

Facilitating Consumer Learning in Insurance Markets—What Are the Welfare Effects?

Johan N. M. Lagerlöf

University of Copenhagen

Christoph Schottmüller*

University of Copenhagen and TILEC

May 4, 2016

Abstract

We model a monopoly insurance market where initially uninformed consumers can privately learn their accident risks at a cost c . We then ask: What are the welfare effects of a policy that reduces the cost c ? We show that if c is sufficiently small ($c < c^*$), the optimal contract is such that the consumer gathers information. For such low values of the cost, both insurer and consumer benefit from a policy that reduces c further. For $c > c^*$, marginally reducing c hurts the insurer and weakly benefits the consumer. Finally, a reduction in c that is “successful,” meaning that the consumer gathers information after the reduction but not before it, can hurt both parties. The reasons for this are that, after the reduction, (i) the cost is actually incurred and (ii) the contracts can be more distorted.

Keywords: adverse selection, consumer policy, genetic testing, information acquisition, insurance

JEL codes: D82, I13

*Address to both authors: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark; email: johan.lagerlof@econ.ku.dk and christoph.schottmueller@econ.ku.dk. We are grateful for helpful comments from the editor Paul Klein, two anonymous referees, and seminar participants at the University of Copenhagen.

1. Introduction

What are the welfare effects of public policies that help insurance customers to learn, privately and covertly, about their accident risks? Although it may be tempting to presume that such policies are unambiguously welfare enhancing, the question is not straightforward, in particular not if insurance companies have market power. While it is true that more accurate information can help consumers choose the most appropriate insurance policy, a reduction in the information gathering cost is also likely to affect the insurer's optimal design of the policy, possibly in a way that hurts consumers and overall welfare. In this paper, we study formally the welfare consequences of an exogenous change in insurance customers' cost of learning, privately and covertly, about their accident risks. We do this by first endogenizing the information structure in Stiglitz's (1977) classic model of a monopoly insurance market with adverse selection, and we then carry out comparative statics exercises with respect to the information gathering cost.

Although our analysis is consistent with many kinds of insurance problems, our preferred interpretation of the model is in terms of health insurance. In recent years, genetic tests that detect certain risk factors have become available to consumers. For example, the companies 23andme and Genome Liberty offer such tests. The tests allow customers to obtain private information concerning their health risks. As an illustration, if the marker rs12425791 equals "A," this increases the risk of having a stroke by a factor of 1.3 (see Ikram et al., 2009). These tests are associated with monetary but also non-monetary costs: Studies have shown that people often feel anxiety in connection with tests that reveal their true health risks to them. (For empirical evidence, see Lerman et al., 1996; Meiser and Dunn, 2000; Levy et al., 2011. For a theoretical approach, see Hoy et al., 2014.) In this context, we are interested in the welfare effects of medical counseling and assistance that are targeted at reducing the test anxiety. In the case of costs that are monetary, we are interested in the welfare consequences of subsidizing the tests.

Another example within the health insurance context, which fits our model, is an individual with private information about her extended family's health history. However, she still needs to learn about how that information translates into a change in her own health risk. If the individual indeed gathers the additional information—e.g., spends time reading books or searching the Internet in order to learn about inheritance patterns of certain diseases—she will obtain access to private information about her own health risk.¹ Examples of public policies that would facilitate

¹This is even true if the additional information is known by the insurance company, since the company is not privy to the health history of the individual's family. Of course, insurers have also additional (private) information on risk which they can use to sort consumers in different risk categories. Our analysis should be thought of being

the gathering of additional information include the funding and launch of a website with relevant information, a telephone line with free expert advice, and an informational campaign.

In our model, there is an insurance company that has a monopoly in its market.² This company and the consumer initially both know only the prior distribution of the latter's accident risk. The company first offers a menu of insurance policies. The consumer then chooses whether or not to, at a cost c , acquire private information about her accident risk. Thereafter she chooses an insurance policy from the menu or decides to remain uninsured. The timing is inspired by health insurance markets where insurers offer several standard insurance contracts to choose from and consumers typically have some time to consider which option to take. Individualized contracts and renegotiation (after information acquisition) are not feasible on these mass markets because the resulting transaction costs would be excessive.

Our analysis shows that if the cost of information acquisition, c , is sufficiently small ($c < c^*$, for a cutoff value $c^* > 0$), the consumer indeed gathers information and thus, at the optimum, is privately informed. In particular, the qualitative features of the optimal menu of insurance policies are the same as in the Stiglitz (1977) model. For such values of the cost c , we can show that the naive intuition is confirmed: A (further) reduction in the cost benefits *both* the insurer and the consumer. In other words, a policy that facilitates consumer learning is a Pareto improvement if, already at the outset, the consumer finds it worthwhile to learn about her risk. In practice, this may be precisely the kind of situation in which such a policy would not be undertaken, since policymakers would believe it is not needed.

If the cost of information gathering is very high ($c > c''$, where $c'' > c^*$), it is common knowledge between the insurer and the consumer that the latter will never have an incentive to gather information and, therefore, the cost does not matter for the design of the optimal insurance policy. More interestingly, however, there are intermediate values of the cost ($c^* < c < c''$) for which the consumer does not incur the cost but it does affect the insurance policy. For this range of cost levels, the insurance company is always worse off from a local reduction in the cost and the consumer is, depending on parameter values, either indifferent to a local cost reduction or better off. That is, for these cost levels, there is much weaker support for the naive intuition discussed

conditional on the insurer's information, i.e. within a given risk category.

²The important feature of the insurance market that we want to capture with this assumption is that the insurer has at least some market power. For empirical evidence suggesting that health insurance markets are not perfectly competitive, see Dafny (2010). To assume a monopoly is an extreme but also very tractable way of incorporating this feature into the model. It would be interesting to study the questions that we look at in an oligopoly setting, but that must await future research.

above: Either there is a conflict of interest concerning the desirability of a cost reduction or such a reduction would lead to a Pareto inferior outcome (with one party being unaffected and the other being strictly worse off). Again, however, in practice this may be precisely the kind of situation in which a policy that facilitates consumer learning is indeed undertaken, since policymakers would believe it is needed.

