceptions that are permutations (obvious deceptions) under rule (1) either, other than the revelation strategy, since it is possible to obtain exactly the same outcome with a concealed deception under rule (3). The weak social norm against lying guarantees that all agents prefer a concealed deception to an obvious one.

Next we prove that this mechanism implements $f$. The proof proceeds in the following way. First we show that a strategy profile $\sigma = (\sigma_1, \ldots, \sigma_n) = ((\alpha_1, s_1), \ldots, (\alpha_n, s_n))$, where $\alpha_j : \Theta_j \to \Theta_j$ is the deception and $s_j : \Theta_j \to \{\Theta_{j-1} \cup \{T\}\} \times \Delta_j \times \{0,1,2\} \times \mathbb{N}$ is the auxiliary strategy, can be a Bayes–Nash equilibrium with a social norm against lying if, and only if, the outcome is always selected using rule (1). After this we show that at any equilibrium under rule (1) all deceptions $\alpha_i$ must be identity mappings and therefore the outcome coincides with the SCF $f$. This is done by verifying two things: (a) if some agent, say agent $i$, were using a concealed deception, then agent $i+1$ would rather deviate to rule (2), and (b) it is not possible that agents are using obvious deception either, since then they would rather deviate to rule (3) themselves.

PROOF Suppose that $\sigma$ is a Bayes–Nash equilibrium with a social norm against lying. The outcome is never selected using rule (2), (3), or (4). First of all, in any state the outcome cannot be selected using rule (4), since some agent could improve his position by increasing the fifth component and leaving everything else intact. This would guarantee that the outcome is exactly as before except that it would be better under rule (4). Furthermore, there is always some agent who can deviate from rules (2) and (3) to rule (4) and improve his position in so doing.[24] Therefore, all equilibria of this mechanism (if any) must select the outcome using rule (1) in all possible states.

The fact that $f$ is IC guarantees that if the outcome is always selected using rule (1) and all agents are truthful in all states, then $\sigma$ is a Bayes–Nash equilibrium with a social norm against lying. Thus, there is at least one good equilibrium, since in this case the outcome coincides with $f$ by definition. Moreover, the mechanism does not have any other equilibria besides this one. To argue this, let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be any profile of deception that can be used as a part of $\sigma$. Suppose that agent $j$ is not truthful, so that $\alpha_j$ is not an identity mapping, and that the outcome is always selected using rule (1). There are two possibilities: (a) $\alpha_j$ is a concealed deception or (b) $\alpha_j$ is an obvious deception. In the first case agent $j+1$ could deviate prof-

[23] This does not mean that agent $i$ has to know the type of agent $i-1$. Instead, agent $i$ only needs to have a belief that certain types are never part of any message sent by agent $i-1$.

[24] Strictly speaking, this requires that $\lambda$ be large enough.

itably to rule (2), since there must be some type, say $\theta_j \in \Theta_j$, that does not belong to the range of $\alpha_j$.[25] This means that agent $j+1$ could announce $\theta_j$ as the second component and 1 as the fourth component of his strategy and deviate to rule (2) without any fear of suffering the penalty $A$. The outcome would be exactly as before except that the transfer would be higher. In the second case it is agent $j$ himself that wants to deviate. This is because he can start telling the truth and deviate to rule (3) by adjusting the third and fourth components of his strategy in such a way that he gets exactly the same outcome as before. He prefers to do this because of the social norm against lying. This completes our proof. *Q.E.D.*

### *References*


Abeler, J., A. Becker, and A. Falk (2014), "Representative Evidence on Lying Costs," *Journal of Public Economics*, 113, 96–104.

Aumann, R. J. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1(1), 67–96.

Bergin, J. (1995), "On Some Recent Results in Incomplete Information Implementation," *The Canadian Journal of Economics / Revue canadienne d'Economique*, 28(1), 108–138.

