Durability, Regulation, and Incentives: Essays in Applied Microeconomics

Salem Saljanin

Doctoral Thesis

Accepted by the Faculty of Economics in partial fulfilment of the requirement for the degree of *Doctor Rerum Politicarum* at the University of Hamburg

Thesis Committee

Chairman: Prof. Dr. Anke Gerber, University of Hamburg

1st Examiner: Prof. Dr. Wilhelm Pfähler, University of Hamburg

2nd Examiner: Prof. Dr. Manfred J. Holler, University of Hamburg

Submitted: May 31, 2010

Date of Defence (Disputation): July 4, 2011

Mojim roditeljima

Acknowledgments

This thesis has benefited from comments and criticisms of many people. I am particularly grateful to my supervisor Wilhelm Pfähler for his continuous support and profound advice. He managed to provide intellectual freedom with the right guidance.

I am indebted to Johannes Bruder, Thomas Haertel, Adi Husovic, Jasmina Husovic, Annette Hofmann, Lars Just, Leon Leschus, Helena Meier, Theodoros Sarigiannis, and Evelyn Schäfer for feedback and support throughout the process of writing this thesis.

Contents

1 Introduction

Durability is one of the main characteristics of many goods and services which represent a large fraction of output of developed economies. Despite its apparent relevance, the topic of durability has not received very much attention from the economics profession.

Durable goods pose a number of specific question for microeconomists. First, there is a set of questions concerned with the durability choice and the related issue of "planned obsolescence." Second, there is a set of questions around the topic of timing. How are current prices and the choice of other marketing instruments affected by a firm's future actions? For example, should a producer of a durable good sell or just rent? Finally, there is a complex of issues around adverse selection. The seminal contribution of Akerlof (1970) used the case of second-hand cars as a motivating example concerning the problem of quality uncertainty and its consequences. Until recently, however, Akerlof's argumentation in the context of durable goods has been mainly ignored.

Intrinsically tied to the questions concerning durability are government interventions. Regulation is one of the main issues in (modern) economics. The ubiquitousness of regulation concerning various activities around the world is apparent, yet standard economic theory predicts that it should be rather uncommon, that is, the government should refrain from regulation—except in very specific cases of market failure. According to traditional Pigouvian welfare economics, government regulation makes perfect sense. Markets may fail because of externalities, asymmetric information, and market power, and governments need to regulate them to counter these failures.

This perspective, however, has lost its popularity over the last few decades, under pressure from the law and economics tradition originating with the seminal contribution of Coase (1960). According to Coase's view, contracts are a (perfect) substitute for regulation. If the costs of transaction are not prohibitively high, no government regulation is necessary. Yet, we even observe extensive government regulation of contracts themselves. One of the

many questions in this context is: Why do we observe the pervasiveness of certain modes of regulation? Further, why do we observe seemingly counterproductive interferences in the working of markets?

Within this wide spectrum of questions concerning durability and regulation, the present analysis highlights seven specific questions spanning from the question whether to sell or rent a durable good to the role of incentives in the choice of originality. Each of the seven chapters presented in the following is a self-contained essay.

Chapter 2 deals with the question whether a durable-good monopolist should sell or rent under conditions of uncertainty. Contrary to the literature based on the assumption of certainty, we show that selling without commitment may dominate selling with commitment under conditions of moderate uncertainty, that is, uncertainty without the possibility of a shutdown in the future. Allowing for the possibility of a shutdown, we show that selling without commitment may even dominate renting.

In chapter 3 we offer three new reasons for the phenomenon of planned obsolescence without referring to time inconsistency as the classical explanation. The driving force behind all three reasons are informational concerns. These concerns are present in a broad range of situations leading to the following three rational reasons for the practice of planned obsolescence: (1) learning by the firm through repeat purchases, (2) option value to the firm and/or consumers, and (3) signaling innovative ability. Once all the factors are taken into account, practicing planned obsolescence may be profitable and even increase social welfare.

Chapter 4 deals with an important question concerning regulation: Why are quantity restrictions so ubiquitous? Economists generally prefer taxes. We argue that quantity regulation allows for enforcement of complementary regulations at no significant additional cost. In contrast, the regulation by taxes usually requires additional effort for every regulated activity separately. When the enforcement of a tax rule is sufficiently costly, quantity regulation may become desirable from a social point of view even if it eliminates some efficient conduct. The general argument is also useful in the justification of

clear rules (refereed to as bright-line rules).

Why do we often observe soft budgets (the refinancing of loss-making enterprises)? That is the topic of chapter 5. We argue that various projects of economic agents (e.g., firms) are often the more profitable the more confident the agents are that the principal (e.g., a government) is competent or supports the project or both. The type of the principal is often not known to the agents. The confusion becomes even worse in times of a reform and/or crisis. Soft budgets may then be used to signal support and/or competence by the principal.

Chapter 6 deals with another important topic from regulatory economics: with the role of intermediaries. We analyze the consequences of (possibly) fraudulent information intermediaries in a market for credence goods characterized by monopolistic competition. The honest counseling may require side payments from the firms to the information intermediaries. The analysis suggests that strict regulation of the relationship between firms and intermediaries may be crucial to make these markets operate more efficiently.

Finally, chapter 7 analyzes a simple model concerning the choice of originality to solve a given task. This is of great importance for regulatory activities for at least two reasons: (1) imposed restrictions will often change the general approach to a regulated task, and (2) regulations itself can be of varying originality. A less original approach is more likely to succeed, but the potential benefit is at the same time generally smaller. On the other side, employing a more original approach is less likely to succeed, but it may lead to much higher benefits. The model predicts that the level of originality may increase with the number of potential beneficiaries, decreases with the number of people to be persuaded, and generally increases with the ability level of the agent. The incorporation of uncertainty concerning the exogenous variables leads to ambiguous predictions. The role of originality as a signal of ability or talent is also discussed.

Taken together, these essays use standard economic tools to shed light on some important questions concerning durability and regulation. They are at the intersection of several branches of economics such as Industrial Organi-

zation, Regulatory Economics, Law and Economics, and Political Economy.

The models in the following essays are always the simplest formalizations of the underlying phenomenon. The reason for this is the belief that a good model is supposed to reveal the essential components of the problem at hand; it should be reduced to just those pieces that are required to make the point. As Einstein (1934) said "It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience."

2 Durable-Good Monopoly under Uncertainty: A Case for Non-Commitment and Selling

Abstract

This paper shows that uncertainty alters some of the basic propositions that dominated the literature on durable-good monopolies. We demonstrate that uncertainty can make commitment to an output path undesirable for a monopolistic seller of a durable good. The paper also shows that selling can dominate renting in the case of a shutdown possibility in the future. The results of the paper correspond to the fact that many durable goods are sold on the basis of simple sales contracts.

Keywords: *Durable good, commitment, uncertainty, monopoly*

JEL classification: D42, L12, D80

2.1 Introduction

Coase (1972) has first described the commitment problem facing a monopolist seller of a durable good in a deterministic environment. He conjectured that a monopolist competes with future incarnations of himself. Even though the most profitable course of action is to sell the monopoly quantity immediately and then never sell again, the monopolist cannot resist selling more to consumers with lower valuations of the good in the future once the monopoly profit is earned, that is, he acts in a time-inconsistent way. Subgame perfection condemns the monopolist to low profits. This can lead to a situation with no market power at all. Thus, a seller of a durable good looks for ways to make precommitments about production levels. The Coase conjecture was later formally proved by Stokey (1981) and Gul et al. (1986).

Coase has discussed several ways to avoid the problem. The most prominent alternative is to rent rather than sell the good. Much of the literature following Coase on durable good markets focuses on other instruments which a

durable goods monopolist can use in order to avoid the negative consequences of the problem of non-commitment to a certain output path (equivalent to a certain price path under certainty).

Since even casual observation reveals that many durable goods are often sold on the basis of simple simple sales contracts, one may ask why many firms obviously don't perceive time-inconsistent behavior and the selling strategy as threats to their profitability. There are several possible explanations for this fact. One obvious answer is that the monopolist can use special contractual arrangements in order to commit himself. The best example for such a behavior are best-price clauses, but not all sellers of durable goods use them. Bulow (1986) and Bucovetsky and Chilton (1986) have shown that the threat of entry of new competitors leads to an incentive to simultaneously sell and rent in order to discourage the entry of competitors by making it less profitable through the reduced residual demand. Rust (1986) has offered another explanation based on the fact that renting a durable good creates a moral hazard problem. That is the case because the monopolist is often unable to monitor the level of maintenance exercised by the renter.

This article offers a simple explanation based on uncertainty. It goes almost without saying that uncertainty is a fundamental element of real economic decision making. One of the main consequences of uncertainty is that it creates a trade-off between commitment and flexibility. A firm which is flexible to delay a certain decision until it learns more about the uncertain environment may greatly benefit from doing so. The objective of this paper is to characterize the trade-off between commitment and flexibility for the case of a durable-good monopolist; it is shown that the incorporation of uncertainty can lead to exactly the opposite conclusions concerning the desirability of commitment and renting. Using a simple model, this paper shows that with levels of uncertainty beyond a certain threshold, a committed monopolist seller makes lower profits than a non-committed seller. It is also shown that selling may dominate renting if there is a possibility of a shutdown in the future, that is, the chance of no additional output in future periods.

Uncertainty is particularly relevant in the case of durable goods, for their

usage typically extends far into the future. The literature on durable goods markets is, however, predominantly concerned with models based on the assumption of certainty. Exceptions are (among others) Bhatt (1989), Biehl (2001), Board (2008) and Usategui (2007). Most relevant to this paper is the work of Bhatt (1989), who shows that in the presence of demand uncertainty and risk-averse firms, the monopolistic supplier of a durable good has an incentive to sell rather than rent in order to transfer the risk of future demand shocks to consumers.

Another literature that is related to this paper is the one thoroughly presented and surveyed by Dixit and Pindyck (1994) on investment under uncertainty. This literature, however, generally does not focus on the strategic aspects of decisions. In the language of this literature the lack of commitment by the monopolist could be interpreted as a real option. The real option gives the monopolist the possibility of adapting the output decision in the second period to the circumstances in that period. If the amount of uncertainty is high enough, the real option has a positive value.

The rest of the paper is organized as follows. Section 2.2 describes the basic assumptions of the model. Section 2.3 presents the results for the case without the possibility of a shutdown. The results for the case with a chance of a shutdown are presented in section 2.4. A short conclusion follows.

2.2 The Basic Model

The model presented in this section is suited to capture situations in which a durable-good monopolist faces a trade-off between commitment and flexibility: on the one hand, the monopolist has an incentive to avoid commitment in order to get better information about the profitability of the market; on the other hand, he is tempted to eliminate the problem of time inconsistency by committing in the first period to output levels for future periods, but if demand turns out to be high, the monopolist is stuck with lower profit.

The standard two-period framework of Bulow (1986) is used. Technically, the two-period framework is used only because it permits the calculations by backward induction very easily. The same type of analysis can be carried

out in a multiple-period framework with a known terminal date. Such an extension would, however, not change the qualitative nature of the results in a fundamental way.

The monopolist offers a single homogeneous product. The firm chooses an output level in each period x_t, $t \in \{1, 2\}$. The output of the monopolist is either rented or sold. The quantity of first-period output still available in period 2 is x_1, that is, the good is assumed to be perfectly durable. In the second period, a perfect resale market shall exist in which the sold units of the durable good from the first period can be resold to other consumers. The essential difference between the seller and the renter is that if a renter "produces too much" he suffers the capital loss on old units; thus, the costs of additional production are internalized. If a seller overproduces, the losses are suffered by old purchasers whose welfare the seller does not internalize.

For simplicity, assume that there are no fixed cost of production, and the marginal cost are constant which implies that they can be, without loss of generality, set equal to 0—at least for the first period. Discounting is ignored, without loss of generality. Suppose, further, that no other firm can enter the market in period 1 and 2, that is, there is no threat of potential competition.

The environment in the second period is not assumed to be known with certainty at the beginning of period 1. Uncertainty is incorporated in the form of an additive shock u in the second period. The shock can be interpreted as an effect of a change in demand and/or cost of production. At the end of the first period, the monopolist observes the realization of u, denoted by $\tilde{u}$. Thus, output decisions at $t = 1$ are made under uncertainty about u and the output decision at $t = 2$ are made under complete information. The firm is assumed to be risk neutral, implying that the monopolist maximizes expected profit. The consumers' willingness to pay per period of service is uniformly distributed over the interval $[0, 1]$. This implies a linear rental demand function $x(p) = 1 - p$. Thus, the inverse rental demand in each period is assumed to be a linear function of the total stock available for use in the period. If we assume (as is standard) the stock of output is zero initially in the absence of any production, the stock in period 1 is equal to first

period production. Denoting price by p, the inverse rental demand curves for services in period 1 and 2 can then be written as:

$$p_1 = 1 - x_1, \tag{1}$$

$$p_2 = 1 - (x_1 + x_2) + u. \tag{2}$$

We assume that u is a continuous random variable with $E[u] = 0$ and variance $Var[u] = \sigma^2 > 0$. The density $f(u)$ is defined on the support $[\underline{u}, \overline{u}]$, with $\overline{u} > \underline{u}$. We consider two cases. In the first case, $\underline{u}$ is sufficiently large such that the second-period output will always be positive. This rules out the possibility of uncertainty serving as a commitment to no output in the second period. In the second case not all states of the world are sufficiently beneficial so that the monopolist will always sell or rent a positive amount of the good in the second period. That is, a shutdown possibility is considered. This possibility can be motivated in a variety of ways. For example, there could be a severe drop in demand due to an economic crisis.

Consumers and the monopolist anticipate that the monopolist will no longer be interested in the value of the products already sold. The monopolist maximizes expected profits by setting quantities. The firm has an opportunity to commit to a certain sales path in both periods at the beginning of period 1 before the realization of u is known (commitment), but may choose to postpone the output decision for the second period until period 2, when the realization of u is known (no commitment).

2.3 The Case Without a Shutdown Possibility

This section examines the trade-off between commitment and flexibility in the case of a monopolistic seller of a durable good. The relevant equilibrium concept for the analysis of an uncommitted seller is that of subgame-perfect equilibrium. That is, we use backward induction to compute the equilibrium. The monopolist does not care about the loss in value of previously sold units in the second period due to his additional output in that period, which

implies that he solves:

$$\max_{x_2} \pi_2 := \left[\left(1 - \left(x_1 + x_2 \right) + \tilde{u} \right) x_2 \right], \tag{3}$$

Solving the first-order condition yields the optimal second-period output (as a function of $\tilde{u}$ and the first-period output):

$$x_2^* := \arg \max_{x_2} \pi_2 \left(x_2, x_1, \tilde{u} \right) = \frac{1 + \tilde{u} - x_1}{2}. \tag{4}$$

If the consumers are rational, they will take into account the firm's behavior in the second period when purchasing first-period output. Using π^{NC} to denote profit of the non-committed monopolist the monopolist maximizes:

$$\max_{x_1} \pi^{NC} := \left[\left(1 - x_1 \right) x_1 + \left(1 - \left(x_1 + x_2^* \right) \right) \left(x_1 + x_2^* \right) \right]. \tag{5}$$

Solving the first-order condition for x_1 and substituting back in (4) yields the equilibrium output levels

$$x_1^{NC} = \frac{2}{5} \text{ and } x_2^{NC} = \frac{1}{2} \left(\frac{3}{5} + \tilde{u} \right). \tag{6}$$

Since $E\left[u\right] = 0$, the expected values are easily computed:

$$E\left[x_1^{NC} \right] = \frac{2}{5} \text{ and } E\left[x_2^{NC} \right] = \frac{3}{10}. \tag{7}$$

The profit of the monopolist is *ex post*

$$\pi^{NC} = \frac{1}{4} \left(1 + \tilde{u} \right)^2 + \frac{1}{5}. \tag{8}$$

Note that this profit function is continuous, monotonically increasing, and convex in the realization of u. An increase in $\tilde{u}$ has a more-than-proportional effect on profit because a firm with market power responds to higher demand by increasing both output and prices.

Since $E\left[u^2\right] = Var\left[u\right]$, the expected value of (8) is

$$E\left[\pi^{NC}\right] = \frac{1}{4}\left(\frac{9}{5} + \sigma^2\right). \tag{9}$$

This result indicates that the expected value of profit is increasing in the variance of demand.

A committed monopolistic seller of a durable good determines his output levels for the first and the second period already in the first period. Thus, Nash equilibrium is used as the relevant equilibrium concept. Using π^C to denote the profit of a monopolist with commitment, the firm then maximizes:

$$\max_{x_1,x_2} E\left[\pi^C\right] := E\left[\left(1 - x_1\right)x_1 + \left(1 - x_1 - x_2 + u\right)\left(x_1 + x_2\right)\right]. \tag{10}$$

Solving the first-order conditions for x_1 and x_2 yields the equilibrium output levels:

$$x_1^C = \frac{1}{2} \text{ and } x_2^C = 0, \tag{11}$$

Note that the monopolist does not supply additional units in the second period.