The comparisons discussed above concern changes in c within the range $0 < c < c^*$ and within the range $c^* < c < c''$. In both cases, we concluded that the consumer is never strictly worse off from a reduction in the cost parameter (although the insurer may be). What about a cost reduction that makes c drop from some level above c^* to a level below it? With the help of numerical examples, we can show that such a cost reduction can make the consumer strictly worse off.³ Indeed, the level of c that maximizes the consumer's expected utility, in these examples, is greater than zero and it is high enough to ensure that information is not acquired. There are two reasons why the consumer's expected utility is lower after the cost reduction: For cost levels below c^* (i) the cost is actually incurred as consumers acquire information and (ii) the contract is sometimes more distorted compared to cost levels above that cutoff value. In our examples, both (i) and (ii) contribute to the expected utility loss that the cost reduction gives rise to. Our analysis thus suggests that a public policy that "successfully" facilitates consumer learning (in the sense that it makes the consumers start to gather information) can be counterproductive in that it lowers consumer surplus.

Our paper is related to three broad strands of literature. The first one is a relatively small but growing literature that endogenizes the information structure in models of adverse selection (or screening). Instead of assuming an asymmetry between the principal and the agent in terms of access to the information itself, this literature takes as its primitive an asymmetry in the *ability to gather* information. The contract that the principal offers must therefore serve the double role of appropriately incentivizing the agent's information gathering and ensuring truthful revelation. Whether or not the principal chooses to induce information gathering, this double role of the contract often means that the contracts look different relative to what they do in the standard models of asymmetric information.

Three early contributions to this literature are Barzel (1977) and Craswell (1988) in the field of law and economics and Demski and Sappington (1987) on delegated expertise. There are also some early papers on auction theory and information acquisition (e.g., Lee 1982). Somewhat more recent contributions include a series of papers by Crémer and Khalil (1992) and Crémer et al.

³The insurer's profits may either increase or decrease, depending on parameter values.

(1998a). All these papers are based on a procurement setting à la Baron and Myerson (1982), where the information that can be gathered concerns the supplier’s production cost. Our insurance application differs from that setting in that, in our framework, the information that the agent can acquire is not relevant for the first best allocation (as full insurance is always desirable). In Crémer and Khalil (1992), the production cost is, regardless of whether information gathering takes place, assumed to become common knowledge after the contract is signed but before production takes place, which makes the setting a bit more similar to ours (since in neither setting information gathering is required for allocative efficiency). Still, information gathering is never induced in that model, which is in contrast to our results. In Crémer et al. (1998a), the production cost does not automatically become known before production, but can only be learned by the agent if incurring the cost before signing the contract.⁴ Apart from the differences in setting, our analysis also puts more emphasis on the comparative statics exercises compared to the papers by Crémer and coauthors (and compared to other papers in the literature).⁵

The second broad strand of related literature, recently surveyed by Dionne et al. (2013), is the one on insurance and adverse selection. A significant part of that literature, which was initiated by Rothschild and Stiglitz (1976) and Wilson (1977), is concerned with competitive markets. However, in an early and classic contribution Stiglitz (1977) modeled a monopolist insurer facing privately informed consumers; for a recent treatment of this problem, see Chade and Schlee (2012).

The paper in the insurance literature that is perhaps the closest to ours is Doherty and Thistle (1996). Their analysis was extended by Hoel et al. (2006) to a setting with heterogenous anxiety costs of acquiring information. Both these papers study the incentives of consumers in an insurance market to gather information about their accident risks. However, in contrast to us, Doherty and Thistle (1996) and Hoel et al. (2006) assume a perfectly competitive insurance market. The equilibrium contracts in such a market are determined by zero profit and interim incentive compatibility constraints that do not depend on the information gathering costs. Hence, comparative statics with respect to the information gathering cost are not particularly interesting. The authors instead focus on the effect of observability and verifiability of information gathering.⁶ Snow

⁴The authors refer to this case as *productive*, as opposed to *strategic*, information gathering.

⁵Crémer and Khalil (1994) and Crémer et al. (1998b) study settings where the agent must gather information before the contract is offered, which make those papers less related to our analysis. The analysis in Crémer et al. (1998a) has recently been extended by Szalay (2009) and Iossa and Martimort (2015).

⁶Hoy and Polborn (2000) investigate similar questions to Doherty and Thistle but assume a setting with life instead of health insurance (still in a competitive market). Ligon and Thistle (1996) model a monopoly insurance problem with both adverse selection and moral hazard. The authors compare a situation where the consumer knows her risk with one where she does not know it. However, Ligon and Thistle do not, as we do, model any information acquisition.

(2015) assumes a monopoly insurer and compares welfare under different information structures. However, he considers neither endogenous nor exogenous and costly information acquisition. In a companion paper (Lagerlöf and Schottmüller, 2015), we investigate a problem that is similar to the one in the present paper, but in an alternative economic environment where the consumer can choose the precision of a noisy signal from a continuum. We discuss that paper and how it relates to the present one in the concluding section (Section 6).

Our paper is also related to the literature on the value of information. The broad point that, in a market environment, more information can hurt is well understood. In some early work, Hirshleifer (1971) pointed out that information can eliminate opportunities for mutually beneficial trade (in an insurance setting, information about the future accident status makes the risk uninsurable).⁷ This so-called Hirshleifer effect plays a role in our result about the welfare effect of lowering the cost c across the threshold c^* . Such a change in c leads to an outcome where the consumer is privately informed, which in turn implies that the optimal contract is more distorted (in particular, involves more underinsurance). However, whether and exactly how the Hirshleifer effect matters for the policy questions that we study is far from straightforward. A contribution of our paper is to clarify this. Moreover, our analysis of the effect of a local cost reduction below c^* and above c^* , respectively, is unrelated to the results in Hirshleifer (1971). That analysis is instead driven by a binding information gathering constraint, which is novel relative to Stiglitz (1977) and to the rest of the insurance literature.

The remainder of the paper is organized as follows. The next section describes our model. Section 3 studies two benchmarks, which will be useful for the analysis of our main model. Section 4.1 solves for the optimal menu of insurance policies in our main model, conditional on the insurer inducing information gathering; Section 4.2 does the same for the case when information gathering is not induced. Section 5 compares those two cases and solves for the overall optimum. That section also presents and discusses our main results. Section 6 concludes. Most of the proofs are relegated to the appendix.

2. The model

We model a monopoly insurance market as a principal-agent problem with adverse selection. Our setup builds on Stiglitz's (1977) classic framework, but we endogenize the information structure.

The principal of the model (P) is a risk neutral and profit maximizing monopoly insurer, and

⁷See Dreze (1960) for an even earlier related contribution and, for example, Marshall (1974) and Schlee (2001) for extensions and variations of Hirshleifer's (1971) analysis.

the agent (A) is a risk averse insuree who has initial wealth $w > 0$. With exogenous probability $\theta \in (0, 1)$, A has an accident, which means that she incurs (the equivalent of) a monetary loss $d \in (0, w)$. A has the opportunity to purchase an insurance policy from P . An insurance policy consists of a premium p and a net indemnity a that is paid in the event of an accident. A 's Bernoulli utility function is twice continuously differentiable and denoted by u (with $u' > 0$ and $u'' < 0$). Hence, provided that A purchases the insurance policy (p, a) , her utility is $u(w - d + a)$ if having an accident and $u(w - p)$ if not having an accident.