Dasgupta, P., P. Hammond, and E. Maskin (1979), "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *The Review of Economic Studies*, 46(2), 185–216.

d'Aspremont, C., and L.-A. Gérard-Varet (1979), "Incentives and Incomplete Information," *Journal of Public Economics*, 11(1), 25–45.

— and — (1982), "Bayesian Incentive Compatible Beliefs," *Journal of Mathematical Economics*, 10(1), 83–103.

de Clippel, G. (2014), "Behavioral Implementation," *The American Economic Review*, 104(10), 2975–3002.

—, R. Saran, and R. Serrano (2014), "Mechanism Design with Bounded Depth of Reasoning and Small Modeling Mistakes," Working Paper 2014-7, Brown University, Providence (RI).

DellaVigna, S. (2009), "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47(2), 315–372.

Diamond, P., and H. Vartiainen (2007), *Behavioral Economics and its Applications*, Princeton University Press, Princeton (NJ).

Diekmann, A., W. Przepiorka, and H. Rauhut (2011), "Lifting the Veil of Ignorance: An Experiment on the Contagiousness of Norm Violations," Discussion Paper 2011-004, Nuffield Centre for Experimental Social Sciences, Oxford.

Doghmi, A., and A. Ziad (2013), "On Partially Honest Nash Implementation in Private Good Economies with Restricted Domains: A Sufficient Condition," *The B.E. Journal of Theoretical Economics*, 13(1), 415–428.

— and — (2015), "Nash Implementation in Private Good Economies with Single-Plateaued Preferences and in Matching Problems," *Mathematical Social Sciences*, 73, 32–39.

Duggan, J. (1995), "Bayesian Implementation," Ph.D. dissertation, California Institute of Technology, Pasadena (CA).

Dutta, B., and A. Sen (2012), "Nash Implementation with Partially Honest Individuals," *Games and Economic Behavior*, 74(1), 154–169.


[25] Here we need the assumption that $\Theta_j$ is finite. Indeed, to be exact, we have used this assumption already. For the existence of $\epsilon$ and $A$ we need $\Theta$ to be finite.

Epstein, L. G., and M. Peters (1999), "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 88(1), 119–160.

Feldman, A. M., and R. Serrano (2006), *Welfare Economics and Social Choice Theory: 2nd Edition*, Springer Science+Business Media, New York.

Fischbacher, U., and F. Föllmi-Heusi (2013), "Lies in Disguise—An Experimental Study on Cheating," *Journal of the European Economic Association*, 11(3), 525–547.

Gibbard, A. (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41(4), 587–601.

Glazer, J., and A. Rubinstein (2014), "Complex Questionnaires," *Econometrica*, 82(4), 1529–1541.

Gneezy, U. (2005), "Deception: The Role of Consequences," *The American Economic Review*, 95(1), 384–394.

Greene, J. D., and J. M. Paxton (2009), "Patterns of Neural Activity Associated with Honest and Dishonest Moral Decisions," *Proceedings of the National Academy of Science of the United States of America (PNAS)*, 106(30), 12506–12511.

Hao, L., and D. Houser (2011), "Honest Lies," Working Paper 11-16, Interdisciplinary Center for Economic Science, George Mason University, Fairfax (VA).

Harris, M., and R. M. Townsend (1981), "Resource Allocation under Asymmetric Information," *Econometrica*, 49(1), 33–64.

Harsanyi, J. C. (1967), "Games with Incomplete Information Played by 'Bayesian' Players, Part I. The Basic Model," *Management Science*, 14(3), 159–182.

— (1968a), "Games with Incomplete Information Played by 'Bayesian' Players, Part II. Bayesian Equilibrium Points," *Management Science*, 14(5), 320–334.

— (1968b), "Games with Incomplete Information Played by 'Bayesian' Players, Part III. The Basic Probability Distribution of the Game," *Management Science*, 14(7), 486–502.