The profit of the monopolist has the expected value

$$E\left[\pi^C\right] = \frac{1}{2}. \tag{12}$$

Thus, uncertainty has no effect on the output levels and profit.

The comparison of (9) and (12) gives rise to the following

Proposition 1. *The expected profit of a monopolist seller with no commitment is higher than the expected profit of a monopolist seller with commitment for levels of σ^2 beyond a certain threshold.*

Proof. This can be easily shown using the expressions for the expected levels of profit under a monopolist seller with and without commitment derived in (9) and (12). Direct calculations show that

$$E\left[\pi^{NC}\right] - E\left[\pi^{C}\right] > 0 \; for \; \sigma^2 > \frac{1}{5}.$$

□

This result suggests why many producers of durable goods don't seem to worry much about the negative consequences of a lack of precommitment to future output levels.

It is easy to show that renting remains more profitable than selling without commitment under the assumption of no shutdown in the second period (see Bhatt (1989)). The next section considers the consequences of a shutdown possibility in the second period. We show that the chance of a shutdown can make selling more profitable than renting.

2.4 The Case with a Shutdown Possibility

So far all states of the world were sufficiently beneficial that the monopolist will always sell or rent a positive amount of the good in the second period. This section studies the consequences of a shutdown possibility in the future, that is, the possibility that there is no (additional) supply after the first period. The monopolist will not supply existing—as in the case of renting—or additional output if the realization a shock is sufficiently negative to make additional profits negative. There are several potential (and not mutually exclusive) reasons for a shutdown in the future. Here are some possibilities:

(a) The firm could declare bankruptcy.

(b) Demand could drop severely.

(c) The firm could experience a drastic rise in the cost of production.

The effects of a shutdown depend on how it is introduced. An obvious way would be to widen the support $[\underline{u}, \overline{u}]$ of u, leaving the expected value unchanged at 0, but changing the variance of u (mean-preserving spread), and leaving all other parameters unchanged. This approach leads to a critical

value of u, denoted by u^*, at or below which the monopolist will not sell or rent output. The shutdown possibility exists if $\underline{u} < u^*$. That leads to a shutdown probability of $\int_{\underline{u}}^{u^*} f(u)\, du > 0$. Note that the critical value of u—and therefore the shutdown probability—differs between a monopolistic seller and a monopolistic renter. In both cases, the probability of a shutdown is increasing in the first-period output, but at a different rate.

It is now possible to formulate the following

Proposition 2. *The expected profit of a seller with no commitment can be higher than the expected profit of a renter in the case of a shutdown possibility in the second period.*

Proof. The proof is an existence proof; one single example is sufficient to prove the proposition. For simplicity, assume that u is uniformly distributed on the support $[-\overline{u}, \overline{u}]$. It is straightforward but cumbersome to show that there is a critical value for $\overline{u}$—and therefore a critical value for the variance of the shock u, above which the expected profit from selling without commitment is higher than the expected profit from renting.[1] □

The intuition behind this result is quite simple. The possibility of a shutdown in the future offers the durable-good monopolist another source of commitment to less production in the future, which is not present in a deterministic environment. The chance of a shutdown provides much less benefits to a renter of a durable good. This effect seems to suggest a reason why durable goods tend to be sold rather than rented.

The associated testable prediction is: Selling will be more common than renting for durable goods whose demand and/or cost conditions are characterized by a high variance. This prediction corresponds to the prevalence of simple sales contracts for many durable goods.

[1] A MATHEMATICA file that shows this result is available from the author upon request.

2.5 Conclusion

This paper shows that—contrary to the literature based on deterministic models—commitment concerning future output levels is not always desirable for a durable-good monopolist. If the amount of uncertainty overcomes a certain threshold, he benefits from a lack of commitment. The reason is that non-commitment offers the possibility to react to new information. The result is in accord with the fact that firms in many markets for durable goods don't try to establish commitment to a certain output path. A further result of this paper is that selling can dominate renting in the case of a shutdown possibility.

These results are obtained with a linear demand curve, two periods, and risk neutrality. The incorporation of nonlinearities or an infinite time horizon could introduce some ambiguities, but would not undermine the basic point that commitment and renting can have adverse effects on the profit of a durable-good monopolist in an uncertain environment.

3 Another Look at Planned Obsolescence

Abstract

This paper offers three novel and complementary reasons for the practice of planned obsolescence without referring to time inconsistency. The underlying force behind all three reasons are informational concerns. These concerns are present in a broad range of situations leading to the following three rational reasons for the practice planned obsolescence: (1) learning by the firm, (2) option value to the firm and/or consumers, and (3) signaling commitment to (continued) introductions of innovations. Once all the factors are taken into account, practicing planned obsolescence may be profitable and even increase social welfare.

Keywords: *Planned obsolescence, learning, options, signaling*

JEL classification: D21, L15

3.1 Introduction

Many manufacturers of durable goods (allegedly) make their products less durable than they could be. This phenomenon is commonly known as "planned obsolescence." The case for planned obsolescence is intuitively convincing for most laymen. In economist's terminology, a typical layman would say that incentive to reduce durability and is clearly "too strong." A typical layperson, however, generally ignores that by slashing the lifetime of products in half, consumers need to buy them twice as often but are willing to pay only half as much.

Of course, just as is true for the ordinary man in the street, economists have also long observed and commented the choice of durability of firms. Why would a firm deliberately make its product less durable? There have been several theoretical attempts to model the idea of planned obsolescence. They

are almost exclusively based on a time-inconsistency problem: the optimal ex-post strategy may differ from the optimal ex-ante strategy. Consider a monopolist of a durable good who sells output in each of two periods. In the second period the seller will not internalize how his current behavior affects the value of units previously sold as a buyer does in the first period. Building on the earlier work of Coase (1972), Bulow (1982) has used this logic to show that in such a setting the monopolist will sell "too much" in the second period. The (correct) anticipation of this reasoning leads to lower prices already in the first period. By decreasing the durability of its good, the firm increases demand in the future, thus keeping prices high. Bulow (1986) has shown that a monopolist's choice of durability will be socially inefficient in the case he cannot commit to future prices.

The papers of Choi (2001) and Strausz (2009) have offered an explanation for the phenomenon of planned obsolescence in the form of lower durability without referring to time inconsistency. Choi (2001) has argued that reduced durability may serve as a signal of quality. Building on a related argument, Strausz (2009) has demonstrated that reduced durability increases the frequency of repurchases and, therefore, enables consumers to punish producers faster for a lack of quality. This strengthens the producers incentives to provide adequate levels of quality. Fishman et al. (1993) have shown that planned obsolescence in the form of reduced durability may be necessary for the achievement of technological progress and that a pattern of rapidly deteriorating products and fast innovation may be preferred to long-lasting products and slow innovation.

In this paper, we argue that, in a broad range of environments, planned obsolescence is actually rational without referring to time inconsistency as the underlying force. Firms may rationally practice planned obsolescence because of informational concerns in three different but complementary forms:

1. *Learning.* Planned obsolescence resulting in repeat purchases may open the possibility of learning about the consumers and/or increase the efficiency of the learning process by more frequent observations of prices

and quantities. Thus, planned obsolescence may be seen as a special form of *active learning*.

2. *Option Value*. Planned obsolescence may also be seen as a real option. The real option gives the firms or consumers or both the possibility of adapting their production and consumption decisions in future periods to the information then available. If the potential gain is high enough, this real option may be of high value. In other words, planned obsolescence may be seen as a special form of *passive learning*.

3. *Signaling*. In a variety of situations, consumers do not directly observe the producers' commitment to (continued) introductions of innovations or the fashion cycle. Practicing planned obsolescence, that is, reducing the durability of the product may then signal that the firm is concerned about the "fit"of the product and the consumers needs and preferences in a generally uncertain future.

The above literature on planned obsolescence shows little direct parallel to our arguments based on informational concerns. Reducing the durability of a durable good may be the only way to offer the firms and also the consumers an opportunity to learn something about each other and/or to adapt their environment and to react on the basis of the new information. For that reason, planned obsolescence may therefore actually increase the profits of the firms *and* even social welfare. The above three concerns are of particular importance since durable goods are (often) quite expensive and their usage extends far into the future.

Notice that the proposed explanation of planned obsolescence is not restricted to a monopolistic market; it may also explain planned obsolescence under the condition of competition. This aspect is of importance since, as discussed in a recent paper by Grout and Park (2005), planned obsolescence may also occur in competitive markets.

The paper is organized as follows. Section 3.2 discusses learning as a possible reason for planned obsolescence. The option value of planned obsolescence

is analyzed in section 3.3. Signaling as yet another reason for planned obsolescence is discussed in section 3.4. A short summary follows.

3.2 Learning

It almost goes without saying that the knowledge of firms concerning various factors in their environment is generally far from perfect. For example, the relevant characteristics of buyers, the desirable features of the products, and the market potential of the underlying technology are often quite uncertain. Consequently, firms generally (have to) devote significant resources to acquiring, processing and synthesizing information.

Microeconomic models often assume that individual firms know all relevant aspects of the demand function. One of the justifications of this critical assumption is that even if the true demand function is initially unknown, but remains fixed as time progresses, firms eventually learn all relevant aspects of demand from various trial-and-error experiments with different prices and/or different characteristics of the good (see Clower (1959)). There are several papers trying to formalize the problem of learning and experimentation. For example, McLennan (1984) has considered the case of a seller who learns the demand for his good by selling one unit per period of a non-durable good to a customer who has an unknown reservation price. Hart and Tirole (1988) have analyzed the case of a seller who doesn't know the buyer's willingness to pay for a durable good without referring to the possibility of planned obsolescence.

We can understand immediately, without any formal mathematical reasoning, that the problem with learning through experimentation is that it may not apply in the case of a durable good. In the extreme case of an infinitely durable good and a single seller, the firm and the consumer may interact only once. In other words, the absence of planned obsolescence may severely limit the potential of trial-and-error experiments with different prices and/or different characteristics of the good. This means that planned obsolescence leading to (more) repeat purchases may increase the producer's quality of

information concerning various important factors influencing demand of a durable good. This may increase profits from selling the good and give valuable information concerning modifications and innovations in the future. It may even increase social welfare. The associated testable prediction is: Planned obsolescence will be more common for durable goods whose demand characteristics are unknown to the supplying firms.

3.3 Option Value

An option in the broadest sense is the possibility to take actions or make adjustments at a future point in time, possibly after getting more information on the true state of nature. In a world of gradually resolving uncertainty, maintaining options may be of great value. Reducing the durability may increase the potential to benefit from new information concerning various factors available only as time progresses *for both* producers and consumers of a durable good. Thus, there may be an option value to planned obsolescence. This is especially the case in highly dynamic environments. Note that this is true for the producers as well as the buyers of a durable good.

In order to formalize the basic idea in a simple way, we consider the following two-period model. The model features a (representative) consumer and a risk-neutral monopolist. Without any loss of generality, we ignore discounting in our model. In period 1, the consumer has a unit demand for a good. The good can be durable or not, that is, the supplying monopolist has to decide whether to produce a durable (D) or a non-durable-good (ND) good, and how to price it. We say that a good is durable if it lasts for both periods. The second period is characterized by (possibly multiple sources of) uncertainty. For example, the willingness to pay for the durable good could unexpectedly increase due to a economic boom, a fashion cycle, or totally vanish due to a new and superior product by some other firm.

Note that the assumption of unit demand eliminates the durable goods monopolist's incentive to practice planned obsolescence in order to mitigate the dynamic-consistency problem, which was first analyzed by Bulow (1986). This modeling approach focuses the analysis strictly on the phenomenon of

planned obsolescence.

Let $V > 0$ denote the per period gross benefit the consumer gets as viewed from the perspective of the first period, and let p^i, $i \in \{D, ND\}$, denote the price of the good. The per period gross benefit in the second period is $\overline{V} > V$ with probability $0 < p_{\overline{V}} < 1$, and 0 with probability $0 < p_0 < 1$, where $p_{\overline{V}} + p_0 = 1$. We will assume that $p_{\overline{V}}\overline{V} = V$, that is, we will assume that the expected per period benefit remains constant in the second period. The net benefit (B) of the consumer in the case of a durable good is $B^D = 2V - p^D$, and in the case of a non-durable good is $B^{ND} = V - p^{ND}$. We assume the consumer receives new relevant information at the beginning of the second period, that is, non-durability offers the consumer an advantage in the form of additional information and the opportunity to tailor the consumption decision to this new information. For example, a new firm with a technological innovation could convince the consumer that its new product is much better suited to his particular needs. Yet another possible interpretation can be based on a fashion cycle, that is, the product may become (really) fashionable or go completely out of fashion.

The unit production cost of a durable good is denoted by c^D, whereas the unit production cost of a non-durable good is represented by c^{ND}, where we assume that $c^D > c^{ND} \geq 0$, that is, we assume that a durable version is more costly to produce than a non-durable version. Notice that the condition $c^D > c^{ND}$ includes the possibility of economies of scale in the "supply" of durability.

It is now straightforward to analyze the optimal behavior of the firm. The profit-maximizing price for the durable version of the good is $p^D \leq 2V$, and the profit-maximizing price for the non-durable version is $p_1^{ND} \leq V$ in the first period, $p_2^{ND} \leq \overline{V}$ with probability $p_{\overline{V}}$ in the second period. The associated profit levels are

$$\Pi^D \leq 2V - c^D, \tag{13}$$

and

$$\Pi^{ND} \leq V - c^{ND} + p_{\overline{V}}\left(\overline{V} - c^{ND}\right) = 2V - \left(1 + p_{\overline{V}}\right)c^{ND}, \tag{14}$$

depending on the bargaining power of the firm, which we take as exogenous.

We now state our main conclusions in this section with regard to the private and social optimum. First, we analyze the optimal behavior of the firm.

Proposition 3. *The production of a non-durable good may be more preferable than the production of a durable good from the firm's perspective.*

Proof. The proof by way of an example is straightforward.

□

In other words, planned obsolescence can be profitable from the firm's perspective. This result is quite simple and intuitive.

In the same vein, it can be shown that planned obsolescence may also increase social welfare defined as the the sum of the consumer's net benefit and the firm's profit, as summarized in the following proposition.

Proposition 4. *The production of a non-durable good may be more preferable than the production of a durable good from the perspective of a social planer.*

Proof. The proof by way of an example is straightforward.

□

This approach also makes a testable prediction: Planned obsolescence will be more common in the case of producers of durable goods faced with a highly dynamic environment.

The theory behind real options is developed in striking generality in a book by Dixit and Pindyck (1994).

3.4 Signaling

There is yet another possible reason for the phenomenon of planned obsolescence related to the previous discussion. In many situations, consumers do not know with certainty the generalized technical competence and innovative ability of a firm producing durable goods, that is, they cannot directly observe the producers' ability and commitment to (continued) introductions of innovations in the subsequent period(s). This is of crucial importance in many industries. The uncertainty of the consumers can be related to the size of the expected innovation as well as the innovation itself. Simply stating to buyers whether a new product is forthcoming cannot be relied on to be a truthful representation of the firm's actual activities in the future. It is a widely observed practice to announce new products well in advance of actual market availability, and even announce products that never reach the market. This phenomenon, especially common in the computer industry, has been referred to as "vaporware."

The same holds true for actual or potential competitors. A producer of durable goods may want to signal to competitors its technical competence and innovative ability. For example, a producer of a durable goods may want to deter the entry of new competitors.

The firm's characteristics and/or intentions, however, may be credibly conveyed by manipulating the durability of the product. This approach to planned obsolescence requires that durability is observable–at least in form of an expected value. Practicing planned obsolescence may therefore serve as a signaling device to convey that the firm has a high level of technical competence and innovative ability. This point suggests that planned obsolescence is present in situations where the innovative ability of the firm(s) is not readily recognized.

While we do not present a model in which planned obsolescence serves as a signal of innovativeness, such a model is straightforward to construct. The signaling approach to the phenomenon of planned obsolescence makes the following testable prediction: Planned obsolescence will be more common for producers of durable goods whose buyers and/or competitors are uncertain

of the innovative potential of the respective firms.

3.5 Summary

This paper has presented three novel but complementary reasons for the practice of planned obsolescence without referring to the standard argument of time inconsistency. The basic factor behind all three explanations are informational concerns. These concerns are present in a broad range of situations leading to the following three rational reasons for the practice of planned obsolescence: (1) learning by the firm, (2) option value to the firm and/or consumers, and (3) signaling innovative ability.

4 Enforcement Cost: A Case for Quantity Regulation[2]

Abstract

This paper considers the role of enforcement cost in the choice between quantity restriction and tax regulation. We argue that a regime of quantity regulation allows for enforcement of additional regulations at no significant cost. In contrast, the regulation by taxes usually requires additional effort for every regulated activity separately. When the enforcement of a tax rule is sufficiently costly, quantity regulation may become desirable even if it eliminates some efficient conduct. The general argument is also useful in the justification of clear rules.