The accident probability is either low ($\theta = \underline{\theta}$) or high ($\theta = \bar{\theta}$), where $\underline{\theta} < \bar{\theta}$. We will refer to an agent who knows that $\theta = \underline{\theta}$ as the *low-risk type*. Similarly, an agent with information that $\theta = \bar{\theta}$ will be called the *high-risk type*. At the outset of the game, neither P nor A knows θ . Their common prior belief is that θ is low with probability $v \in (0, 1)$.

The sequence of events is as follows.⁸

- (i) P commits to a menu of insurance policies, $\{(\underline{p}, \underline{a}), (\bar{p}, \bar{a})\}$, where the policy $(\underline{p}, \underline{a})$ is aimed at the low-risk type and the policy (\bar{p}, \bar{a}) is aimed at the high-risk type.
- (ii) A observes the menu and then makes a choice whether or not to gather information, $x \in \{0, 1\}$. If $x = 1$, A must incur a cost $c > 0$ but receives a signal that perfectly reveals the true value of θ . If $x = 0$, A incurs no cost but does not obtain any new information about θ . We assume that the cost c enters A 's payoff as an additive term.⁹ A 's choice of x is not observed by P . Nor can P observe the signal that A receives if $x = 1$.
- (iii) A decides whether to accept any insurance policy in the menu and, if so, which one.

Notice that the assumption in (ii) that P cannot observe A 's choice whether to gather information adds an element of moral hazard to the model, on top of the adverse selection problem that arises if A indeed gathers private information.

⁸The notation introduced in (i) below *allows* for a menu that contains two distinct insurance policies. However, the notation is also consistent with the possibility that P bundles the two policies ($\underline{p} = \bar{p}$ and $\underline{a} = \bar{a}$) or shuts, say, the low-risk type down ($\underline{p} = \underline{a} = 0$). Moreover, if A chooses $x = 0$, then effectively there is only one type, which we can capture by again imposing $\underline{p} = \bar{p}$ and $\underline{a} = \bar{a}$.

⁹This assumption, which is made also by Doherty and Thistle (1996), can be justified in two ways. First, we can think of c as either a disutility of effort or a cost of feeling anxiety about taking the test, and then suppose that the disutility/cost is independent of the wealth level. Second, we can think of the information gathering choice to be made earlier in time relative to the consumption of the wealth, and suppose again that there are no cross effects. In the concluding section, we discuss the effect on the results of instead letting c enter the argument of the utility function u . Our assumption that A , when incurring the cost, learns her risk perfectly is innocuous in our setting. Under non-expected utility, however, the assumption may have non-trivial implications (see Hoy et. al, 2014).

P 's ex ante expected profit from the menu of insurance policies equals

$$\pi = v [(1 - \underline{\theta}) \underline{p} - \underline{\theta} \underline{a}] + (1 - v) [(1 - \bar{\theta}) \bar{p} - \bar{\theta} \bar{a}]. \quad (1)$$

The choice variables that P has available when maximizing profits are the premium and the net indemnity of the two types. When solving the model, however, it will be more convenient to think of P as (equivalently) choosing A 's ex post utility levels directly. Thus, let \bar{u}_A and \bar{u}_N denote A 's utility if she has an accident and if she does not have an accident, respectively, given that she has purchased the policy (\bar{p}, \bar{a}) . Similarly, let \underline{u}_A and \underline{u}_N denote A 's utility if she has an accident and if she does not have an accident, respectively, when having purchased the policy $(\underline{p}, \underline{a})$.¹⁰ Also, let h be the inverse of the function u (hence $h' > 0$ and $h'' > 0$). We can now, equivalently to (1) above, write P 's ex ante expected profit from the menu of insurance policies as follows:

$$\pi = \hat{w} - v [(1 - \underline{\theta}) h(\underline{u}_N) + \underline{\theta} h(\underline{u}_A)] - (1 - v) [(1 - \bar{\theta}) h(\bar{u}_N) + \bar{\theta} h(\bar{u}_A)], \quad (2)$$

where $\hat{w} \equiv w - [v\underline{\theta} + (1 - v)\bar{\theta}]d$ is A 's wealth net of the ex ante expected monetary loss associated with an accident. The net wealth level \hat{w} thus equals, in expected terms, the total amount of resources available in this economy,¹¹ and it therefore also constitutes an upper bound on the amount of profits that P can earn.

We will solve this game by backward induction.

3. Benchmarks with exogenous information

3.1. BENCHMARK 1: THE STIGLITZ MODEL

Our first benchmark is the original Stiglitz (1977) model, in which A is exogenously and privately informed about the accident probability θ . P 's problem is to choose the utility levels \underline{u}_N , \underline{u}_A , \bar{u}_N , and \bar{u}_A so as to maximize the profits in (2), subject to four constraints. First, for each agent type there is one individual rationality (IR) constraint. This ensures that the agent type prefers the insurance policy aimed at her rather than no insurance at all (i.e., $(p, a) = (0, 0)$):

$$(1 - \bar{\theta}) \bar{u}_N + \bar{\theta} \bar{u}_A \geq \bar{U}^*, \quad (\text{IR-high})$$

$$(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A \geq \underline{U}^*, \quad (\text{IR-low})$$

where

$$\bar{U}^* \equiv (1 - \bar{\theta}) u(w) + \bar{\theta} u(w - d) \quad \text{and} \quad \underline{U}^* \equiv (1 - \underline{\theta}) u(w) + \underline{\theta} u(w - d)$$

¹⁰Formally, $\bar{u}_N \equiv u(w - \bar{p})$, $\bar{u}_A \equiv u(w - d + \bar{a})$, $\underline{u}_N \equiv u(w - \underline{p})$, and $\underline{u}_A \equiv u(w - d + \underline{a})$.