Holmström, B. R. (1977), *On Incentives and Control in Organizations*, Ph.D. dissertation, Stanford University, Cambridge (MA).

Houser, D., S. Vetter, and J. Winter (2012), "Fairness and Cheating," *European Economic Review*, 56(8), 1645–1655.

Hurkens, S., and N. Kartik (2009), "Would I Lie to you? On Social Preferences and Lying Aversion," *Experimental Economics*, 12(2), 180–192.

Jackson, M. O. (1991), "Bayesian Implementation," *Econometrica*, 59(2), 461–477.

Kartik, N. (2009), "Strategic Communication with Lying Costs," *The Review of Economic Studies*, 76(4), 1359–1395.

—, O. Tercieux, and R. Holden (2014), "Simple Mechanisms and Preferences for Honesty," *Games and Economic Behavior*, 83, 284–290.

Korpela, V. (2014), "Bayesian Implementation with Partially Honest Individuals," *Social Choice and Welfare*, 43(3), 647–658.

Lombardi, M., and N. Yoshihara (2011), "Partially-Honest Nash Implementation: Characterization Results," Discussion Paper 43, CCES, Hitotsubashi University, Tokyo.

— and — (2013), "Natural Implementation with Partially Honest Agents in Economic Environments," Working Paper 48294, Munich Personal RePEc Archive (MPRA), Munich.

Matsushima, H. (1993), "Bayesian Monotonicity with Side Payments," *Journal of Economic Theory*, 59(1), 107–121.

— (2008a), "Behavioral Aspects of Implementation Theory," *Economics Letters*, 100(1), 161–164.

— (2008b), "Role of Honesty in Full Implementation," *Journal of Economic Theory*, 139(1), 353–359.

Mazar, N., O. Amir, and D. Ariely (2008), "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance," *Journal of Marketing Research (JMR)*, 45(6), 633–644.

Mookherjee, D., and S. Reichelstein (1990), "Implementation via Augmented Revelation Mechanisms," *The Review of Economic Studies*, 57(3), 453–475.

Myerson, R. B. (1979), "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47(1), 61–74.
— (1981), "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.
— (2008), "Perspectives on Mechanism Design in Economic Theory," *The American Economic Review*, 98(3), 586–603.
— (2009), "Learning from Schelling's *Strategy of Conflict*," *Journal of Economic Literature*, 47(4), 1109–1125.
Ortner, J. (2015), "Direct Implementation with Minimally Honest Individuals," *Games and Economic Behavior*, 90, 1–16.
Palfrey, T. R. (1992), "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design," in: J.-J. Laffont (ed.), *Advances in Economic Theory: Sixth World Congress, Vol. I*, Cambridge University Press, Cambridge, pp. 283–323.
— and S. Srivastava (1993), *Bayesian Implementation*, Harwood Academic Publishers, Chur.
Peters, M. (2001), "Common Agency and the Revelation Principle," *Econometrica*, 69(5), 1349–1372.
Radner, R., and D. Ray (2003), "Robert W. Rosenthal," *Journal of Economic Theory*, 112(2), 365–368.
Rosenthal, R. W. (1978), "Arbitration of Two-Party Dispute under Uncertainty," *The Review of Economic Studies*, 45(3), 595–604.
Rubinstein, A. (1998), *Modeling Bounded Rationality*, The MIT Press, Cambridge (MA).
Saporiti, A. (2014), "Securely Implementable Social Choice Rules with Partially Honest Agents," *Journal of Economic Theory*, 154, 216–228.
Saran, R. (2011), "Menu-Dependent Preferences and Revelation Principle," *Journal of Economic Theory*, 146(4), 1712–1720.
Schelling, T. C. (1960), *The Strategy of Conflict*, Harvard University Press, Cambridge (MA).

Ville Korpela
Turku School of Economics
Rehtorinpellonkatu 3
20500 Turku
Finland
vipeko@utu.fi