Keywords: *Regulation, environment, enforcement*

JEL classification: H23, D62, K32, L51

4.1 Introduction

Modern societies are characterized by growing regulation. The spectrum of regulated activities as well as the degree of regulation have increased significantly over the past 100 years (for a recent discussion of this phenomenon, see Glaeser and Shleifer (2003)). It is therefore no surprise that questions concerning regulation have received much attention by economists.

The most common way to regulate activities characterized by negative externalities is to impose quantity restrictions (e.g., permits) rather than to use prices (e.g., taxes). Some prominent examples of quantity restrictions in the everyday life are zoning laws restricting commercial construction activity in residential areas, blue laws prohibiting sales of goods during certain times, anti-smoking laws, international trade law (quotas), and anti-trust laws. Hunting and fishing rules also employ quantity limitations. The same

[2]An earlier version of this essay was presented at the 37th European Association for Research in Industrial Economics (EARIE) Conference in Istanbul (September 2010).

holds true for safety and health regulations which generally require compliance with quantity restrictions. The most important case these days seems to be environmental protection which is also dominated by quantity restrictions around the world.

The imposition of quantity restrictions is in sharp contrast to the general predisposition of economists for the price mechanism as the best allocation device. The implementation of a corrective tax is possible in all the situations mentioned above, but the question is, why is that not the case? The literature has offered several arguments. The most obvious explanation is that quantity restrictions are natural in situations where a zero level of the respective activities is intended. This is the case when these activities are regarded as morally unacceptable, such as drug consumption, smoking, or poisoning food. This moral argument, however, has clearly not enough explanatory power for the apparent pervasiveness of quantity regulations.

It is well known to economists that there is an equivalence between a quantity limit and a corrective tax in achieving any target level of activity if the following conditions are satisfied:

1. No uncertainty (i.e., free, complete, and common information).
2. No rent seeking through lobby activities.
3. No enforcement problems and, hence, no enforcement cost.

Any argument for the superiority of quantity regulation must therefore be based on the violation of at least one of these conditions.

Following a seminal paper by Weitzman (1974), the bulk of the literature has focused on the first condition in order to explain the existence of quantity regulation (for a thorough discussion of the topic, see Kaplow and Shavell (2002)). Influenced by Weitzman and a large number of subsequent papers, economists now often argue that in a situation of uncertainty concerning costs and benefits of economic agents and regulators either instrument, tax or quantity restriction, could be superior to the other. Nowadays, this insight can be found even in basic undergraduate textbooks of microeconomics (see,

for example, Pindyck and Rubinfeld (2009)). Weitzman's motivating example, namely the choice of an optimal mode of pollution control, continues to attract the highest attention in the literature.

The second condition in the above list has so far attracted much less attention in the literature. Buchanan and Tullock (1975) have argued that quantity regulations are favored by incumbent firms in order to deter entry of new firms. Finkelshtain and Kislev (1997) have analyzed the choice of a regulation regime subject to lobbying in more general terms. They have shown that—as in the case with uncertainty—the preferred control instrument cannot unambiguously be determined, that is, either instrument can be favored to the other.

The problem of a lack of enforcement has also not attracted much attention in the literature. This is clearly unjustified considering the empirically well-established fact that under-enforcement represents a permanent problem of all kinds of laws. At least since the seminal papers of Becker (1968) and Stigler (1970), economists recognize that enforcement cost and their optimal level pose a genuine economic problem. Building on insight that enforcement incurs cost, Glaeser and Shleifer (2001) have offered a justification for the implementation of a quantity restriction rather than a corrective tax. They have argued that a quantity regulation of a certain harmful activity can reduce enforcement costs by greatly simplifying enforcement since it often requires visual inspection only. This, in turn, encourages private enforcement, that is, enforcement by agents (consumers, competitors, et cetera) other than public enforcement agents or agencies.

This paper also argues that quantity regulation may lower enforcement cost. But, in contrast to most of the literature, we do not restrict the attention to a single activity to be regulated. It is an obvious fact that many regulated activities are often interrelated with, or embedded in, other regulated activities. It is also an empirical regularity that an enforcement authority (e.g., an enforcement agency) is typically responsible for the enforcement of a whole spectrum of regulations. Thus, a new economic argument comes into play: the potential economies of scope in enforcement costs of multiple

activities to be regulated and their trade-off with the well-known negative efficiency effects of quantity restrictions. This argument implies that a case for quantity regulation based on cost may be impossible if we restrict our attention to just one regulated activity.

The literature of law and economics has long recognized the phenomenon of "general enforcement." Enforcement may be general in the sense that several different types of regulations can be enforced by an enforcement agent's activity (see Shavell (1991)). The classical example for general enforcement is a police officer who may detect a variety of infractions when he patrols. We argue that the possibility of general enforcement is particularly convenient in a regime of quantity regulation and, therefore, analyze the implications of economies of scope in enforcement cost for the choice of quantity rather than tax regulation (for a discussion of the concept of economies of scope, see Panzar and Willig (1981)).

A simple example inspired by a fascinating book by Wilson (1989) illustrates the basic idea in a straightforward way. When health and safety inspectors enter factories, they come to enforce a variety of quantity regulations. For example, they look for ladders that are unsafe, slippery floors, guardrails that are missing, fumes that are toxic, et cetera. This example shows that an enforcement officer carrying out a regular inspection is often able to detect violations of different regulations without (significant) additional effort in the case of quantity regulations. In contrast, a tax rule often requires additional investigations for every violation separately—which may be quite substantive. These additional costs in the case of a tax regulation my discourage an enforcement agent from (careful) investigations, with the result that substantial violations may take place unnoticed. If there is a quantity restriction, the detection of violations may often be a by-product of other enforcement efforts. This aspect of enforcement is of particular importance since the implementation of effective incentive structures for an enforcement agent is generally quite difficult. The enforcement process is not easy to monitor and that rewards which agents receive for enforcing regulations are typically intangible. Additionally, enforcement capacity is generally very

scarce so that every possibility for productivity improvements must be used.

4.2 A Simple Model

This section presents a simple model based on Shavell (1991). Consider a situation in which a benevolent government wishes to regulate various activities creating negative externalities (e.g., pollution). This task is assigned to an enforcement agency. The question we address is whether the government wants this agency to enforce a corrective tax or a quantity restriction. For simplicity, assume that the there are only two regulated activities. The extension to more than two activities is straightforward.

We consider an enforcement agency examining possible violations of regulation rules (taxes or quantity restrictions) by risk-neutral individuals with a total population size equal to 1. There are two types of acts: the fraction $0 < \theta < 1$ are of type 1; $0 < 1 - \theta < 1$ are of a type 2. Activity 1 leads to a low level of harmful externality of $h_1 \geq 0$, and activity 2 results in a high level of harmful externality of $h_2 \geq 0$; with $h_2 > h_1 \geq 0$.[3] The allowed levels for the both activities are denoted by $\overline{h}_1 \geq 0$ and $\overline{h}_2 \geq 0$ in the case of a system of quantity restrictions.

Individuals' benefits $b_i \geq 0$, $i \in \{1, 2\}$, from both activities have a continuous density function $f(b_i) > 0$ on $[0, \infty)$. We assume that the benefits b_1 and b_2 of individuals are independently distributed.[4]It is clear that the that first-best behavior is for an individual to commit act i if and only if $b_i \geq h_i$. We assume that the regulation authority knows only the distribution of individuals benefits.

If an individual engages in a harmful activity, he will suffer a sanction with a certain probability. For simplicity, assume that the sanction is solely monetary. Let $s_i(h_i)$ be the monetary sanction for engaging in activity

[3]Considering the fact that each individual could choose whether to commit either act 1 or act 2 or both would significantly increase the complexity of the model. No substantial additional insights, however, could be obtained from a generalization of the model along these lines.

[4]This assumption is inconsequential. Since it significantly simplifies notation, it is made here for expositional convenience.

h_i, $i \in \{1,2\}$, and $p(e_i)$ the probability of detection. Our crucial assumption is that only quantity regulation of both activities allows for general enforcement, that is, for cost reductions induced by economies of scope, meaning that the enforcement of both activities in the case of quantity regulations requires enforcement expenditures of only $e > 0$, resulting in a probability of apprehension of $0 < p(e) < 1$, where $p(0) = 0$, $p^{'}(e) > 0$, and $p_i^{''}(e) < 0$. If taxes are used to regulate both activities, general enforcement is not possible, that is, enforcement of both activities costs $e_1 + e_2 > 0$, resulting in a probability of apprehension of $0 < p_i(e_i) < 1$, $i \in \{1,2\}$, where $p_i(0) = 0$, $p_i^{'}(e_i) > 0$, and $p_i^{''}(e_i) < 0$. It is assumed that individuals know the probability of detection and applicable monetary sanctions for both activities. Because individuals are risk neutral, an individual thus will commit a harmful activity i if and only if the benefit at least exceeds the expected monetary sanction.

Social welfare equals the benefits individuals derive from both activities less the harm done less enforcement costs. In calculating social welfare, we ignore potential tax revenues, collected fines, and potential rewards to enforcers, assuming all of these as pure transfers. Accordingly, social welfare in the case of a tax regulation (W_T) is equal to

$$W_T = \theta \int_{p_1(e_1)s_1}^{\infty} (b_1 - h_1) f(b_1)\, db_1 + (1-\theta) \int_{p_2(e_2)s_2}^{\infty} (b_2 - h_2) f(b_2)\, db_2 - (e_1 + e_2). \quad (15)$$

Society's problem is to choose e_1, e_2, s_1, and s_2 to maximize (15).

When the regulatory authority imposes a system of quantity regulations,

social welfare (W_Q) is equal to

$$W_Q = \theta \int_0^{\overline{h}_1} (b_1 - h_1)\, f(b_1) + \theta \int_{\overline{h}_1 + p_1(e)s_1}^{\infty} (b_1 - h_1)\, f(b_1)\, db_1 + (1-\theta) \int_0^{\overline{h}_2} (b_2 - h_2)\, f(b_2)\, db_2 + (1-\theta) \int_{\overline{h}_2 + p_2(e)s_2}^{\infty} (b_2 - h_2)\, f(b_2)\, db_2 - e. \tag{16}$$

Society's problem in this case is to choose s_1, s_2, $\overline{h}_1$, $\overline{h}_2$, and e that maximize (16).

We can now state

Proposition 5. *Social welfare may be higher in a regime of quantity regulations.*

Proof. The difference in social welfare between the two regimes is $W_T - W_Q \gtreqless 0$–even for the optimal values of the endogenous variables. □

As proposition 5 indicates, the comparison of the two regimes is rather ambiguous: general statements on the relative desirability of one of the two regimes require strong assumptions concerning the involved functions. However, some predictions can me made. As one would expect, the relative desirability of a regime of quantity regulations is greater, the lower the optimal level for e is. In addition, the greater the discrepancy between the two activities, that is, the greater the difference between the optimal values for h_1 and h_2, and the lower the level of θ is, the more desirable a regime of tax regulations becomes because it offers the possibility of differentiation in enforcement effort. Further, if most individuals have very high benefits, and thus engage in harful activities, a regime of quantity restrictions will be more efficient because of lower enforcement costs.

The proposition 5 implicitly assumes that a particular enforcement agent has a real incentive to enforce regulation, that is, we ignore principal-agent problems. This is a critical assumption since enforcement agent generally do not

take social benefits of enforcement (fully) into account. Further, the government and the management of the enforcement agency typically do not have full control over the incentives facing law enforcement agents. For example, many of the rewards that enforcement agents receive for implementing regulations are often intangible, including self-esteem and the respect of one's peers. In other words, the government may only influence the enforcement agents only through the choice of the mode of regulation–at least in the short run.

Let us assume that a representative enforcer derives an exogenously given payoff of $r_i > 0$ from enforcing the law concerning activity $i \in \{1, 2\}$, which here implies punishing a violator. Since quantity regulations make general enforcement feasible, it is evident that the private returns to enforcement for the enforcer are higher under that regime of regulation; that is, the payoff from enforcement is higher without increasing the costs. Thus, we can state the following

Proposition 6. *Enforcement agents prefer a system of quantity regulations.*

Proof. The proof is straightforward. □

This analysis captures in a simple way the difference between a regime of quantity regulations with general enforcement and a tax regultions associated with specific enforcement. The regulation by taxes may yield a first-best solution, but may be to costly to enforce. In contrast, quantity regulation may eliminate some efficient conduct, but may be much cheaper to implement because it allows for general enforcement.

4.3 Summary and Extensions

This paper offers a simple explanation for the prevalence of quantity regulations in a variety of areas. The basic point is that general enforcement is much easier in the case of quantity regulation, and may therefore lower enforcement costs. When the enforcement of a tax rule—which is usually

based on specific enforcement efforts—is sufficiently costly, quantity regulation may become desirable even if it eliminates some efficient conduct.

The basic model allows for extensions in a variety of ways. For example, the case for quantity regulation becomes even stronger when we introduce the possibility of some amount of private enforcement. This is of great importance since private enforcers often play a crucial role in many areas. For instance, the enforcement of anti-trust laws relies to a great extent on information supplied by direct competitors. The basic argument of the paper is also valid in a world with private enforcement. The argument is the same as in the case of public enforcement, that is, quantity restrictions make the enforcement of a wide range of regulations by private agents generally much easier.

It is even possible to generalize the basic point of this paper. The discussion has focused on the topic of regulation, but it can be generalized to the following question that appears in any system of governance and is characterized by tremendous importance: How clear should rules be? First, they can be formulated that allows for many interpretations in order to leave room for subjective judgment adapted to particular circumstances in each case. Second, they can be stated in way that leaves little or no room for varying interpretations. It is clear that either way requires enforcement.

The main area where this issue comes into play is the legal sector. Legal systems, particularly those of civil-law countries, describe legality of conduct in terms of relatively clear rules, incorporated into legal codes. The existence of clear rules (referred to as bright-line rules) in contrast to broader principles can also be understood from the perspective of general enforcement activities. The simpler the rules for an array of activities, the easier general enforcement of these rules is. That may lead to considerable reductions in enforcement costs or more enforcement with a given budget.

5 A Signaling Approach to Soft Budgets[5]

Abstract

This paper offers a new explanation for soft budget constraints (refinancing of loss-making enterprises), a phenomenon often observed in various economic systems. Investment projects of an economic agent (e.g., a firm) are often the more profitable, the more confident the agent is that the principal (e.g., a government) is competent or supports the project or both. At the outset of an investment project, agents will often be unable to fathom the true motivations and/or competence of the principal influencing the success of the project. The confusion becomes even worse in times of an economic crisis or ongoing (radical) reforms. Soft budget constraints may then be used to signal support and/or competence by the principal.

Keywords: *Soft budget constraint, hard budget constraint, signaling*

JEL classification: D02, E42, G28,

5.1 Introduction

The term "soft budget constraint" (SBC) is now familiar to most economists. It was originally coined by Kornai (1979) to denote the various forms of refinancing of loss-making enterprises in socialist economies. He has focused his attention mainly on the consequences of SBC, namely the emergence of pervasive shortages under socialism. Kornai saw the wish of "paternalistic" governments to avoid social and/or political unrest as the primarily reason for SBC. The phenomenon of the SBC, however, has been present in most countries in some form, though clearly being more prevalent in socialist economies.

The relevance of the phenomenon became strikingly evident in the global financial crisis starting in 2007. Governments around the world have been forced to refinance loss-making banks and firms from various other sectors

[5]Published in *Economics Letters, 111(3), June 2011, p. 272-274*

in order to save the whole economy from collapsing and/or to prevent unemployment rate to rise significantly. Even entire countries need a bailout. Kornai et al. (2003) offer a recent and thorough presentation of the concept and the related theoretical and empirical literature.

We follow their definition by stating that the SBC phenomenon is present if one or more supporting organizations (S-organizations), typically the (central) government, cover at least a part of the deficit of an organization which has a "budget constraint" (BC-organization). The term "budget constraint" means that the organization is supposed to cover all its expenditures out of its initial endowment and/or revenue.

The interest of a BC-organization in a bailout does not need further discussion. In contrast, the motives of the S-organization are generally diverse and less clear. Some common motives for bailouts are business interests, political considerations, the influence of interest groups, and (macro)economic concerns. The theoretical literature on the SBC phenomenon has emphasized the following commitment problem. At the beginning, the prospects of the BC-organization look good. To additionally boost incentives, the S-organization may even promise not to help in the case of financial difficulties. But a declaration is *per se* not credible ("cheap talk"); the S-organization may ex post be tempted to help the BC-organization. Common experience and basic economic reasoning indicate that lack of credibility often creates undesirable consequences. The anticipation of a rescue may diminish efforts on the side of the BC-organization. In other words, the ex ante incentives and the ex post incentives of the S-organization may be conflicting (time-inconsistency problem). For example, a SBC may lead loss-making enterprises to circumvent a necessary restructuring activities because the threat of bankruptcy is reduced or even completely eliminated. Further, the macroeconomic stability of a whole country may be in danger because (continued) SBCs make it hard to keep government expenditures under control.