¹¹The resources may, however, be further reduced by A 's information gathering cost, c .

are the utility levels associated with the two types' outside option of not purchasing any insurance policy at all (clearly, $\bar{U}^* < \underline{U}^*$). In addition, P 's optimal menu must satisfy two incentive compatibility (IC) constraints. These ensure that each agent type prefers the policy aimed at her to the other policy:

$$(1 - \bar{\theta}) \bar{u}_N + \bar{\theta} \bar{u}_A \geq (1 - \bar{\theta}) \underline{u}_N + \bar{\theta} \underline{u}_A, \quad (\text{IC-high})$$

$$(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A \geq (1 - \underline{\theta}) \bar{u}_N + \underline{\theta} \bar{u}_A. \quad (\text{IC-low})$$

Depending on parameter values, there are two qualitatively different ways in which P 's optimal menu of insurance policies may look like. One possibility is that P prefers to shut down the low-risk type, meaning $(\underline{p}, \underline{a}) = (0, 0)$. If so, P effectively interacts with only one type of agent, with a known accident probability $\bar{\theta}$, and of the four constraints only IR-high is relevant. In order to maximize the profits from this single agent type, P should first provide her with full insurance and then extract all of her rents. Thus, at the optimum, the high-risk type receives the same utility whether she has an accident or not, and this utility level equals her outside option utility, \bar{U}^* . It follows (cf. (2)) that P 's optimized profit in this case equals

$$\pi_{SD}^{SZ} = (1 - v) \left[w - \bar{\theta}d - h(\bar{U}^*) \right] \quad (3)$$

(where the superscript SZ is short for Stiglitz and the subscript SD stands for shutdown).

The other possibility is that P optimally interacts with both types. Here the optimal menu of insurance policies is such that the constraints IR-low and IC-high both bind, whereas IR-high and IC-low are lax.

Proposition 1 (The Stiglitz model). *Consider the Stiglitz model and suppose that P optimally interacts with both agent types. Then, at the optimum, the high-risk type is fully insured ($\bar{u}_A^{SZ} = \bar{u}_N^{SZ} \equiv \bar{u}^{SZ}$) and the low-risk type is underinsured ($\underline{u}_A^{SZ} < \underline{u}_N^{SZ}$). The ex post utility levels at the optimum (\underline{u}_A^{SZ} , \underline{u}_N^{SZ} , and \bar{u}^{SZ}) are implicitly defined by the two binding constraints (IR-low and IC-high) and by the first-order condition*

$$v\underline{\theta}(1 - \underline{\theta}) [h'(\underline{u}_N^{SZ}) - h'(\underline{u}_A^{SZ})] = (1 - v)(\bar{\theta} - \underline{\theta}) h'(\bar{u}^{SZ}). \quad (4)$$

Proof. See Stiglitz (1977) and the supplementary material to the present paper (available on our websites).

Given the optimal utility levels \underline{u}_N^{SZ} , \underline{u}_A^{SZ} , and \bar{u}^{SZ} that are characterized in Proposition 1, P 's ex ante expected profits can be written as (cf. (2))

$$\pi^{SZ} = \hat{w} - v [(1 - \underline{\theta}) h(\underline{u}_N^{SZ}) + \underline{\theta} h(\underline{u}_A^{SZ})] - (1 - v) h(\bar{u}^{SZ}). \quad (5)$$

3.2. BENCHMARK 2: NEITHER PARTY INFORMED

Our second benchmark is identical to the model in Section 2 except that here there is no possibility for A to learn the true state. In this model, there is effectively only one type of agent. The relevant outside option payoff for this single type of agent equals

$$EU^{NI} = v\underline{U}^* + (1 - v)\overline{U}^* \quad (6)$$

(where NI is short for no information). The optimal insurance policy in this benchmark provides full insurance for the single agent type and also leaves her with no rents; thus, $u_N = u_A = EU^{NI}$. It follows (cf. (2)) that P 's maximized ex ante expected profit in this case equals

$$\pi^{NI} = \hat{w} - h(EU^{NI}). \quad (7)$$

It is straightforward to verify (see the supplementary material on our websites) that $\pi^{NI} > \pi_{max}^{SZ} \equiv \max\{\pi_{SD}^{SZ}, \pi^{SZ}\}$, which we will make use of later in the analysis.

4. Inducing and not inducing information gathering

In this section, we will start to solve the model described in Section 2. We first consider the possibility that P wants to induce information gathering (Section 4.1). Thereafter we study the possibility that P does not want to induce information gathering (Section 4.2). In Section 5, we will compare these cases and investigate under what circumstances P optimally induces, and does not induce, information gathering.

4.1. OPTIMAL MENU CONDITIONAL ON INFORMATION GATHERING

Suppose P wants to provide incentives for A to gather information ($x = 1$).¹² P 's problem is then to choose $(\underline{u}_N, \underline{u}_A, \overline{u}_N, \overline{u}_A)$ so as to maximize π as stated in (2), subject to altogether seven constraints. The first four (interim) constraints are identical to the ones in the Stiglitz model, namely IR-high, IR-low, IC-high, and IC-low, as stated in Section 3.1. In addition, three further constraints must be satisfied. These concern A 's incentives at the ex ante stage prior to having learned her type.

¹²As in the Stiglitz model, P may prefer to offer one insurance policy to each type of agent or to interact only with the high-risk type (meaning that the low-risk type remains uninsured). We can study the two cases in a single analytical framework. To obtain the case where the low-risk type is shut down, we simply think of that type's insurance policy as being equal to her outside option (i.e., $\underline{u}_N = u(w)$ and $\underline{u}_A = u(w - d)$). The analysis presented below is thus valid for both cases.

First, A must prefer to gather information about her type to not doing that and then (regardless of her type) choose her outside option:

$$EU_{x=1} \equiv v[(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A] + (1 - v) [(1 - \bar{\theta}) \bar{u}_N + \bar{\theta} \bar{u}_A] - c \geq v \underline{U}^* + (1 - v) \bar{U}^*. \quad (\text{IR-ante})$$

The left-hand side of IR-ante, $EU_{x=1}$, is A 's ex ante expected payoff if gathering information, the last term being the information acquisition cost c (the interim constraints ensure that, after having learned her type, A indeed chooses the policy aimed at that particular type); the right-hand side is A 's ex ante expected outside option utility. The second constraint, which we will refer to as an *information gathering* (IG) constraint, requires that A must prefer to acquire information about her type to not doing so and instead choose the policy aimed at the low-risk type (without gathering information):

$$EU_{x=1} \geq v[(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A] + (1 - v) [(1 - \bar{\theta}) \underline{u}_N + \bar{\theta} \underline{u}_A]. \quad (\text{IG-low})$$

The third constraint says that A must prefer to gather information about her type to not doing so and instead choose the policy aimed at the high-risk type:

$$EU_{x=1} \geq v[(1 - \underline{\theta}) \bar{u}_N + \underline{\theta} \bar{u}_A] + (1 - v) [(1 - \bar{\theta}) \bar{u}_N + \bar{\theta} \bar{u}_A]. \quad (\text{IG-high})$$

There is a strong relationship between the two IG constraints, on the one hand, and the IC constraints, on the other. In particular, from inspecting the constraints, it is clear that IG-low implies IC-high and that IG-high implies IC-low.¹³ This is quite intuitive: If A weakly preferred one contract, say the low coverage contract, no matter which risk type she has, then she should buy the low coverage contract directly without gathering information as this saves the information gathering costs. This behavior would, of course, violate the information gathering constraint. Moreover, as in the standard Stiglitz model, IR-low and IC-high jointly imply IR-high (see Appendix A). Finally, one can show (see again Appendix A) that IR-ante is implied by IR-low and IG-low. In summary we have:

Lemma 1. *The constraints IC-low, IC-high, IR-high, and IR-ante are implied by the other constraints.*

P 's profit maximization problem thus amounts to maximizing π with respect to $(\underline{u}_N, \underline{u}_A, \bar{u}_N, \bar{u}_A)$, subject only to IR-low, IG-low, and IG-high. In the proof of Proposition 2, we show that the first-order conditions of this maximization problem imply that, at the optimum, (i) IG-low and IR-low

¹³Also, in the limit as $c \rightarrow 0$, IG-low and IC-high coincide, as do IG-high and IC-low.

are both binding¹⁴ and that (ii) the low-risk type is underinsured ($\underline{u}_N > \underline{u}_A$). Under the assumption that inducing information gathering ($x = 1$) is actually optimal for P , we also show that IG-high is lax and that the high-risk type is fully insured (i.e. $\bar{u}_N = \bar{u}_A$).

Proposition 2. *Suppose that it is optimal for P to induce information gathering and interact with both types. Then the binding constraints are IR-low and IG-low. The high-risk type is fully insured ($\bar{u}_A^{SB} = \bar{u}_N^{SB} \equiv \bar{u}^{SB}$) and the low-risk type is underinsured ($\underline{u}_A^{SB} < \underline{u}_N^{SB}$). The utility levels at the optimum (\underline{u}_A^{SB} , \underline{u}_N^{SB} , and \bar{u}^{SB}) are implicitly defined by the two binding constraints and by the equality*

$$v\underline{\theta}(1 - \underline{\theta}) [h'(\underline{u}_N^{SB}) - h'(\underline{u}_A^{SB})] = (1 - v)(\bar{\theta} - \underline{\theta}) h'(\bar{u}^{SB}), \quad (8)$$

and ex ante expected profits are given by

$$\pi^{SB} = \hat{w} - v [(1 - \underline{\theta}) h(\underline{u}_N^{SB}) + \underline{\theta} h(\underline{u}_A^{SB})] - (1 - v) h(\bar{u}^{SB}). \quad (9)$$

If it is optimal for the principal to induce information gathering and to interact only with the high-risk type, then this type is fully insured with $u_N^{SD} = u_A^{SD} = u^{SD} = \bar{U}^* + \frac{c}{1-v}$ and ex ante expected profits are

$$\pi_{SD}^{SB} = (1 - v) \left[w - \bar{\theta}d - h \left(\bar{U}^* + \frac{c}{1-v} \right) \right]. \quad (10)$$

Proof. See Appendix A.

(The superscript SB used in the proposition is short for “second best”.)

How does the outcome in Proposition 2 compare with that of the Stiglitz model? The difference is essentially that here IG-low replaces IC-high as one of the binding constraints. Moreover, as noted above, IG-low is more stringent than IC-high. Because of this, the distortion of the low-risk type’s policy is more severe than in the Stiglitz model. Intuitively, the distortion of the low coverage contract allows P to extract more rents. In the Stiglitz model, the distortion makes the low coverage contract less attractive for the high-risk type which allows to increase the premium on the high coverage contract. In our model, the distortion makes the low coverage contract less attractive for A from an ex ante point of view when she has the expected risk $\theta^e \equiv v\underline{\theta} + (1 - v)\bar{\theta}$. The distortion has to be bigger than the one in the Stiglitz model because the ex ante risk is lower than the high risk and therefore A is ex ante not as easily deterred by underinsurance as a high-risk type. The benefit of the distortion is that the high coverage premium can be higher (without inducing consumers to buy the low coverage contract without gathering information).

¹⁴One implication of the fact that IG-low is binding is that A will not get an informational rent if P finds it optimal not to interact with the low-risk type. This follows because IG-low coincides with IR-ante if $\underline{u}_N = u(w)$ and $\underline{u}_A = u(w - d)$.

Note also that the contracts that are optimal in the Stiglitz model cannot induce information gathering—they would violate IG-low: We know (since IC-high binds) that the high type is indifferent between the contracts in the optimal Stiglitz menu. Hence, A can obtain the same gross utility, while saving the costs of information gathering, if she refrains from gathering information and simply buys the contract aimed at the low type directly. This reasoning is valid for all $c > 0$.

Depending on which option is more profitable, P chooses to interact only with the high-risk type or with both types. P 's optimized profits (conditional on inducing information gathering) will thus equal

$$\pi_{x=1}^* \equiv \max \{ \pi^{SB}, \pi_{SD}^{SB} \}. \quad (11)$$

The following lemma, which will be useful when identifying the overall optimum in Section 5, relates P 's profits $\pi_{x=1}^*$ to the profits in the Stiglitz model.

Lemma 2. *Suppose that it is optimal for P to induce information gathering. Then P 's ex ante expected profits are decreasing in c : $\partial \pi_{x=1}^* / \partial c < 0$. Moreover, as the information gathering cost goes to zero, the profit level approaches the one in the Stiglitz model: $\lim_{c \rightarrow 0} \pi_{x=1}^* = \pi_{max}^{SZ}$.*

Proof. See Appendix A.

The result that profits are decreasing in the cost is due to the IG constraints: it is harder to induce information gathering when the cost of information gathering is higher. Hence, a higher cost reduces the set of feasible contract menus that induce information gathering. An alternative way of viewing the problem of higher c is that the low coverage contract has to be distorted more in order to prevent A from buying the low coverage contract without information gathering. Either way, the effect on profits is negative.

4.2. OPTIMAL MENU CONDITIONAL ON NO INFORMATION GATHERING

Now suppose that P does not want to provide incentives for A to gather information ($x = 0$). Note that this situation is different from the one studied in Benchmark 2 in Section 3.2, where A was simply not able to learn about her risk. Here A indeed has the option of gathering information, which means that P must incentivize A not to do so. As a consequence, P will now face additional constraints.