An obvious question is why the S-organization would refinance a financially troubled BC-organization despite potential drawbacks? A possible answer is that the S-organization cannot credibly commit not to refinance bad projects

ex post and cannot distinguish bad from good ex ante. If a project is launched, it will be often refinanced. The reason is that termination is often not sequentially rational, that is, once a project has been started, its net continuation value may be positive so that, in the absence of a credible commitment device, SBC arises. The BC-organization knows this, and submits bad projects for financing in the first place. This is the basic argument in Dewatripont and Maskin (1995) which has become the standard modeling approach of the SBC. They have been primarily concerned with ways of hardening budget constraints of enterprises. This topic has attracted the most attention in the by now large literature on transition (for example, see the book by Roland (2000)). For instance, in Eastern Europe and the former Soviet Union, mass privatization is the method chosen to harden budget constraints.

Modeling the SBC in such a way is surely justified and valuable. But it reduces the phenomenon to a dynamic commitment problem, that is, to a time-inconsistency problem. This paper offers a new approach to the SBC phenomenon. We give new answers to the following two questions by using a signaling argument:

1. Why are there SBC?

2. Why are some economies much more affected by the SBC phenomenon?

The success of various projects often depends on the support or competence (in political and/or economic matters) or both of the government, which is often not completely known. This is especially the case for large and long-term investments. For example, large investments in the exporting sector are unreasonable if the government does not support trade and financial liberalization. The government thus has to signal its support and/or competence in some way. There are a variety of possibilities and examples for such actions, as we will see in more depth in the next section. For example, the German Federal Government offers export guarantees (known as Hermes-Bürgschaften) for the very important exporting sector of the country. These guarantees imply an expenditure by the public sector if the foreign customer

does not meet its obligations.

The success of a project may also depend on the support or competence or both of the government if it has significant political and/or social implications. For example, projects with (potentially) considerable effects on the political stability of a country may also require a SBC in order to get implemented.

The confusion about the true objectives and the true competence of the government is a recurring and widespread phenomenon. Incomplete information of this sort is likely to be particularly prevalent in countries where governments (as the S-organization) are of great importance, often act in in-transparent ways, or change quite frequently or all at the same time, and in less developed countries in particular. Launching of a new or (the) continuation of an existing enterprise is therefore often not very promising without some form of assurance by the government.

The confusion created by incomplete information often becomes even worse when the government's assessment of the current economic situation or policy objectives or both are (at least temporarily) unclear as a consequence of an acute (economic) crisis or an ongoing (fundamental) reform. The resulting reduction in economic activity may then exert (additional) strong recessionary effects. In order to at least stabilize the situation, the government may be forced to use signaling, even to the point of wasting resources. The implementation of a SBC may be used as a signaling device. Signaling in this fashion is of course potentially very costly for a government, and its ultimate desirability will depend on a number of idiosyncratic factors. However, the main point is that a SBC may not be negative *per se* from an ex ante perspective. This approach may explain the existence of the SBC observable in different countries. Notice that this interpretation of the SBC is radically different from the time-inconsistency specification above wherein the BC-organization understands the behavior of the S-organization only too well.

Though in the following we shall speak of a government and a firm, the reasoning can be applied far more broadly. For example, the central government

may want the local government to invest in a certain project, and therefore implements a SBC. Another example is a (large) firm wanting a subsidiary to launch a new or to continue with an existing project.

5.2 The Model

In order to formalize the idea of SBC as a signaling device, we consider a simple signaling game with three players that allows a straightforward analysis of the issue. To focus the analysis, we will abstract from many real-world aspects.

The first two players refer to the government of a different type (as the S-organization and the sender of the game), and the third player is a manager of a (state-owned) firm (as the BC-organization and the receiver of the signal) with an investment project. The project of the firm with the support of and/or with a competent government in charge is more likely to succeed or at least continue. The manager's incomplete information on the government's preferences or competence or both may generate considerable uncertainty. This may lead to a situation where profitable projects do not get realized. A competent and/or a supportive government may therefore benefit from signaling its type. In this model, we do not distinguish between the manager and the shareholder of the firm, and assume that the manager serves the interests of the shareholder.

The model features two periods. The manager can invest an amount $I > 0$ in the first period in order to receive a discounted amount $V > 0$, with $V > (1 + i) I$, or 0 in the second period; that is, the project could lead to a total loss for the investing firm. The parameter $i \geq 0$ denotes the opportunity cost of capital. A successful project has a discounted value of $W > 0$ to the government. We assume that the value of W is exogenous. The value of W allows for many interpretations. For example,

$$W = f(I, V), \tag{17}$$

where we assume that W is determined as the political-market value of I

and V. No substantial additional insights, however, could be obtained from a generalization of the model along these lines.

Assume that if the project fails, the government can refinance the firm with a discounted amount $s \geq 0$. The government can be of two types: type–A with prior probability $0 < a < 1$, or type–B with prior probability $0 < 1 - a < 1$, whereby the probability a is common knowledge. A type–A government implies a higher probability $0 < p_A < 1$ that the project will succeed over a type–B government with $0 < p_B < 1$; that is, $p_A > p_B$. Informally, this assumption implies that the manager may be interested in the actual type of government, and that a SBC is more costly for a type–B government. Technically, this assumption implies that the single-crossing condition is satisfied. A strategy for the government prescribes a probability distribution over actions s for each type $j \in \{A, B\}$. In contrast, a strategy for the manager prescribes a probability distribution over I for each signal s.

The manager tries to maximize its expected profits, and the government tries to maximize its expected benefits, defined in this case as the expected return from the perspective of the government minus potential refinancing costs incurred by the government. Note that we do not assume that the government is necessarily benevolent.

Altogether, there are three players in the signaling game: a government of type–A, a government of type–B , and an investing firm. The game they play is characterized by the following sequence of events:

1. "Nature" determines the type of government (either A or B), the government learns its type, and nobody else does.

2. The government announces or implements $s \geq 0$, which then becomes common knowledge.

3. The manager observes the value of s, and decides whether to invest $I > 0$, contingent upon the value of s.

4. "Nature" determines the profitability of the project, which then becomes common knowledge.

5. The government potentially refinances a failed project.

6. The payoffs are realized.

The announcement or implementation of s allows for different interpretations. Here are some possibilities:

1. The government can use some form of earmarking.

2. The government could implement of some form of "soft" taxation–mainly through the tolerance of tax arrears.

3. The government could issue a guarantee.

4. The government could establishment (special) lines of credit.

5. The government could signal its support for the project of a given firm by bailing out a similar project and/or a similar firm with an amount s.

All these options to soften the budget constraint are observable events and all well-known from everyday practice of governments around the world. It is of crucial importance to realize that all the options in the above list are *not* costless, non-binding, non-verifiable claims; that is, they are *not* cheap talk due to the fact that they are in fact all directly or indirectly costly, verifiable, and binding–typically though explicit contracts enforceable by third parties. For example, if a government issues a guarantee concerning a specific firm, as in the third point in the above list, the firm could enforce it through the judicial system. The same is true even for the fifth point in the above list. A government generally cannot deny help to one firm if it simultaneously helps a similar firm without the threat of a legal dispute concerning this discrimination.

The situation is often somewhat different in transition and developing economies due to the lack of a reliable and independent court system that is able to

enforce contracts. In such situations, governments establish a system characterized by the fact that firms can often get away with postponing payments beyond the agreed-upon deadline–especially in the case of tax payments ("soft" taxation–the second point in the above list of options). There are many examples of such arrangements in a variety of countries. It is at least very costly–or even impossible–for the government to eliminate soft taxation in the short run due to the lack of institutional capacity and the induced political resistance to such measures.

In is important to emphasize that the actual forms of the SBC, however, often differ significantly from country to country and period to period, as the extensive empirical literature on various methods to soften the budget constraint clearly shows. General predictions concerning the implemented form of the SBC are thus hard to make. Kornai et al. (2003) offer a concise and recent survey on the empirical literature concerning the SBC.

The structure of the game is typical for signaling games. The solution concept is that of a perfect Bayesian equilibrium (PBE). For a detailed discussion of this concept, see the book by Fudenberg and Tirole (1991).

To highlight the effect of the manager's incomplete information about the type of the government on the investment decision, we first analyze the choice under complete information as a reference point. That is, we temporarily assume that the government's type is common knowledge rather than privately known by the government. The problem of the firm in a world of complete information is very simple; the firm invests—regardless of the type of government—if the following condition holds:

$$p_j V > (1+i)\, I,\ j \in \{A, B\}\,. \tag{18}$$

The firm invests under a type–A government only if the following conditions hold:

$$p_A V > (1+i)\, I, \tag{19}$$

$$p_B V < (1+i)\, I. \tag{20}$$

We will now return to the case of a game with incomplete information. The solution of the game depends upon the beliefs of the manager concerning the relationship between the observable signal s and the unobservable type of the government. Such beliefs are described by the following probability function

$$\mu(s) := Pr\{government\ is\ of\ type - A|s\}, \tag{21}$$

which gives the probability that the firm assigns to the government's having type–A, conditional on observing the choice of s.

The solution of the game is a collection of strategies and beliefs that have the following characteristics:

1. The action of each player in each information set is optimal, given his belief and the strategies of all players in the future.

2. The beliefs are obtained from equilibrium strategies and observed actions using Bayes' rule.

Signaling games generally allow for an infinite number of equilibria. We will restrict our attention to just a subset of all pure-strategy equilibria of which there are two kinds: separating and pooling solutions. In a separating equilibrium, different types choose different signals. In contrast in a pooling equilibrium, both types choose the same signal, that is, in a pooling equilibrium the manager does not update its beliefs when observing a signal. Both types may occur in this game.

First, we turn our attention to the possibility of a pooling PBE. Such an equilibrium is a situation in which both types of government send the same signal, that is, choose the same value for s and are therefore indistinguishable (the signal is not informative at all). The immediate conclusion is the following: If the government sends the same signal independent of its type, the signal must be $s = 0$. A situation in which both types of government send the signal $s > 0$ cannot be optimal because of its unsustainability, that is, because of the incentives to deviate and send the signal $s = 0$. Given this, we shall only study pooling equilibria with $s = 0$.

The agent invests—regardless of the government's type—if the following condition holds:

$$[ap_A + (1-a)\,p_B]\,V > (1+i)\,I. \tag{22}$$

The following proposition describes one simple pooling equilibrium without a SBC on the equilibrium path as a point of reference.

Proposition 7. *Suppose that condition (22) is satisfied, then the following beliefs and strategies constitute a pooling perfect Bayesian equilibrium:*

1. The manager has beliefs given by $\mu(s) = \begin{cases} a & if \;\; s = 0 \\ \in [0,1] & if \;\; s > 0 \end{cases}$.
2. The government chooses $s^* = 0$, regardless of its type. In other words, this means that there is no SBC.
3. The manager starts the project, that is, the firm invests I.

Proof. The proof is straightforward. □

We now turn our attention to the more interesting possibility of a separating PBE with a SBC on the equilibrium path. In the case of a separating equilibrium the firm does not invest if it thinks the government is of type–B, and the firm does invest if it thinks the government is of type–A (the signal is thus informative). The idea that a signal can be informative as to a government's type depends on the fact that the firm will form different beliefs with regard to the government according to which signal is sent.

Assume that condition (22) is violated. Further, assume the conditions (19) and (20) are satisfied. Note that inequalities (19) and (20) are a subset of all cases where (22) is violated. The following proposition describes a separating equilibrium.

Proposition 8. *Suppose that condition (22) is not satisfied, and the conditions (19) and (20) hold, then the following beliefs and strategies constitute a separating perfect Bayesian equilibrium.*

1. The firm has beliefs given by $\mu(s) = \begin{cases} 1 & if \ \ s \geq s^* \\ 0 & if \ \ s < s^* \end{cases}$.

2. The government chooses $s^* \in [\frac{p_B W}{1-p_B}, \frac{p_A W}{1-p_A}]$ in the case of a type–A government, and the government chooses $s = 0$ in the case of a type–B government. In other words, this means that there is a SBC.

3. The manager starts the project if $s \geq s^*$.

Proof. In this case the manager does not invest if he thinks the government is of type–B, and the manager does invest if he thinks the government is of type–A. The condition for a type–A government to prefer committing s over not committing is

$$p_A W - (1 - p_A)\, s \geq 0. \tag{23}$$

Therefore, the maximum commitment the type–A government would make is $\frac{p_A W}{1-p_A}$.
A type–B government prefers no commitment (which gives zero net benefits) over committing to spend at least s if

$$p_B W - (1 - p_B)\, s \leq 0. \tag{24}$$

The last expression is satisfied as an equality if $s = \frac{p_B W}{1-p_B}$. In other words, if a type–A government is expected to commit to spend $\frac{p_B W}{1-p_B}$ or more, then a type–B government would not commit to spending, and therefore the equilibrium separates the two types. □

This proposition shows that a SBC may be in the best interest of a government in order to increase social welfare. Note that the model is also compatible with an interpretation that a SBC may be in the best interest of a government even if it is not beneficial from a social point of view.

Since the threshold level s^* can assume any value between $\frac{p_B W}{1-p_B}$ and $\frac{p_A W}{1-p_A}$, there is an infinite number of such separating equilibria. However, since the government has no benefit from raising its bailout cost in this model,

these equilibria can be ranked by the Pareto criterion. The unique Pareto-dominant separating PBE (also referred to as the *least-cost separating* equilibrium) is obtained by setting $s^* = \frac{p_B W}{1-p_B}$.

Note that there may be a separating equilibrium in which a type–A government may commit to spend more than the project's cost in the case of a failure, or $\frac{p_B W}{1-p_B} > I$.

What happens when there are no binding contracts? It is clear that in the absence of binding contracts, there cannot be a separating equilibrium, and there exists a pooling equilibrium. Were there a separating equilibrium, one type of government would benefit by mimicking the other type, destroying the equilibrium. The existence of a pooling equilibrium is evident. For example, a case without a SBC represents a pooling equilibrium under these circumstances.

5.3 Summary and Extensions

The purpose of this paper is to offer a new and yet simple explanation for the (continued) existence of SBCs. The argument offered here is that a SBC may have the consequence of distinguishing a genuinely supportive and/or competent government from its counterpart. That, in turn, has the effect of implementing (more) and/or continuing existing projects than it would otherwise have been, alleviating many problems introduced by credibility concerns. Though resources may sometimes be wasted, a SBC can indeed increase expected social welfare.

The approach in this paper can explain the pervasiveness of soft budgets in socialist, developing, and transitional economies, since the success of investment projects in such economies depends to a great extent on the support or competence of the government or both, which is often not (completely) known. That is, the managers of firms will often be unable to fathom the true type of the government influencing the success of the project through its actions to a significant degree. The confusion concerning the type of government becomes particularly relevant in times of an economic crisis or ongoing fundamental reforms. SBCs may then be used to signal support or

competence or both.

We conclude by discussing some of the various possible extensions of the basic model. One possible extension would allow for more than one firm. These firms—which might have multidimensional strategy sets—could be related in various ways. Another interesting extension would a discussion of additional consequences of a SBC. For example, a SBC could possibly influence the costs of the actual project, and it could influence the characteristics and prospects of future projects.

The incorporation of various asymmetries and/or political economy arguments is also quite plausible. Politicians generally differ in their valuation of welfare of different groups and/or organizations, and in their regional/ethnic/industry attachment. For example, the model can also be extended, and the effects strengthened, in a way that differentiates the two types of government concerning their valuations of the firm's investment project. For example, a type–A government can place a value of W_A and a type–B places a value of W_B, with $W_A > W_B$.

6 (Mis)Counseling in Monopolistic Competition: A Case for Regulation[6]

Abstract

We analyze the impact of (possibly) fraudulent independent experts in a market for credence goods characterized by monopolistic competition. This setting applies to various industries such as repair markets, health care markets or financial services markets where consumers are usually uninformed about which product best fits their individual needs and therefore use an expert. We analyze the outcome under honest and fraudulent, whereby honesty may require side payments from the firms to the experts. Rigorous regulation of the relationship between firms and experts may be essential in order to make these markets operate more efficiently.