Given that no information is gathered, both types must be offered the same insurance policy, (u_N, u_A) . P 's profit function is thus given by (2) but with $\bar{u}_N = \underline{u}_N = u_N$ and $\bar{u}_A = \underline{u}_A = u_A$. The problem P faces is to maximize this function with respect to u_N and u_A , subject to three constraints. First, without having access to information about her type, A must prefer to purchase

the insurance policy to the alternative of remaining uninsured, which would yield the expected outside option utility:

$$EU_{x=0} \equiv [v(1-\underline{\theta}) + (1-v)(1-\bar{\theta})] u_N + [v\underline{\theta} + (1-v)\bar{\theta}] u_A \geq v\underline{U}^* + (1-v)\bar{U}^*. \quad (\text{IR-ante})$$

Second, A must be willing to purchase the insurance policy without access to information about her type instead of gathering information and then buy the policy only if being the high-risk type:

$$EU_{x=0} \geq v\underline{U}^* + (1-v)[(1-\bar{\theta})u_N + \bar{\theta}u_A] - c. \quad (\text{IG-high})$$

The third constraint requires that A must be willing to purchase the insurance policy without access to information about her type instead of gathering information and then buy the policy only if being the low-risk type. Unsurprisingly, however, this constraint will never bind and we can thus safely disregard it (see Lemma A6 in Appendix B).

To solve P 's problem we will make use of a graphical analysis. To this end, note that the two constraints IR-ante and IG-high can be rewritten as $u_N \geq \varphi_{ante}(u_A)$ and $u_N \geq \varphi_{high}(u_A)$, respectively,¹⁵ where the two φ functions are linear in u_A and satisfy $\varphi'_{ante}(u_A) < \varphi'_{high}(u_A) < 0$. Panels (a) and (b) of Figure 1 depict the two constraints in the (u_A, u_N) -space. The two panels also show the 45-degree line along which A is fully insured. One can verify¹⁶ that the crossing of φ_{ante} and φ_{high} (point C in the figures) is given by the point

$$(u_A, u_N) = (u(w-d) + (1-\theta^e)k, u(w) - \theta^e k), \quad (12)$$

where $\theta^e \equiv v\underline{\theta} + (1-v)\bar{\theta}$ is the ex ante expected risk and $k \equiv c/[v(1-v)(\bar{\theta}-\underline{\theta})]$. With the help of some further algebra, one can confirm that this crossing point is located at or below the 45-degree line if and only if

$$c \geq v(1-v)(\bar{\theta}-\underline{\theta})[u(w) - u(w-d)] \equiv c''.$$

Panel (a) depicts the case where C is located at or below the 45-degree line, and Panel (b) shows the opposite case.

An isoprofit curve has the following slope:

$$\varphi'_{ante}(u_A) \frac{h'(u_A)}{h'(u_N)}. \quad (13)$$

¹⁵We have,

$$\varphi_{ante}(u_A) \equiv \frac{v\underline{U}^* + (1-v)\bar{U}^*}{v(1-\underline{\theta}) + (1-v)(1-\bar{\theta})} - \frac{v\underline{\theta} + (1-v)\bar{\theta}}{v(1-\underline{\theta}) + (1-v)(1-\bar{\theta})} u_A, \quad \varphi_{high}(u_A) \equiv \frac{v\underline{U}^* - c}{v(1-\underline{\theta})} - \frac{\underline{\theta}}{1-\underline{\theta}} u_A.$$

¹⁶See supplementary material (available on our websites).

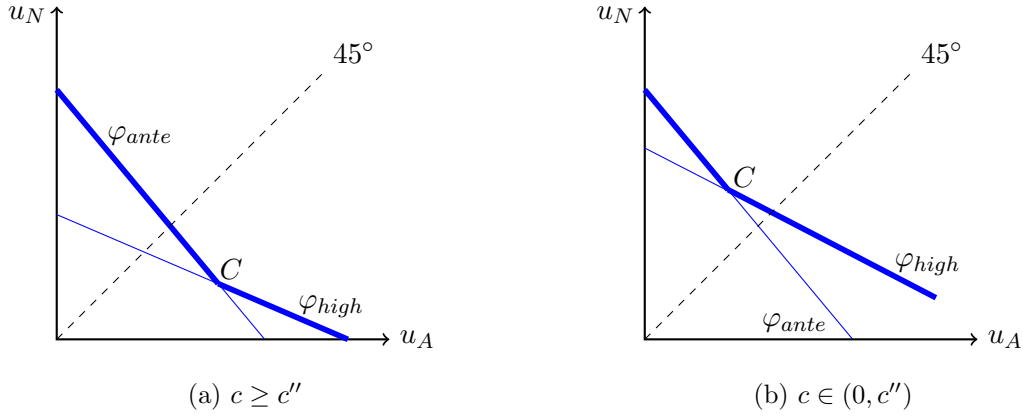


Figure 1: Not inducing information gathering

That is, any isoprofit curve is downward-sloping and, whenever A is fully insured ($u_A = u_N$), its slope equals the slope of IR-ante; above (below) the 45-degree line the slope of the isoprofit curve is flatter (steeper) than the slope of IR-ante. Now consider Panel (a). P 's profits are higher, the closer to the origin A 's utility levels are located. P 's optimal choice of insurance policy must of course satisfy both constraints; that is, it must lie north-east of, or exactly at, the relatively thick segments of the two straight lines in the figure. It is clear that in Panel (a), the optimum must be located on the 45-degree line at a point where the isoprofit curve is tangent to IR-ante and only that constraint is binding. At any other point on the edge of the feasible set, the isoprofit curve's slope would differ from the slope of the binding constraint and P would thus be able to increase the profits by moving closer to full insurance. Intuitively, in Panel (a) the cost c is so high that it is common knowledge between P and A that the latter will never have an incentive to gather information. Distortion is therefore unnecessary for rent extraction and we are effectively back in Benchmark 2, discussed in Section 3.2.

Next consider Panel (b). By using the same argument as above, we can conclude that the optimum cannot be at a point where $u_A \geq u_N$: Then only IG-high would bind and this constraint is flatter than the isoprofit curve. By a similar argument, the optimum cannot be located anywhere strictly north-west of C . The only remaining possibilities are that the optimum is exactly at C or at a point between C and the 45-degree line. Intuitively, offering a contract with partial coverage relaxes IG-high: if A deviates by gathering information and buying the contract only when being a high risk, then she values coverage as a high-risk type. If she sticks to not gathering information she values coverage as an average type with risk θ^e and therefore less than a high-risk type. Consequently, A is hurt more by underinsurance in the deviation than when sticking to no information gathering – the IG-h constraint is therefore relaxed by underinsurance. Note that the deviation in IR-ante is not affected by distorting the contracts and therefore full coverage was

optimal in panel (a).

Let us now investigate under what circumstances the optimum is exactly at C . This will be the case if and only if, at C , the slope of the isoprofit curve is weakly steeper than the slope of IG-high:

$$|\varphi'_{ante}| \frac{h' [u(w-d) + (1-\theta^e)k]}{h' [u(w) - \theta^e k]} \geq |\varphi'_{high}|. \quad (14)$$

At $c = c''$, this inequality is satisfied strictly (because at that value of c , $u_A = u_N$ and thus the slopes of IR-ante and the isoprofit curve are identical). Moreover, it is satisfied for all $c > 0$ if and only if it is satisfied for $c = 0$. Evaluating (14) at $c = 0$ and rearranging, we have

$$\frac{u'(w)}{u'(w-d)} \geq \frac{v\underline{\theta}(1-\underline{\theta}) + (1-v)\underline{\theta}(1-\bar{\theta})}{v\underline{\theta}(1-\underline{\theta}) + (1-v)\bar{\theta}(1-\underline{\theta})}. \quad (15)$$

This inequality may or may not hold—for values of d sufficiently close to zero it does, whereas for d 's sufficiently close to w it does not. We can conclude that if the parameters are such that (15) holds, then the optimum is exactly at C for all $c \in (0, c'')$. Suppose that (15) is violated. Then—since (14) becomes less stringent, the higher c is—there must be a critical value of c , call it c' , such that for $c \in (0, c')$ the optimum is at a point along IG-high strictly between C and the 45-degree line;¹⁷ moreover, we have $0 < c' < c''$.

The following proposition summarizes the results.

Proposition 3. *Assume that P chooses the optimal menu, conditional on A being induced not to gather information.*

- a) *Suppose $c \geq c''$. Then only IR-ante binds and there is full insurance: $u_A^{NI} = u_N^{NI} \equiv u^{NI} = v\underline{U}^* + (1-v)\bar{U}^*$.*
- b) *Suppose that (15) holds and $c < c''$; or that (15) is violated and $c \in [c', c'')$. Then IR-ante and IG-high bind and A is underinsured, with (u_A^{NI}, u_N^{NI}) given by (12).*
- c) *Suppose that (15) is violated and $c \in (0, c')$. Then only IG-high binds and A is underinsured, with (u_A^{NI}, u_N^{NI}) being implicitly defined by the binding constraint and by*

$$\frac{h'(u_A^{NI})}{h'(u_N^{NI})} = \frac{v\underline{\theta}(1-\underline{\theta}) + (1-v)\underline{\theta}(1-\bar{\theta})}{v\underline{\theta}(1-\underline{\theta}) + (1-v)\bar{\theta}(1-\underline{\theta})}. \quad (16)$$

¹⁷The cutoff value c' is implicitly defined by

$$\varphi'_{ante} h' \left[u(w-d) + \frac{(1-\theta^e)c'}{v(1-v)(\bar{\theta}-\underline{\theta})} \right] = \varphi'_{high} h' \left[u(w) - \frac{\theta^e c'}{v(1-v)(\bar{\theta}-\underline{\theta})} \right].$$

Proof. See Appendix B.

We let $\pi_{x=0}^*$ denote P 's profit given the optimum in Proposition 3:

$$\pi_{x=0}^* = \widehat{w} - [v(1 - \underline{\theta}) + (1 - v)(1 - \bar{\theta})] h(u_N^{NI}) - [v\underline{\theta} + (1 - v)\bar{\theta}] h(u_A^{NI}). \quad (17)$$

The results stated in the following lemma will, in the next section, help us characterize the overall optimum.

Lemma 3. *Assume that P chooses the optimal menu, conditional on A being induced not to gather information.*

- a) *Suppose that $c \in (0, c'')$. Then P 's ex ante expected profits are increasing in c : $\partial\pi_{x=0}^*/\partial c > 0$.*
- b) *Suppose that $c > c''$. Then P 's ex ante expected profits are constant with respect to c : $\partial\pi_{x=0}^*/\partial c = 0$.*
- c) *As the information gathering cost goes to zero, the profit level approaches a value strictly below the ex ante expected profits in the Stiglitz model: $\lim_{c \rightarrow 0} \pi_{x=0}^* < \lim_{c \rightarrow 0} \pi_{x=1}^* = \pi_{max}^{SZ}$.*

Proof. See Appendix B.

Parts a) and b) of Lemma 3 say that P 's profits when not inducing information gathering are increasing in c if IG-high is binding; and they are constant with respect to c if IG-high is lax. The reason for this result is similar to the reason for why P 's profits in Section 4.1 are decreasing in c . There, a larger c made the IG constraints more stringent, which had a negative impact on the profits. Here, in contrast, an increase in c *relaxes* the IG constraints (P wants to ensure that A does not incur the cost, which is easier to achieve if the cost is high). Therefore, whenever IG-high is binding, P 's profits benefit from an increase in the information gathering cost.

Part c) of Lemma 3 says that, for small values of c , the profits that P can earn when not inducing information acquisition are lower than those in the Stiglitz model. The reason for this is that if A does not know her type, then P can offer only a pooled insurance policy to A . It is well known that, in the Stiglitz model with two types, offering the same contract to the two types is always dominated by separating the contracts.

5. The overall optimal menu

P will induce information gathering if and only if the profits from doing this are higher than the profits from not inducing information gathering. In order to find the overall optimum to P 's problem, we therefore must compare the profits in (11) with those in (17).