Keywords: *Imperfect markets, miscounseling, monopolistic competition, collusion.*

JEL classification: D43, L13, L15

6.1 Introduction

Miscounseling is a phenomenon of great empirical interest. Although there are a variety of examples, we may focus our attention on rather spectacular cases. In 2004, for instance, New York Attorney General Eliot Spitzer sued Marsh & McLennan, the largest insurance broker in the world, for collusion with major insurance companies. In the civil complaint filed in State Superior Court in Manhattan, Spitzer alleges that Marsh "steered unsuspecting clients to insurers with whom it had lucrative payoff agreements, and that the firm solicited rigged bids for insurance contracts." For a long time in the United States, insurance brokers have received high payments from insurance

[6]Co-authored with Johann Buhne and Annette Hofmann. This essay was presented at the 36th annual conference of the European Association for Research in Industrial Economics (EARIE) in Ljubljana (September 2009).

companies they ostensibly evaluate objectively on behalf of customers who pay fees in good faith in order to have their coverage arranged. The brokers insist that the payments from the insurers—called contingent commissions or placement-service agreements—do not influence their advice, and that they regard the money as an ex-post reward for a job well done. Since the late 1990s, brokers have been disclosing these payments in general terms.[7]

Economic models usually assume that consumers know which goods or services they need. Many goods, however, have characteristics which cannot be revealed by inspection or ordinary use. These goods are known as credence goods. In such a case consumers often have to rely on an expert's advice to assess their needs. This creates the possibility of miscounseling. There are many examples of situations were consumers face the problem of receiving goods or services which are not necessarily suitable to them. Common examples include all sorts of repair markets, health care markets as well as legal and financial services markets. In view of the actual financial crisis, a very important example in these days are mortgage brokers. Many reviewers of the ongoing financial crisis assign them a great deal of responsibility.

Obviously, there is a need for means—besides legal restraints—to fight miscounseling. One possibility is to establish reputation. This approach, however, requires repeated interaction. Another possibility is a kind of consumer protection institution, which again produces a bunch of new incentive problems. The most straightforward way to ensure honest counseling is to separate counseling and purchase. But is that enough? Introspection and the story above clearly tell no.

The main purpose of this paper is to analyze the consequences of fraudulent independent experts in a model based on the familiar paradigm of monopolistic competition. We use the well-known circular product space introduced by Salop (1979). In the original Salop model, all consumers know their position in the product space. We examine the effects of relaxing this crucial assumption and incorporate uninformed consumers and fraudulent independent experts into this framework. The contribution of the paper is

[7] See Spitzer (2004)

threefold. First, we study the question of how much miscounseling should be expected in equilibrium. Second, we analyze optimal pricing and the number of offered varieties under miscounseling. Third, we analyze the impact of side payments—payments to experts in order to prevent them from business stealing. We find that miscounseling will not happen in equilibrium with side payments. The side payments, however, lead to fewer firms and higher prices in market equilibrium compared to the standard full-information Salop equilibrium.

The paper is organized as follows. Section 6.2 contains a literature review. The basic assumptions of the analysis are presented in section 6.3. Section 6.4 derives the conduct of the firms. Side payments are considered in section 6.5. Section 6.6 studies the brokers' behavior. Concluding remarks follow.

6.2 Literature Review

There are basically two strands of literature we build upon. On the one side, our model is loosely related to the literature on credence goods. This literature is surveyed by Dulleck and Kerschbamer (2006). Contrary to this literature, however, we accentuate the consequences of bribery in a model of monopolistic competition which—to the best of our knowledge—has not previously been used to answer the questions we pose in this paper. The monopolistic-competition assumption is convenient since product differentiation of credence goods is a common phenomenon. This approach offers the possibility to study the consequences of miscounseling on the number of offered brands in equilibrium. A better understanding of incentive problems and miscounseling may clearly be very helpful in designing better public and regulation policy. On the other side, there is a large literature related to financial intermediation. There are two main strands in this literature that are related to our model. The first is concerned with capital markets in general while the second is related particularly to insurance markets. In the first strand of literature, various authors analyze the benefits of delegating some informational task to an intermediary who is presumed to have some transaction cost advantage over consumers.

There is significant literature (that cannot be reviewed here) dealing with informational asymmetries in banks and financial services markets, particularly commercial bank markets. Those markets are studied, for instance, by Chan (1983), Leland and Pyle (1977), Campbell and Kracaw (1982), Diamond (1984) and Williamson (1987) as well as Allen and Santomero (1997). Most authors stress the role of transaction costs. For example, fixed costs of asset evaluation imply that intermediaries have an advantage over individuals because they allow such costs to be shared. In a similar way, trading costs imply that intermediaries can more easily be diversified than individuals. In the second strand of literature on financial intermediation, authors explicitly address insurance markets. Brokers as search agents whose function is to match trading partners in insurance markets are studied by Rubinstein and Wolinsky (1987). The authors analyze time-consuming negotiations between buyers, sellers and middlemen. Posey and Yavas (1995) as well as Posey and Tennyson (1998) study search cost models while Cosimano (1996) looks at a monopolist intermediary who lowers the probability of an unsuccessful trade. Seog (1999) offers a dynamic model and focuses on price search. Intermediation is found to solve adverse selection problems in insurance markets as illustrated by Biglaiser (1993) and Cummins and Doherty (2006). Among all these studies, our paper is most related to Gravelle (1994). Gravelle is, however, not concerned with matching. In his setting, products are homogeneous and consumers differ in their willingness to pay for insurance coverage which is determined by some unknown random variable. This variable is only observed by brokers. Thus the brokers' main service is to advise consumers on whether they should purchase insurance or not. In contrast to Gravelle, our focus here is on corrupt independent experts. We allow for product differentiation and product-specific mismatching which allows for different degrees of consumer miscounseling.

The most recent study we build upon is by Inderst and Ottaviani (2009). These authors analyze the optimal compensation structure for a direct marketing agent who must be incentivized to sell and—simultaneously—not to miscounsel. Inderst and Ottaviani (2009), however, are concerned with the internal agency problem and do not address market structure as a result

of competition among different firms and independent brokers, which is the main focus of this paper.

6.3 The Assumptions

We adopt a spatial model of product differentiation based on Salop (1979). Consumers are uniformly distributed on a circle of unit circumference: $(0, 1]$ is the space of differentiated preferences. A single good is produced at constant and uniform marginal cost $c > 0$ by firms $i = 1, 2, ..., n$, each of which decides whether to enter the market and then selects a symmetric location in the unit circle. All firms have uniform fixed cost $f > 0$. Consumers have linear unit transport cost proportional to the distance to firms, and buy one unit of the good from the firm for which price plus travel cost is lowest. Transport cost is due to the "mismatch" between consumers' individual preferences and offered product varieties. The model is a three-stage game. In the first stage, all firms (simultaneously) decide whether to enter the market; firms entering the market are equidistantly located on the circle. In the second stage, brokers decide upon their fees. In the third stage, having observed the locations selected and the brokers' fees, each firm simultaneously offers a price and decides about how to handle the business with the brokers, i.e., business stealing or side payments strategy (which we specify below). The solution concept is subgame-perfect equilibrium.

We assume that consumers do not know their position on the circle ex ante. Consider the individual decision problem of an uninformed consumer who decides upon becoming an informed consumer: In case he becomes informed, he can choose between the nearest firms located in the product space left and right to his position on the circle. Without loss of generality, we denote these two firms by i and $i + 1$, respectively. An informed consumer located at $x \in \left(0, \frac{1}{n}\right)$ faces mismatch x from purchasing a product from firm i (and respectively $\left(\frac{1}{n} - x\right)$ when he purchases at $i + 1$), so his mismatch ranges from 0 to $\frac{1}{2n}$ implying an expected mismatch of $\frac{1}{4n}$. An uninformed consumer's mismatch is between 0 and $\frac{1}{2}$, which implies an expected mismatch of $\frac{1}{4}$. Since consumers do not know their position on the Salop circle, they

base their decision to become informed about the adequate product variety on the advantage of being informed versus the cost of becoming informed. A consumer $l \in [0,1)$ may engage in costly search for her best-fitting variety. Suppose consumers' search cost $\theta_l \in [0,\infty)$[8] is governed by some distribution function $F(\theta_l)$. Given consumers are rational, searching for an optimal product variety takes place if the expected reduction in transport cost is greater than individual search cost, hence if

$$\theta_l + \frac{1}{4n} \leq \frac{1}{4}. \tag{25}$$

As a consequence, there is an indifferent consumer with search cost $\overline{\theta}$, where

$$\overline{\theta} := \frac{n-1}{4n}. \tag{26}$$

such that a share of $F\left(\overline{\theta}\right)$ consumers becomes informed while $1 - F\left(\overline{\theta}\right)$ prefer to stay uninformed. Uninformed consumers are rational and follow the "principle of insufficient reason" and purchase at the cheapest provider (since all product varieties are equally likely to be associated with each price).[9]Therefore, firms engage in local competition for informed consumers and in Bertrand competition for uninformed consumers. Given this type of competitive environment, as has been shown by Polo (1991) using Hotelling's street, no pure-strategy equilibrium exists but only a mixed-strategy equilibrium where prices are lower than in the full-information pure-strategy case. In the face of costly product differentiation, it seems forceful that providers seek to inform every consumer on the "advantage" of their own variety. Since advertising alone is hardly efficient in matters of complex products, engaging intermediaries—who act as information brokers—is common practice.

Suppose there is a market of $m > 1$ identical information brokers[10] ad-

[8]The search cost may also include expected personal cost resulting from miscounseling.

[9]The "principle of insufficient reason" was first expressed by Jacob Bernoulli. It states that if an agent is ignorant of the ways an event might occur (and therefore has no reason to believe that one way will occur rather than another), the event will occur equally likely.

[10]The number of brokers is assumed exogenous since we are mainly concerned with the consequences of miscounseling on the firms. A restriction concerning the number of brokers could be due to government regulations. For example, the requirement for licenses

vising a consumer about his individual position in the circle and hence of his best-fitting product variety against payment of some fee $k_j > 0$, where $j = 1, 2, ..., m$.[11] Given that a representative broker j operates along the whole market $[0, 1)$, she engages in pairwise Bertrand competition with $m - 1$ brokers at any location $x \in [0, 1)$. If consumers can observe all k_j, the information market becomes competitive, so $k_j = k$ for all j, where k is marginal cost of advising a consumer. Now, the decision problem of an uninformed consumer upon becoming informed is threefold. Consumers with $\theta < \min\left\{k, \overline{\theta}\right\}$ engage in individual search, while consumers with $\theta \geq \min\left\{k, \overline{\theta}\right\}$ consult an information intermediary or stay uninformed depending on the fee k. If the brokers' fee is above $\overline{\theta}$, those consumers prefer to stay uninformed since the cost of information is above its expected benefit. Then there is no demand for the brokers' service. If, however, $k \leq \overline{\theta}$ holds, all consumers become informed about product varieties. Making the reasonable assumption that this inequality indeed holds and hence the brokers' fee is below the expected benefit of information, there exists a broker market in which providers engage in purely spatial competition. Since brokers are identical, we further assume that consumers' demand for information distributes symmetrically among all m intermediaries for any location $x \in [0, 1)$.

Given that all consumers undertake an informed purchase decision, we may derive firm i's demand. Suppose that $\overline{x}_i$ is the location of a consumer who is indifferent between varieties offered at i and $i + 1$, that is

$$\overline{x}_i := \frac{p_{i+1} - p_i + \frac{1}{n}}{2} \quad i = 1, 2, \ldots, n. \tag{27}$$

As a point of reference, assume that brokers are *honest*. This means brokers always give their customers the right advice about their actual position on the unit circle. Then, as in the standard full-information setting, firm i's

and permits may limit the number of brokers.

[11] The fee covers the brokers' cost. In insurance markets, for instance, a risk analysis is usually expensive since it requires expertise not only in finance, but also in actuarial science, law, and engineering. For details, see Cummins and Doherty (2006), p. 392.

demand is given by

$$D_i\left(p_{i-1}, p_i, p_{i+1}\right) = \frac{p_{i+1} + p_{i-1} - 2p_i + \frac{2}{n}}{2}. \tag{28}$$

Firm i seeks to maximize $\Pi_i\left(p_{i-1}, p_i, p_{i+1}\right) = \left(p_i - c\right) D_i - f$. Differentiating with respect to p_i and setting $p_i = p_{i-1} = p_{i+1} = p$ yields $p^* = c + \frac{1}{n}$, which leads to $D_i = D_{i-1} = D_{i+1} = D = \frac{1}{n}$ implying $\Pi\left(n\right)_i = \frac{1}{n^2} - f$. We may then determine the equilibrium number of firms from the zero-profit condition for existing firms, $\left(p - c\right)/n - f = 0$, which implies the equilibrium number of firms $n_0 = \sqrt{\frac{1}{f}}$ and equilibrium market price $p_0 = c + \sqrt{f}$. This equilibrium seems, however, unrealistic from the brokers' view. Why should they earn zero profits and act honestly? Therefore, we assume in the following that brokers are corrupt and miscounsel consumers in favor of a firm against some price.

6.4 The Conduct of the Firms

As discussed above, many sales practices observed in markets for complex products sail fairly close to the wind. Hence, one is probably concerned to what degree competitors might dishonestly play on informational asymmetries leaving consumers ripped off? Given that products are complex and some consumers require the service of information brokers, a representative provider may bribe brokers into advising consumers to purchase its own variety instead of some better-fitting competitor's. If this is a rational strategy, any broker will encounter such fraudulent attempts. It is not unrealistic to assume that when a broker participates in such collusive agreements with firms, miscounseling consumers is associated with additional efforts. This is because brokers need to convince consumers about their best-fitting product variety which becomes more difficult the less a product actually fits a consumer's preferences. We assume that brokers try to convince consumers to be "located" at the respective indifference border $\overline{x}_i$ instead of their "real" position. For the sake of simplicity, suppose that the cost of miscounseling increases linearly with the degree of miscounseling, i.e. the distance

$|x - \overline{x}_i|$. Interestingly, though purely competitive for consumers' information demand, brokers enjoy market power in their miscounseling activities. Since consumers' demand for information from brokers is given by $1 - F(k)$ at any $x \in [0, 1)$ and distributes symmetrically among m brokers, every one is left with control on $\nu_x(m, k) := (1 - F(k))/m$ of demand density. In respect of miscounseling cost—which is increasing linearly in the degree of miscounseling—consumers who are located closely to an indifferent consumer $\overline{x}_i$ are more attractive victims. Every broker j holds control on a share $\frac{1}{m}$ of potential victims. Thus, if firms refuse a broker's miscounseling fee, they cannot revert to a different broker who miscounsels potential victims of other firms. As a consequence, any broker acts as a monopolist offering miscounseled victims to firms. The miscounseling market can therefore be seen as quasi monopolistically competitive.

We do not specify the brokers' behavior in more detail here, but instead suppose only that brokers charge some markup $\sigma > 0$ on their cost of miscounseling.[12] The brokers' behavior is studied later in section 6.6.

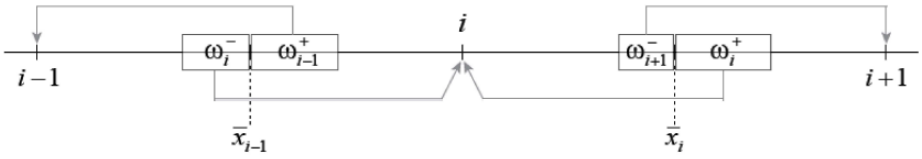

Figure 1: Consumers near the indifference border fall victim to miscounseling.

A representative firm i may decide to attract an upstream ω_i^+ and downstream ω_i^- amount of consumers from its adjacent competitors through miscounseling. Mathematically, taking into account that merely a fraction

[12] Firms could offer location-specific "bids" to brokers for miscounseling. We assume, however, that such location-specific "bids" are too costly to implement.

$1-F(k) =: \lambda_k$ of consumers even asks for informational advice from a broker, brokers' miscounseling intervals are expanded by a factor $\frac{1}{\lambda_k}$ in order to satisfy a provider's request for business stealing. In order to capture the market share $\left(\omega_i^+ + \omega_i^-\right)$, an upstream $\left[\overline{x}_i, \overline{x}_i + \omega_i^+/\lambda_k\right]$ and downstream $\left[\overline{x}_{i-1} - \omega_i^-/\lambda_k, \overline{x}_{i-1}\right]$ interval of miscounseling is required. Consequently, providers' cost of business stealing are given by

$$\mu\left(\omega_i^+, \omega_i^-\right) = \int_{\overline{x}_i}^{\overline{x}_i + \omega_i^+/\lambda_k} (1+\sigma)\left(x - \overline{x}_i\right) dx + \int_{\overline{x}_{i-1}}^{\overline{x}_{i-1} - \omega_i^-/\lambda_k} (1+\sigma)\left(x - \overline{x}_{i-1}\right) dx \tag{29}$$

which reduces to

$$\mu\left(\omega_i^+, \omega_i^-\right) = \frac{(1+\sigma)\left(\left[\omega_i^+\right]^2 + \left[\omega_i^-\right]^2\right)}{2\lambda_k^2}. \tag{30}$$

Given the locally competitive environment $P_i := (p_{i-1}, p_i, p_{i+1}) \in [0, \infty)^3$ and $\Omega_i := \left(\omega_i^+, \omega_i^-, \omega_{i-1}^+, \omega_{i+1}^-\right) \in [0, \frac{1}{2})^4$, demand of a representative firm is given by

$$D_i\left(P_i, \Omega_i\right) = \frac{p_{i+1} + p_{i-1} - 2p_i + \frac{2}{n}}{2} + \omega_i^+ + \omega_i^- - \left(\omega_{i-1}^+ + \omega_{i+1}^-\right) \tag{31}$$

which implies the following profit maximization problem for firm i:

$$\underset{\left\{p_i, \omega_i^+, \omega_i^-\right\}}{max} \Pi_i\left(P_i, \Omega_i\right) = \left(p_i - c\right) D_i\left(P_i, \Omega_i\right) - \mu\left(\omega_i^+, \omega_i^-\right) - f. \tag{32}$$

The associated first-order conditions are:

$$\frac{p_{i-1} + p_{i+1} - 4p_i}{2} + \omega_i^+ + \omega_i^- - \left(\omega_{i-1}^+ + \omega_{i+1}^-\right) = -\frac{1+nc}{n}, \tag{33}$$

$$\lambda_k^2\left(p_i - c\right) = (1+\sigma)\,\omega_i^+, \tag{34}$$

$$\lambda_k^2\left(p_i - c\right) = (1+\sigma)\,\omega_i^-. \tag{35}$$

Note that from the adjacent competitors' first-order conditions, we know

ω_{i-1}^{+} and ω_{i+1}^{-}. Inserting ω_i^{+} and ω_i^{-} into (33) gives

$$\underbrace{-\frac{2\left(1+\sigma-\lambda_k^2\right)}{(1+\sigma)}}_{=:\phi} p_i + \underbrace{\frac{1+\sigma-2\lambda_k^2}{2\left(1+\sigma\right)}}_{=:\psi}\left(p_{i-1}+p_{i+1}\right) = -\frac{1+nc}{n}. \tag{36}$$

As there are $i = 1, 2, ..., n$ firms in the market, we have a n-dimensional algebraic systems of first-order conditions:

$$\left\{\phi p_i + \psi\left(p_{n+1-1[n]} + p_{i+1[n]}\right) = b\right\}_{i=1,2,...,n}. \tag{37}$$

If system (37) is written in its matrix representation

$$\mathbf{A}\cdot\mathbf{p} = \mathbf{b}, \tag{38}$$

where

$$\mathbf{p} = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \in \mathbb{R}^n, \mathbf{A} = \begin{pmatrix} \phi & \psi & 0 & \cdots & 0 & \psi \\ \psi & \phi & \psi & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \psi & \phi & \psi \\ \psi & 0 & \cdots & 0 & \psi & \phi \end{pmatrix} \in \mathbb{R}^{n\times n}, \mathbf{b} = \begin{pmatrix} b \\ \vdots \\ b \end{pmatrix} \in \mathbb{R}^n. \tag{39}$$

The tridiagonal structure of $\mathbf{A}$ implies that the solution is a symmetric equilibrium $\left\{p_i = p,\, \omega_i^{+} = \omega_i^{-} = \omega,\, \Pi_i = \Pi\right\}_{i=1,2,...,n}$ where

$$p_{bs} = \frac{b}{\phi+2\psi} = c + \frac{1}{n}, \tag{40}$$

$$\omega_{bs} = \frac{\lambda_k^2}{(1+\sigma)\,n},\ and \tag{41}$$

$$\Pi_{bs} = \frac{1}{n^2}\left(1 - \frac{\lambda_k^2}{1+\sigma}\right) - f. \tag{42}$$

As can easily be seen from the above equations, equilibrium profits are lower than in the standard full-information model. Given a market with free entry,

long-term equilibrium profits are zero. Thus, there are

$$n_{bs} = \sqrt{\frac{1}{f}}\sqrt{1 - \frac{\lambda_k^2}{1+\sigma}} < \sqrt{\frac{1}{f}} = n_0. \tag{43}$$

firms in the zero-profit market equilibrium. Subscript bs indicates that n_{bs} is the number of firms in a business-stealing equilibrium. Apparently, the opportunity of business stealing through miscounseling uninformed consumers leads to less firms in long-run equilibrium. Dishonest brokers decrease product variety in the market. This implies higher prices in market equilibrium under miscounseling, i.e. $n_{bs} < n_0$ leads to $p_{bs} = c + \frac{1}{n_{bs}} > p_0 = c + \frac{1}{n}$. This implies

Proposition 9. *Firms are in a prisoners' dilemma of business stealing. Gross profits are lower compared to the standard Salop model under full consumer information. If binding contracts on "fair competition" were possible, miscounseling could be abandoned and consumers and firms would be made better off.*

As a result, bribery implies that each firm is left with greater market power. We also conclude that the impact of business stealing via intermediary miscounseling on market equilibrium is lower if brokers' markups for miscounseling are higher $\sigma \uparrow$ or if consumers' search cost is lower $\lambda_k \downarrow$.

Misselling leads, of course, to welfare losses. In the following, we specify total social cost of miscounseling. When the number of firms decreases, equilibrium price and aggregated mismatch cost of consumers increase. Since aggregated demand is normalized to one, social cost of the price increase is

$$\triangle p = c + \frac{1}{n_{bs}} - \left(c + \frac{1}{n_0}\right) = \sqrt{f}\left(\underbrace{\sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}}}_{>1} - 1\right) > 0 \tag{44}$$

while additional mismatch cost is

$$\triangle m = 2n_f \int_0^{\frac{1}{2n_{bs}}-\omega} x\,dx + 2n_{bs} \int_0^{\frac{1}{2n_{bs}}+\omega} \frac{1}{2n_{bs}}\,x\,dx - 2n_0 \int_0^{\frac{1}{2n_0}} x\,dx, \quad (45)$$

or

$$\triangle m = \frac{\left(\frac{8\lambda_k^3}{(1+\sigma)^2}+1\right)}{4n_{bs}} - \frac{1}{4n_0}, \quad (46)$$

which is equal to

$$\frac{\sqrt{f}}{4}\left(\sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}}\left(1+\frac{8\lambda_k^2}{(1+\sigma)^2}\right)-1\right) > 0. \quad (47)$$

Treating brokers' earnings from miscounseling as transfers, total social cost of business stealing (SC_{bs}) is

$$SC_{bs} = \triangle p + \triangle m = \frac{\sqrt{f}}{4}\left(\sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}}\left(5+\frac{8\lambda_k^2}{(1+\sigma)^2}\right)-5\right) > 0. \quad (48)$$

We summarize these findings in

Proposition 10. *Total social cost of business stealing is lower if the brokers' miscounseling markup is higher* $\uparrow$ *or if consumers' search cost is lower* $k \downarrow$.

6.5 The Impact of Side Payments

In the business-stealing equilibrium, every firm gains 2ω demand from its adjacent competitors. Yet, in an equal manner, every firm looses 2ω demand to its adjacent competitors. Hence, one might ask why firms—being worse off than without miscounseling—cannot fight business stealing? Given the business-stealing equilibrium, a firm might envisage to counter intermediary miscounseling through side payments. Under what conditions is this strategy profitable? Brokers would accept side payments if their profits under side payments are higher than in the business-stealing equilibrium.

Suppose an anti-miscounseling agreement between a firm and a broker can be made on condition that brokers receive compensatory payments in return

for lost profits. From (30), setting $\omega_i^+ = \omega_i^- = \omega$, follows that the brokers' profits of miscounseling 2ω consumers in disfavor of a representative firm are

$$\pi(2\omega) = \frac{\sigma\omega^2}{\lambda_k^2}, \tag{49}$$

which represents the minimum side payment required to prevent miscounseling. The firm's profits generated by 2ω demand are

$$\Pi(2\omega) = 2\omega(p-c) = \frac{2\omega}{n}. \tag{50}$$

Side payments are taken into account if generated profits outweigh the firm's expenses, i.e., if $\Pi(2\omega) \geq \pi(2\omega)$. Consequently, an agreement is in force if

$$\sigma \leq \frac{2\lambda_k^2}{\omega n}, \tag{51}$$

which—using (41)—reduces to $\sigma \geq -2$ which is always true for any $\sigma \geq 0$. We summarize this result in

Proposition 11. *The business-stealing equilibrium is not stable when firms envisage side payments in order to incentivize brokers to give consumers right advice. Bribing brokers into correctly counseling consumers is a dominant strategy. As a result, the market is free of miscounseling. There are, however, high side payments which increase prices and make consumers worse off compared to the standard Salop full-information market equilibrium.*

Given that $\sigma \geq -2$, all firms fight business stealing through side-payments. Then (49) together with (41) implies the following profit maximization problem

$$\max_{p_i} \left[(p_i - c) D_i(P_i) - \frac{\sigma\lambda_k^2}{(1+\sigma)^2 n^2} - f \right]. \tag{52}$$

The first-order condition is

$$-\frac{2}{t}p_i + \frac{1}{2}(p_{i-1} + p_{i+1}) = -\frac{1+nc}{n} \quad i = 1, 2, \ldots, n. \tag{53}$$

Solving (53) analogously to (37), we obtain a symmetric equilibrium $\Pi_i = \Pi$

and $p_i = p$, given $i = 1, 2, \ldots, n$, where

$$p_{sp} = c + \frac{1}{n}, \tag{54}$$

$$\Pi_{sp} = \frac{1}{n^2}\left(1 - \frac{\sigma\lambda_k^2}{(1+\sigma)^2}\right) - f, \tag{55}$$

implying

$$n_{sp} = \sqrt{\frac{1}{f}}\sqrt{1 - \frac{\sigma\lambda_k^2}{(1+\sigma)^2}} < \sqrt{\frac{1}{f}} = n_0 \tag{56}$$

as the number of firms in the zero-profit side-payments equilibrium. Again, side-payments reduce the equilibrium number of firms leading to higher prices and additional mismatch cost, i.e., $n_{sp} < n_0$ leads to $p_{sp} = c + \frac{1}{n_{sp}} > c + \frac{1}{n_O} = p_0$, even though these are less compared to miscounseling (since all consumers buy their most preferred product variety). Indeed, social cost of the price increase is

$$\triangle p = \left(c + \frac{1}{n_{sp}} - c - \frac{1}{n_0}\right) = \sqrt{f}\left(\underbrace{\sqrt{\frac{(1+\sigma)^2}{(1+\sigma)^2 - \sigma\lambda_k^2}}}_{>1} - 1\right) > 0, \tag{57}$$

and additional mismatch cost is

$$\triangle m = 2n_{sp}\int_0^{\frac{1}{2n_{sp}}} x\,dx - 2n_0\int_0^{\frac{1}{2n_0}} x\,dx, \tag{58}$$

which is equal to

$$\frac{1}{4}\left(\frac{1}{n_{sp}} - \frac{1}{n_0}\right) = \frac{\sqrt{f}}{4}\left(\sqrt{\frac{(1+\sigma)^2}{(1+\sigma)^2 - \sigma\lambda_k^2}} - 1\right) > 0. \tag{59}$$

Therefore, total social cost generated by side-payments is

$$SC_{sp} = \frac{5\sqrt{f}}{4}\left(\sqrt{\frac{(1+\sigma)^2}{(1+\sigma)^2 - \sigma\lambda_k^2}} - 1\right) > 0. \tag{60}$$

This leads us to the following

Proposition 12. *Side payments are welfare improving compared to miscounseling. Social costs under business stealing and miscounseling are always higher than social cost under side-payments because all consumers buy their most preferred product variety.*

Proof. To show that social cost under side-payments is lower, we consider the difference

$$\begin{aligned} SC_{bs} - SC_{sp} &= \frac{5\sqrt{f}}{4}\left(\sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}}\left(1+\frac{8\lambda_k^2}{5\left(1+\sigma\right)^2}\right)-\sqrt{\frac{\left(1+\sigma\right)^2}{\left(1+\sigma\right)^2-\sigma\lambda_k^2}}\right) \\ &\geq \frac{5\sqrt{f}}{4}\left(\sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}}-\sqrt{\frac{\left(1+\sigma\right)^2}{\left(1+\sigma\right)^2-\sigma\lambda_k^2}}\right), \end{aligned}$$

which is equivalent to show that

$$\begin{aligned} \Leftrightarrow \sqrt{\frac{1+\sigma}{1+\sigma-\lambda_k^2}} &\geq \sqrt{\frac{\left(1+\sigma\right)^2}{\left(1+\sigma\right)^2-\sigma\lambda_k^2}} \\ \Leftrightarrow \left(1+\sigma\right)^2-\sigma\lambda_k^2 &\geq \left(1+\sigma\right)\left(1+\sigma-\lambda_k^2\right) \\ \Leftrightarrow \lambda_k^2 &\geq 0. \end{aligned}$$

Since λ_k^2 is always positive, the above inequality always holds. □

6.6 The Conduct of the Brokers

Finally, we haven't analyzed the brokers' behavior in detail in the preceding sections. In the preceding sections, the brokers' miscounseling markup has been some variable parameter allowing us to conclude that externalities of business stealing are lower when brokers' markup is higher. Moreover, we found that business stealing becomes unprofitable under side payments. Certainly, when business stealing is more expensive, such kind of unfair competition inducing externalities on consumers becomes less attractive. To complete the analysis, using broker profit maximization, let us determine the

actual markup σ. Since the business stealing market is monopolistically competitive, we find that brokers earn monopoly profits. From (55), monopoly profit of a representative broker j is

$$\pi_j(\sigma_j) = \frac{n\sigma_j\omega_{bs}^2}{m\lambda_k^2}. \tag{61}$$

Substituting for (41) gives the following profit maximization problem

$$\max_{\sigma_j}\left[\frac{\lambda_k^2}{\sqrt{\frac{1}{f}}}\frac{\sigma_j}{(1+\sigma_j)^2\sqrt{1-\frac{\lambda_{2_k}}{1+\sigma_j}}}\right]. \tag{62}$$

This function has a unique maximum at

$$\sigma^* = \frac{1}{4}\left(\lambda_k^2 + \sqrt{\lambda_k^4 - 16\lambda_k^2 + 16}\right). \tag{63}$$

Thus, $\sigma^* \in [0,1]\ \forall\lambda \in (0,1)$. Finally, the number of firms in market equilibrium under side payments is then given by

$$n^* = \sqrt{\frac{1}{f}}\sqrt{1 - \frac{\frac{1}{4}\left(\lambda_k^2 + \sqrt{\lambda_k^4 - 16\lambda_k^2 + 16}\right)\lambda_k^2}{\left(\frac{1}{4}\left(\lambda_k^2 + \sqrt{\lambda_k^4 - 16\lambda_k^2 + 16}\right) + 1\right)^2}} < \sqrt{\frac{1}{f}} = n_0. \tag{64}$$

Equilibrium prices and profits are given by (54) and (55) together with n^*.

In summary, we have found that a long-run market equilibrium—given a market where bribable brokers might miscounsel consumers against some fee and firms might counter business stealing through side payments to brokers—will probably be characterized by the following properties:

1. Every consumer purchases the product that actually matches his preferences.

2. There is no miscounseling in equilibrium but every firm makes side payments to brokers to counter them from business stealing.

3. Prices are higher than in a standard full-information equilibrium since firms include side payments to brokers in their prices.

4. Due to side payments, equilibrium gross profits are lower and thus the number of firms in market equilibrium is also lower than in the standard full-information equilibrium.

5. Brokers make high profits from side payments.

6.7 Concluding Remarks

In this paper, we have analyzed the impact of fraudulent independent experts on market performance and efficiency in a Salop-market of monopolistic competition. In such a market, experts may want to give the wrong advice to their customers in order to receive collusive business stealing payments for doing so by firms, thereby preventing consumers from searching for a better product variety offered by other firms. This argument applies to various industries such as repair markets, health care markets, and financial services markets where consumers are usually uninformed about which of the offered product varieties best fits their individual needs.

Our analysis provides several interesting insights. First, fraudulent independent experts imply fewer firms in market equilibrium, which is associated with greater market power and higher prices compared to the standard Salop market equilibrium under full consumer information. Second, as a consequence, bribable independent experts entail welfare losses for both consumers and firms. More generally, due to the presence of fraudulent experts, the parties are caught in a prisoner's dilemma: firms and consumers would be better off without fraudulent experts. We also show that an equilibrium without miscounseling may require side payments from the firms to the brokers. Therefore, rigorous regulation of the relationship between firms and intermediaries as well as severe fines in case of detection of bribery seem necessary in order to make these markets operate more efficiently.

7 Some Economics of Originality

Abstract

This paper formulates a simple model concerning the choice of originality. The model assumes that an agent (e.g., a researcher) chooses the level of originality of his work (e.g., a research project). Less original work is more likely to succeed, but the potential benefit is at the same time smaller. More original work is less likely to succeed, but it can break totally new and fruitful ground. The model predicts that the level of originality may increase with the number of potential beneficiaries, decreases with the number of people to be persuaded, and generally increases with the ability level of the agent. The introduction of uncertainty related to the exogenous parameters are indeterminate in sigh, except under special circumstances. The role of originality as a signal of ability is also discussed.

Keywords: *Originality, uncertainty, signaling*

JEL classification: A11, B41

7.1 Introduction

> "Originality means difference, not improvement, and one may invent new errors as well as new truths." George Stigler (1955, p. 301)

Originality is a fascinating subject. Virtually every aspect of life is in some form affected. The complexity of our general environment and its rate of change are rapidly increasing. Many of today's world problems require (more) original approaches than those provided in the past. Additionally, originality is generally portrayed as a desirable personal trait. In contrast, historic records offer numerous examples of original ideas that were keenly rejected. A layman's belief about originality is shaped by the ambiguous and inaccessible nature of the concept, and is generally pessimistic about the science's potential to shed some light on it.

A central element of economics is respect for the power of incentives. The economic approach explains behavior by pointing at the incentives that lie behind that behavior, and economists predict that the extent of an activity will increase when the returns to that same activity rise. This paper aims to apply the economic approach to the choice of originality by analyzing the choice problem of an agent considering different degrees of originality in order to solve a given problem.

The motivating example is the case of a researcher and his choice problem concerning the level of originality of his research. He can undertake standard research, which refines, extends, and tests previously accepted research, and he can undertake original research, which may break new ground, but is also more likely to incur significant costs, to fail, and/or to be rejected. For example, consider the situation of an economic theorist. He can choose to be not very original, making a conceptually minor but mathematically challenging extension to some well-known model. On the other side, he can start with novel assumptions, and build original models in order to make novel predictions, or explain previously discovered empirical results. In general, originality is best described as a continuous variable.

The model in this paper assumes that a higher level of originality is more risky, in terms of the probability of success, than a more standard approach, but we also assume that the benefit to the researcher increases with the level of originality. The main predictions of the model under certainty are that in equilibrium the level of originality may increase with the number of potential beneficiaries, decreases with the number of people to be persuaded and the cost of persuasion, and generally increase with the ability level of the researcher. More "risky" values of the exogenous parameters may decrease or increase the level of originality in equilibrium.

Originality may also serve as a signaling device. Ability is in many situations not readily recognized—at least in the short run. Originality may therefore often serve as a signal of ability or talent.

Though in the following we shall speak of a researcher and his research approach, the reasoning can be applied far more broadly; that is, it can be

applied to any decision maker faced with a choice with respect to originality. For example, a politician concerned with the solution of a given economic problem faces a structurally similar problem. A regulator trying to design a regulatory mechanism is also concerned with the question of originality. Yet another example would be an artist choosing the level of originality of his art.

The paper is organized as follows. Section 7.2 discusses the meaning of originality and its relation to creativity. The formal model is presented in section 7.3. A specific example is discussed in section 7.4. Section 7.5 discusses originality as a signaling device. The conclusion and a short discussion of some possible extensions follow.

7.2 The Meaning and the Literature on Originality

The subject of creativity has attracted much attention. Human beings generally place an enormous emphasis upon original approaches, views, and discoveries in an wide range of activities; the most important examples are scientific discoveries, technological innovations, and arts. On the other hand, it is also a well-known fact that many originators in different fields had become so frustrated by the reaction to their contributions that their lives were characterized by personal tragedies. The fascination with originality is an evident fact despite the lack of either a precise definition or an objective measure of "originality." Until recently, work in the area of creativity has been mainly dominated by psychologists. On the one hand, there is an extensive amount of work trying to measure creativity in individuals by using tests modeled after intelligence tests. On the other hand, there is much work trying to determine the psychological traits of creative individuals. Besides psychology, studying questions related to originality is in the domain of sociology, cognitive sciences, artificial intelligence, philosophy, history, law, economics, arts, and business.

In a first step towards a clarification, it is of crucial importance to distinguish originality from the related concept of creativity. One of the few points of agreement in the relevant psychological literature is that creativity is a

multi-faceted phenomenon (for a discussion, see Mumford and Gustafson (1988) and Boden (1996)). As one of the various facets of creativity, originality is the most widely recognized. Originality is not, however, a sufficient criterion of creativity. By itself originality may imply bizarre and obviously inappropriate thoughts, approaches, or behavior. In other words, creativity requires originality *as well as* (some form or some degree of) correctness, usefulness, and/or acceptance. In economist's terminology, creativity requires in addition to originality some value added. Sternberg (1999) offers a thorough overview of the literature concerned with various questions related to creativity.

To illustrate the point on a well-known example, consider Einstein's general theory of relativity that was obviously an original approach to the analysis of physical reality, and which proved to be better suited to empirical observations than the classical Newtonian physics and also made some novel predictions that were later supported by experiments and observations.

The topic of originality as an object of interest in its own right and a question of rational choice has been mainly ignored by economists. There are only a few papers that discuss and formalize some elements of the process by which originality is evaluated. Stigler (1955) offers a superb discussion of the role of originality in scientific progress illustrated by cases taken from economics. Evenson and Kislev (1976) model the experimentation process in applied sciences. Lazear (1997) looks at the effects of various award schemes in basic research. Ellison (2002) models the academic publishing process. Mialon and Mialon (2005) model the interaction between an author and a reviewer of scientific research.

7.3 The Model

In this section we present a simple economic model concerning the choice of originality. The approach taken here follows the economist's standard analysis of choice and assumes that a researcher employs a more original approach to a given research question if the expected utility to him exceeds the utility he could get by choosing a less original approach. The model

ignores a number of (potentially) relevant factors (e.g., various idiosyncratic parameters). Nonetheless, the model provides a reasonable platform for thinking formally about the issue at hand in economic terms.

Consider the situation of a risk-neutral[13] researcher engaged in scientific research. He can undertake standard research, that is, he can refine, extend, and test previously accepted theoretical and/or empirical research, and he can undertake original research by employing novel assumptions, perspectives, and methods, which may break new ground, but is also much more likely to fail and/or to be rejected.

Let $0 < i < 1$ index the level of originality, where $i \to 0$ indicates the total lack of originality, and $i \to 1$ indicates complete originality. The probability of being creative (being original as well as obtaining a correct result) is $0 \leq p(\theta, i) \leq 1$, and it is a function of i and θ (the probability of failure is thus $1 - p(\theta, i)$), where $0 < \theta < 1$ denotes the exogenous level of the ability or talent of the researcher; higher levels of θ indicate a higher degree of ability or talent. Another possible interpretation of θ could be experience or status. Yet another possible interpretation of θ is the level of intrinsic interest in the given problem.

Research characterized by some degree of originality is subject to following risks and difficulties:

1. *Risk of failure.* Research, by its very nature, often fails. It is plausible to assume that this risk is higher the higher the level of originality.

2. *Risk of rejection.* The results of research efforts can be rejected by the scientific community despite their intrinsic characteristics. It is reasonable to assume that this risk is also higher the higher the level of originality.

3. *Strategic risk.* Research results can be obtained by some other researcher. It this case, it is plausible to assume that this risk is lower the higher the level of originality.

[13]This assumption may seem doubtful since the gains and losses may be huge. But the research sector presumably attracts the least risk-averse individuals, so the assumption may not be so strong after all.

4. *Psychological obstacles.* Overcoming obstacles associated with finding original solutions to (new) problems is often very hard, and we are often unable to break out of our customary solution-seeking processes (see Simon (1969); Tversky and Kahneman (1974)). It is rather uncontroversial that these obstacles increase with the degree of originality.

We assume that $p(\theta, i)$ is twice continuously differentiable in both arguments. Taking all the factors from the above list into account, it is plausible to assume that $E\left[\frac{\partial p(\theta,i)}{\partial i}\right] < 0$ and $E\left[\frac{\partial^2 p(\theta,i)}{\partial i^2}\right] < 0$ for all i, where $E[\bullet]$ denotes the expectation operator with respect to the probability distribution over $p(\theta, i)$ induced by θ and i. The sign of $E\left[\frac{\partial p(\theta,i)}{\partial \theta}\right]$ is *per se* not clear. We will assume that $E\left[\frac{\partial p(\theta,i)}{\partial \theta}\right] \geq 0$ for all θ.

The benefit to the researcher $B(N, i)$ from research is presumably a function of N and i, where $0 < N < \infty$ denotes the size of the population of direct/indirect beneficiaries. The population of beneficiaries can consists of other researchers, academics, students, officials in governmental and non-governmental organizations, firms, et cetera. In other words, N denotes the extent of the market. The extent of the market will, of course, vary with the topic; some academic work, for instance, is often of direct interest only to, often a small number of, other academics. The market here is unusual in many ways, above all because no one pays directly for what academics produce. Journals generally do not compensate contributors for articles and essays. Publishers will pay for the right to publish books, and professors receive royalties. However, little money is usually involved, and the motivation for writing books is rarely monetary for academics. On the other hand, indirect compensation, monetary and non-monetary, such as fame, job opportunities, consulting projects, prizes, medals, memberships in honorary academies, invitations to conferences, and the like, is omnipresent—and it is generally linked to the number of indirect beneficiaries. The researcher's benefit in the case of a failure is assumed to be 0.

We assume that $B(N, i)$ is a twice continuously differentiable in both arguments. Further, assume that $\frac{\partial B(N,i)}{\partial i} > 0$ and $\frac{\partial^2 B(N,i)}{\partial i^2} < 0$ for all i, $\lim_{i \to 0} \frac{\partial B(N,i)}{\partial i} = \infty$ and $\lim_{i \to 1} \frac{\partial B(N,i)}{\partial i} = 0$. Additionally, we assume $\frac{\partial B(N,i)}{\partial N} > 0$

and $\frac{\partial^2 B(N,i)}{\partial N^2} < 0$ for all N.

The lower the number of people to be persuaded by the research results, and the higher the ability of the researcher, the easier is to be original. We will therefore assume that the costs from research, denoted by $C(M, \theta, i)$, are a function of M, θ, and i, where $0 < M \leq N$ denotes the number of people to be persuaded by the research in order to be accepted. For example, M could denote the number of referees to be persuaded in order to publish a paper in a peer-reviewed journal. We assume that $C(M, \theta, i)$ is a twice continuously differentiable in all arguments. Further, assume that costs increase with the level of originality at an increasing rate, that is, we assume that $\frac{\partial C(M,\theta,i)}{\partial i} > 0$ and $\frac{\partial^2 C(M,\theta,i)}{\partial i^2} \geq 0$ for all i, and $\lim_{i \to 0} \frac{\partial C(M,\theta,i)}{\partial i} = 0$, and $\lim_{i \to 1} \frac{\partial C(M,\theta,i)}{\partial i} = \infty$ to ensure an interior solution. Additionally, we assume $\frac{\partial C(M,\theta,i)}{\partial M} > 0$ for all M, $\frac{\partial C(M,\theta,i)}{\partial \theta} < 0$ for all θ, $\frac{\partial^2 C(M,\theta,i)}{\partial i \, \partial M} > 0$ for all i and M, and $\frac{\partial^2 C(M,\theta,i)}{\partial i \, \partial \theta} < 0$ for all i and θ.

The researcher tries to maximize his expected net benefit, denoted by V, as given by the following decision problem:

$$\max_{i \in (0,1)} E[V] := E[p(\theta, i) B(N, i) - C(M, \theta, i)]. \tag{65}$$

The first-order condition of this problem is given by the following expression, whereby the asterisk denotes the optimal value:

$$E\left[\frac{\partial p(\theta, i^*)}{\partial i} B(N, i^*) + p(\theta, i^*) \frac{\partial B(N, i^*)}{\partial i} - \frac{\partial C(M, \theta, i^*)}{\partial i}\right] = 0. \tag{66}$$

Equation (66) yields an interesting comparative statics result, as summarized in the following proposition

Proposition 13. *An increase in the number of beneficiaries (N) may increase or decrease the equilibrium level of originality.*

Proof. Totally differentiating both sides of equation (66) with respect to N

yields:

$$\frac{\partial i^*}{\partial N} = \frac{E[\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial N} + p\left(\theta, i^*\right)\frac{\partial^2 B(N,i^*)}{\partial i\,\partial N}]}{-E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2}B\left(N, i^*\right) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p\left(\theta, i^*\right)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]}.$$

The denominator of this derivative is just the second-order condition multiplied by -1 and is positive by assumption. Thus the numerator determines the sign of $\frac{\partial i^*}{\partial N}$. By assumption, $E\left[\frac{\partial p(\theta,i)}{\partial i}\right] < 0$ and $\frac{\partial B(N,i)}{\partial N} > 0$. Thus, the sign of $\frac{\partial i^*}{\partial N}$ is only positive if $\frac{\partial^2 B(N,i)}{\partial i\,\partial N}$ is also positive and high enough in magnitude. □

In other words, the level of originality will increase with the extent of the market if the marginal benefit to originality is positive and sensitive enough to N.

The perfect knowledge of the exogenous parameters is a very restrictive assumption. This is especially the case for the value of N. For example, basic research is typically characterized by great uncertainty concerning the real value of N. It is therefore interesting to analyze the change in behavior as the parameter becomes "risky", with the mean remaining at the level under certainty. In order to do so, we will employ the concept of a mean-preserving spread. As a distribution undergoes a mean-preserving spread, the variance (if it exists) increases. A mean-preserving spread seems a reasonable way to formalize the notion of an "increase in riskiness" provided one has no particular interest in the behavior of the third and higher order moments (for a discussion of the concept, see Rothschild and Stiglitz (1970)). Operationally, a mean-preserving spread can be obtained by adding a random variable with conditional mean zero to the original random variable.

Let α be a parameter that represents a mean-preserving spread to the distribution of N, that is, we shall assume that the standard deviation of the number of (potential) beneficiaries is some multiple α of its mean, N, i.e., $\sigma = \alpha \times N$. We can now analyze the effect of an mean-preserving spread, as summarized in the following

Proposition 14. *A mean-preserving spread in N may increase or decrease*

the equilibrium level of originality.

Proof. The first-order condition given by equation (66) defines a function $i^* = f(\alpha, \bullet)$. Differentiating both sides of equation (66) with respect to α, we get

$$\frac{\partial i^*}{\partial \alpha} = \frac{\frac{\partial E\left[\frac{\partial p(\theta,i^*)}{\partial i} B(N,i^*) + p(\theta,i^*)\frac{\partial B(N,i^*)}{\partial i}\right]}{\partial \alpha}}{-E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2} B(N,i^*) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p(\theta,i^*)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]}.$$

The denominator of this derivative is just the second-order condition multiplied by -1 and is positive by assumption. Thus the numerator determines the sign of $\frac{\partial i^*}{\partial \alpha}$. Generally, it is not possible to unambiguously determine the sign of the derivative without an additional assumption. If $\frac{\partial V}{\partial i}$ is a strictly convex function of N, then i^* will increase as N undergoes a mean-preserving spread. $\frac{\partial V}{\partial i}$ is a strictly convex in N if $B(N,i)$ is a strictly concave function in N, and if $\frac{\partial B(N,i)}{\partial i}$ is a strictly convex function in N. Similarly, $\frac{\partial i^*}{\partial \alpha} < 0$ if $\frac{\partial V}{\partial i}$ is a strictly concave function of N. That is the case if $B(N,i)$ is a strictly convex function in N, and if $\frac{\partial B(N,i)}{\partial i}$ is a strictly concave function in N. □

Additionally, equation (66) yields further predictions, as summarized in the following proposition.

Proposition 15. *A decrease in the number of persons to be persuaded by the research (M) raises the equilibrium level of originality.*

Proof. Totally differentiating both sides of equation (66) with respect to M yields:

$$\frac{\partial i^*}{\partial M} = \frac{\frac{\partial^2 C(M,\theta,i^*)}{\partial i\,\partial M}}{E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2} B(N,i^*) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p(\theta,i^*)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]} < 0.$$

□

The case of an uncertain value of M is quite possible. It is therefore interesting to analyze the change in behavior as the parameter becomes "risky", with the mean remaining at M. We once again will employ the concept of a mean-preserving spread. Let β be a parameter that represents a mean-preserving spread to the distribution of M. We can now analyze the effect of an mean preserving spread, as summarized in the following

Proposition 16. *A mean-preserving spread in M may increase or decrease the equilibrium level of originality.*

Proof. The first-order condition given by equation (66) defines a function $i^* = f(\beta, \bullet)$. Differentiating both sides of equation (66) with respect to β yields

$$\frac{\partial i^*}{\partial \beta} = \frac{\frac{\partial E\left[\frac{\partial C(M,\theta,i^*)}{\partial i}\right]}{\partial \beta}}{E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2} B(N,i^*) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p(\theta,i^*)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]}.$$

The denominator of this derivative is just the second-order condition and is negative by assumption. Thus the numerator determines the sign of $\frac{\partial i^*}{\partial \beta}$. Generally, it is not possible to unambiguously determine the sign of the derivative without an additional assumption. If $\frac{\partial C(M,\theta,i)}{\partial i}$ is a strictly concave function in M, then the numerator will be unambiguously negative. Thus, $\frac{\partial i^*}{\partial \beta} > 0$. Similarly, $\frac{\partial i^*}{\partial \beta} < 0$ if $\frac{\partial C(M,\theta,i)}{\partial i}$ is a strictly convex function in M. □

In addition, equation (66) yields further and interesting predictions, as summarized in the following proposition.

Proposition 17. *An increase in the level of ability (θ) generally increases the equilibrium level of originality.*

Proof. Totally differentiating both sides of equation (66) with respect to θ yields:

$$\frac{\partial i^*}{\partial \theta} = \frac{E[\frac{\partial^2 p(\theta,i^*)}{\partial i\,\partial \theta} B(N,i^*) + \frac{\partial p(\theta,i^*)}{\partial \theta}\frac{\partial B(N,i^*)}{\partial i} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i\,\partial \theta}]}{-E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2} B(N,i^*) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p(\theta,i^*)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]}.$$

The numerator determines the sign of $\frac{\partial i^*}{\partial \theta}$. The equilibrium level of originality will increase with the level of θ if $E\left[\frac{\partial^2 p(\theta,i)}{\partial i\,\partial \theta}\right] \geq 0$, and even if $E\left[\frac{\partial^2 p(\theta,i)}{\partial i\,\partial \theta}\right] < 0$ but small enough in magnitude. □

In other words, proposition 17 states that originality will increase if the marginal probability of success is positively linked to ability; i.e., when originality and ability are complements or at least not strong substitutes.

Once again, it is worthwhile to analyze the change in behavior as the parameter θ becomes "risky", with the mean remaining at θ. The ability of a researcher need not be constant, but may be determined as the outcome of a stochastic process. Even casual experience reveals that the existence of surprising changes in a researcher's ability is a reasonable assumption. We once again will employ the concept of a mean-preserving spread. Now let γ be a parameter that represents a mean-preserving spread to the distribution of θ. We can now analyze the effect of an mean preserving spread in θ, as summarized in the following

Proposition 18. *A mean-preserving spread in θ may increase or decrease the equilibrium level of originality.*

Proof. The first-order condition given by equation (66) defines a function $i^* = f(\gamma, \bullet)$. Differentiating both sides of equation (66) with respect to γ yields

$$\frac{\partial i^*}{\partial \gamma} = \frac{\frac{\partial E\left[\frac{\partial p(\theta,i^*)}{\partial i} B(N,i^*) + p(\theta,i^*)\frac{\partial B(N,i^*)}{\partial i} - \frac{\partial C(M,\theta,i^*)}{\partial i}\right]}{\partial \gamma}}{-E\left[\frac{\partial^2 p(\theta,i^*)}{\partial i^2} B(N,i^*) + 2\frac{\partial p(\theta,i^*)}{\partial i}\frac{\partial B(N,i^*)}{\partial i} + p(\theta,i^*)\frac{\partial^2 B(N,i^*)}{\partial i^2} - \frac{\partial^2 C(M,\theta,i^*)}{\partial i^2}\right]}.$$

The denominator of this derivative is just the second-order condition multiplied by -1 and is positive by assumption. Thus the numerator determines the sign of $\frac{\partial i^*}{\partial \gamma}$. Once again, it is not possible to unambiguously determine the sign of the derivative without an additional assumption. If $\frac{\partial V}{\partial i}$ is a strictly convex function of θ, then i^* will increase as θ undergoes a mean-preserving spread. $\frac{\partial V}{\partial i}$ is a strictly convex in θ if $\frac{\partial p(\theta,i)}{\partial i}$ and $p(\theta,i)$ are strictly convex

functions in θ, and if $\frac{\partial C(M,\theta,i)}{\partial i}$ is a strictly concave functions in θ. Similarly, $\frac{\partial i^*}{\partial \gamma}$ is unambiguously negative if $\frac{\partial p(\theta,i)}{\partial i}$ and $p(\theta, i)$ are strictly concave functions in θ, and if $\frac{\partial C(M,\theta,i)}{\partial i}$ is a strictly convex functions in θ. □

7.4 An Example

This section analyzes a specific example which illustrates the principles developed in the preceding section and is simple enough to yield a more precise characterization of the solution. Consider once more a risk-neutral researcher engaged in scientific research. He has to choose an approach to a given research question. As in the previous section, let $0 < i < 1$ index the level of originality, whereby $i \to 0$ indicates the total lack of originality, and $i \to 1$ indicates complete originality. Additionally, let $0 < \theta < 1$ indicate the exogenous level of ability or talent of the researcher, whereby higher levels of θ indicate a higher degree of ability or talent. The probability of success with a given value of i and θ is:

$$p(\theta, i) = 1 - i^{\theta}. \tag{67}$$

The population size of potential beneficiaries of the research output is $N > 0$. Let $b \times i$ denote the value of research output per potential beneficiary. Thus, the value of research is assumed to be increasing in the value of i. The total benefit of having research of type i is therefore $B(N, i) = N \times b \times i$.

Undertaking research of type i costs $C(M, \theta, i) = F + c \times M \times i \times (1 - \theta)$. $C(M, \theta, i)$ has a fixed component $F > 0$ (e.g., the costs of a laboratory), and and a variable component $c \times M \times i \times (1 - \theta)$, where $0 < c \times i \leq b \times i$ denotes the persuasion cost per person.

The researcher faces therefore the following decision problem:

$$\max_{i \in (0,1)} \left[\left(1 - i^{\theta} \right) \times N \times b \times i - F - M \times c \times i \times (1 - \theta) \right]. \tag{68}$$

Solving the first-order condition yields the following result, where the asterisk

denotes the optimal value:

$$i^* = \left(\frac{N \times b - M \times c \times (1-\theta)}{N \times b \times (1+\theta)} \right)^{\frac{1}{\theta}}. \tag{69}$$

Equation (76) implies that an increase in the number of beneficiaries increases the level of originality in equilibrium. The reason is that in this particular case $\frac{\partial^2 B(N,i)}{\partial i \, \partial N} = b > 0$, that is, the marginal benefit to originality is increasing in N.

As shown in the previous section, $\frac{\partial i^*}{\partial M} < 0$, that is, the equilibrium level of originality is decreasing in M.

In this example, the equilibrium level of originality may increase or decrease with the level of ability or talent. The reason for this result is that in this particular case $\frac{\partial^2 p(\theta,i)}{\partial i \, \partial \theta} \lesseqgtr 0$, that is, originality and ability are complements or substitutes, depending on the particular values of i and θ.

7.5 Originality as a Signal

What is the reason for the striking fact that originality is of enormous value in some particular areas? For instance, critics in the field of art are often on the side of originality. The previous analysis cannot answer these questions. A new perspective on the phenomenon of originality is therefore needed.

In many situations, ability or talent cannot be directly observed—at least in the short run. Originality may then serve as a signal of ability or talent. Examples of this inference exist in a variety of areas. While we do not present a model in which originality serves as a signal of ability or talent, such a model is straightforward to construct (the single-crossing condition holds here). This point suggests that originality is even more important in situations where individuals' ability or talent is not readily recognized.

7.6 Conclusion and Extensions

We have formulated a simple model concerning the choice of originality, and performed a comparative static analysis with respect to the number

of (potential) beneficiaries, the number of people to be persuaded, and the level of ability or talent. The motivating example was a researcher having to choose the level or originality of his research. We have shown that the level of originality may increase with the number of potential beneficiaries, decreases with the number of people to be persuaded, and may increase with the ability or talent of the researcher. The incorporation of uncertainty concerning the exogenous variables may introduce some ambiguities. We have also discussed the role of originality as a signal of ability or talent.

The basic reasoning can be applied far more broadly, that is, it applies to any decision maker faced with a choice with respect to originality. It would be interesting to test empirically the predictions of the model.

We conclude by discussing some possible extensions of the basic model. One possible extension would allow for an entire research line, that is, a project that involves $n > 1$ interrelated stages, possibly giving rise to multiple potential directions for further research. Another interesting extension would contain a distinction between academic-sector and private-sector research. For example, it is presumably the case that people from academia value "creative control" much more. On the other hand, researcher from academia may end up working on problems that they find interesting, or prestige enhancing, but that are of low value to outsiders.

References

Akerlof, G., 1970. The market for "lemons": Quality uncertainty and the market mechanism. The Quarterly Journal of Economics 84 (3), 488–500.

Allen, F., Santomero, A., 1997. The theory of financial intermediation. Journal of Banking & Finance 21 (11-12), 1461–1485.

Becker, G., 1968. Crime and punishment: An economic approach. The Journal of Political Economy 76 (2), 169–217.

Bhatt, S., 1989. Demand uncertainty in a durable goods monopoly. International Journal of Industrial Organization 7 (3), 341–355.

Biehl, A., 2001. Durable-goods monopoly with stochastic values. The RAND Journal of Economics 32 (3), 565–577.

Biglaiser, G., 1993. Middlemen as experts. The RAND Journal of Economics 24 (2), 212–223.

Board, S., 2008. Durable-goods monopoly with varying demand. The Review of Economic Studies 75 (2), 391–413.

Boden, M., 1996. Dimensions of Creativity. MIT Press, Cambridge.

Buchanan, J., Tullock, G., 1975. Polluters' profits and political response: Direct controls versus taxes. The American Economic Review 65 (1), 139–147.

Bucovetsky, S., Chilton, J., 1986. Concurrent renting and selling in a durable-goods monopoly under threat of entry. The RAND Journal of Economics 17 (2), 261–275.

Bulow, J., 1982. Durable-goods monopolists. The Journal of Political Economy 90 (2), 314–332.

Bulow, J., 1986. An economic theory of planned obsolescence. The Quarterly Journal of Economics 101 (4), 729–750.

Campbell, T., Kracaw, W., 1982. Information production, market signalling, and the theory of financial intermediation: A reply. The Journal of Finance 37 (4), 1097–1099.

Chan, Y.-S., 1983. On the positive role of financial intermediation in allocation of venture capital in a market with imperfect information. The Journal of Finance 38 (5), 1543–1568.

Choi, J. P., 2001. Planned obsolescence as a signal of quality. International Economic Journal 15, 59–79.

Clower, R., Dec. 1959. Some theory of an ignorant monopolist. The Economic Journal 69 (276), 705–716.

Coase, R., 1960. The problem of social cost. Journal of Law and Economics 3, 1–44.

Coase, R., 1972. Durability and monopoly. Journal of Law and Economics 15 (1), 143–149.

Cosimano, T., 1996. Intermediation. Economica 63 (249), 131–143.

Cummins, D., Doherty, N., 2006. The economics of insurance intermediaries. Journal of Risk & Insurance 73, 359–396.

Dewatripont, M., Maskin, E., 1995. Credit and efficiency in centralized and decentralized economies. The Review of Economic Studies 62 (4), 541–555.

Diamond, D., 1984. Financial intermediation and delegated monitoring. The Review of Economic Studies 51 (3), 393–414.

Dixit, A., Pindyck, R., 1994. Investment under Uncertainty. Princeton University Press, Princeton.

Dulleck, U., Kerschbamer, R., 2006. On doctors, mechanics, and computer specialists: The economics of credence goods. Journal of Economic Literature 44 (1), 5–42.

Einstein, A., 1934. On the method of theoretical physics. Philosophy of Science 1 (2), 163–169.

Ellison, G., 2002. Evolving standards for academic publishing: A q-r theory. The Journal of Political Economy 110 (5), 994–1034.

Evenson, R., Kislev, Y., 1976. A stochastic model of applied research. The Journal of Political Economy 84 (2), 265–281.

Finkelshtain, I., Kislev, Y., 1997. Prices versus quantities: The political perspective. The Journal of Political Economy 105 (1), 83–100.

Fishman, A., Gandal, N., Shy, O., 1993. Planned obsolescence as an engine of technological progress. The Journal of Industrial Economics 41 (4), 361–370.

Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge.

Glaeser, E., Shleifer, A., 2001. A reason for quantity regulation. The American Economic Review 91 (2), 431–435.

Glaeser, E., Shleifer, A., 2003. The rise of the regulatory state. Journal of Economic Literature 41 (2), 401–425.

Gravelle, H., 1994. Remunerating information providers: Commissions versus fees in life insurance. The Journal of Risk and Insurance 61 (3), 425–457.

Grout, P., Park, I.-U., 2005. Competitive planned obsolescence. The RAND Journal of Economics 36 (3), 596–612.

Gul, F., Sonnenschein, H., Wilson, R., 1986. Foundations of dynamic monopoly and the coase conjecture. Journal of Economic Theory 39 (1), 155–190.

Hart, O., Tirole, J., 1988. Contract renegotiation and coasian dynamics. The Review of Economic Studies 55 (4), 509–540.

Inderst, R., Ottaviani, M., 2009. Misselling through agents. The American Economic Review 99, 883–908.

Kaplow, L., Shavell, S., 2002. On the superiority of corrective taxes to quantity regulation. Am Law Econ Rev 4 (1), 1–17.

Kornai, J., 1979. Resource-Constrained versus Demand-Constrained systems. Econometrica 47 (4), 801–819.

Kornai, J., Maskin, E., Roland, G., 2003. Understanding the soft budget constraint. Journal of Economic Literature 41 (4), 1095–1136.

Lazear, E., 1997. Incentives in basic research. Journal of Labor Economics 15 (1), 167–197.

Leland, H., Pyle, D., 1977. Informational asymmetries, financial structure, and financial intermediation. The Journal of Finance 32 (2), 371–387.

McLennan, A., 1984. Price dispersion and incomplete learning in the long run. Journal of Economic Dynamics and Control 7 (3), 331–347.

Mialon, H., Mialon, S., 2005. Scientific originality and the economics of publishing. Working Paper.

Mumford, M., Gustafson, S., 1988. Creativity syndrome: Integration, application, and innovation. Psychological bulletin 103 (1), 27–43.

Panzar, J., Willig, R., May 1981. Economies of scope. The American Economic Review 71 (2), 268–272.

Pindyck, R., Rubinfeld, D., 2009. Microeconomics, 7th Edition. Upper Saddle River, NJ.

Polo, M., 1991. Hotelling duopoly with uninformed consumers. The Journal of Industrial Economics 39 (6), 701–715.

Posey, L., Tennyson, S., 1998. The coexistence of distribution systems under price search: Theory and some evidence from insurance. Journal of Economic Behavior & Organization 35 (1), 95–115.

Posey, L., Yavas, A., 1995. A search model of marketing systems in property-liability insurance. The Journal of Risk and Insurance 62 (4), 666–689.

Roland, G., 2000. Transition and Economics. Comparative institutional analysis. MIT Press, Cambridge.

Rothschild, M., Stiglitz, J., 1970. Increasing risk: I. a definition. Journal of Economic Theory 2 (3), 225–243.

Rubinstein, A., Wolinsky, A., 1987. Middlemen. The Quarterly Journal of Economics 102 (3), 581–594.

Rust, J., 1986. When is it optimal to kill off the market for used durable goods? Econometrica 54 (1), 65–86.

Salop, S., 1979. Monopolistic competition with outside goods. The Bell Journal of Economics 10 (1), 141–156.

Seog, H., 1999. The coexistence of distribution systems when consumers are not informed. The Geneva Papers on Risk and Insurance - Theory 24 (2), 173–192.

Shavell, S., 1991. Specific versus general enforcement of law. The Journal of Political Economy 99 (5), 1088–1108.

Simon, H., 1969. The Sciences of the Artificial. MIT Press, Cambridge.

Spitzer, E., 2004. Investigation reveals widespread corruption in insurance industry. Press Release: Offce of New York State Attorney General Eliot Spitzer.
URL `www.oag.state.ny.us/`

Sternberg, R., 1999. Handbook of creativity. Cambridge University Press, Cambridge U.K.

Stigler, G., 1955. The nature and role of originality in scientific progress. Economica 22 (88), 293–302.

Stigler, G., 1970. The optimum enforcement of laws. The Journal of Political Economy 78 (3), 526–536.

Stokey, N., 1981. Rational expectations and durable goods pricing. The Bell Journal of Economics 12 (1), 112–128.

Strausz, R., 2009. Planned obsolescence as an incentive device for unobservable quality. The Economic Journal 119 (540), 1405–1421.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. Science 185 (4157), 1124–1131.

Usategui, J., 2007. Non-stationary demand in a durable goods monopoly. The Manchester School 75 (5), 569–579.

Weitzman, M., 1974. Prices vs. quantities. The Review of Economic Studies 41 (4), 477–491.

Williamson, S., 1987. Costly monitoring, loan contracts, and equilibrium credit rationing. The Quarterly Journal of Economics 102 (1), 135–145.

Wilson, J., 1989. Bureaucracy. Basic Books, New York.