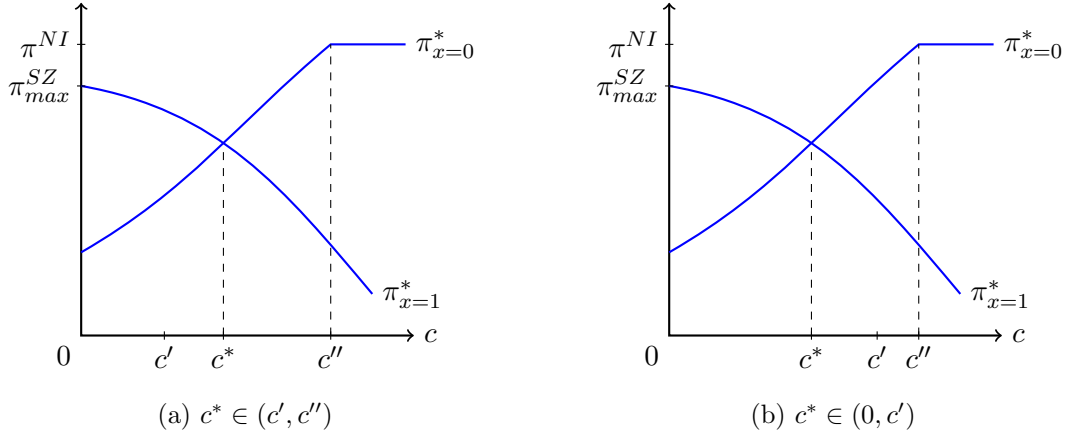


Figure 2: Overall optimum

The comparison is facilitated by Figure 2, showing the $(c, \text{profits})$ -space. First consider Panel (a) of that figure. The downward-sloping curve is the graph of P 's optimized profits, as a function of c , when information acquisition is induced. We know from Lemma 2 that this graph is continuous and downward-sloping; moreover, as c goes to zero, the graph approaches π_{max}^{SZ} , the profits in the Stiglitz model. The other curve in the figure is the graph of P 's optimized profits when information acquisition is not induced. This graph is also continuous and it starts at a point strictly below π_{max}^{SZ} ; it is then initially upward-sloping and after this, for $c \geq c''$, flat. When drawing this graph we could make use of the results in Lemma 3. Notice that, for all $c \geq c''$, P 's profits when information acquisition is not induced equal π^{NI} , the profit level in the benchmark where neither party is informed about A 's type. Moreover, from our analysis in Section 3, we know that $\pi^{NI} > \pi_{max}^{SZ}$.

All in all, it is clear from the figure that there exists a strictly positive cutoff value of c , which we denote by c^* , such that for all $c \in (0, c^*)$ it is optimal for P to induce information gathering and for all $c > c^*$ it is optimal for P not to do that. The cutoff value c^* is implicitly defined by the equality $\pi_{x=0}^* = \pi_{x=1}^*$.

Panel (a) of Figure 2 illustrates the case where $c^* > c'$; that is, here the crossing of the two profit curves occurs at a value of c at which both IG-high and IR-ante are binding when P does not induce information gathering (this case corresponds to part b of Proposition 3). Panel (b) shows the case where $c^* < c'$; that is, here the crossing occurs at a value of c at which only IG-high binds when P does not induce information gathering (part c of Proposition 3).

Proposition 4 below summarizes the results. (When stating the proposition we suppose that, when being indifferent ($c = c^*$), P induces information gathering.)

Proposition 4. *Consider the overall optimum of the model with endogenous information acqui-*

sition.

- a) Suppose that $c \in (0, c^*]$. Then P induces A to gather information ($x = 1$). The menu of insurance policies is as stated in Proposition 2. Moreover, as c goes to zero, the menu of insurance policies approaches the one in the Stiglitz model.
- b) Suppose that $c > c^*$. Then P does not induce A to gather information ($x = 0$). The menu of insurance policies is as stated in Proposition 3.

Many related papers study the “value of information,” which is defined as the utility a player gets when information is acquired minus the utility the player obtains when no information is acquired. As information acquisition is endogenous in our model, this difference could in our context also be called “equilibrium value of information.” Figure 2 shows that the (private) value of information for P is negative (positive) if $c > c^*$ ($c < c^*$). The (private) value of information for A is zero if $c < c''$, because information gathering constraints bind at the optimum. That is, given the optimal contract menu and given $c < c''$, A is indifferent between acquiring and not acquiring information. If $c > c''$, A strictly prefers not to acquire information. From a social perspective, the equilibrium value of information is therefore positive if $c < c^*$ and negative if $c > c^*$, where we use the Pareto criterion to aggregate the private values to a social value.

We will use a superscript asterisk to denote the ex post utility levels at the overall optimum—similarly with P ’s ex ante expected profits (π^*) and A ’s ex ante expected utility (EU^*) at the overall optimum. The following proposition reports a number of comparative statics results concerning these variables.

Proposition 5. *Consider the overall optimum of the model with endogenous information acquisition.*

- a) Suppose $c \in (0, c^*)$. Then the high-risk type is fully insured and the low-risk type is underinsured ($\bar{u}_A^* = \bar{u}_N^* \equiv \bar{u}^*$ and $\underline{u}_A^* < \underline{u}_N^*$). A lower c leads to less underinsurance: $\partial \underline{u}_N^* / \partial c > 0$, $\partial \underline{u}_A^* / \partial c < 0$, and $\partial \bar{u}^* / \partial c > 0$. Moreover, from an ex ante perspective, a lower c (weakly) benefits both P and A : $\partial \pi^* / \partial c < 0$ and $\partial EU^* / \partial c \leq 0$.
- b) Suppose $c \in (c^*, c')$. Then the single type is underinsured ($u_A^* < u_N^*$). Both ex post utility levels are decreasing in c : $\partial u_N^* / \partial c < 0$ and $\partial u_A^* / \partial c < 0$; if $-\frac{u'''}{u''} \geq -2\frac{u''}{u'}$,¹⁸ then a lower c

¹⁸This is a sufficient condition. The condition is equivalent to the requirement that the derivative of risk tolerance is larger than or equal to unity (where risk tolerance is defined as $u'(\cdot)/u''(\cdot)$). With a utility function characterized by constant relative risk aversion, the condition says that the degree of relative risk aversion must not exceed one.

leads to more underinsurance ($\partial u_N^*/\partial c < \partial u_A^*/\partial c$). Moreover, from an ex ante perspective, a lower c hurts P and benefits A : $\partial \pi^*/\partial c > 0$ and $\partial EU^*/\partial c < 0$.

c) Suppose $c > c^*$ and $c \in (c', c'')$. Then the single type is underinsured ($u_A^* < u_N^*$). A lower c leads to more underinsurance: $\partial u_N^*/\partial c < 0$ and $\partial u_A^*/\partial c > 0$. Moreover, from an ex ante perspective, a lower c hurts P and has no effect on A 's utility: $\partial \pi^*/\partial c > 0$ and $\partial EU^*/\partial c = 0$.

d) Suppose $c > c''$. Then the single type is fully insured (and $\partial u^*/\partial c = 0$). Moreover, from an ex ante perspective, a lower c has no effect on P 's profits or on A 's utility: $\partial \pi^*/\partial c = 0$ and $\partial EU^*/\partial c = 0$.

Proof. See Appendix C.

Propositions 4 and 5 tell us that, for sufficiently low values of the information gathering cost, the qualitative features of the optimal menu of insurance policies are the same as in the Stiglitz (1977) model. In particular, there is no distortion at the top and there is full rent extraction at the bottom. Moreover, the values of the ex post utility levels in the limit as c goes to zero approach the corresponding utility levels in the Stiglitz model. We can thus conclude that Stiglitz's classic framework appears to be robust with respect to the endogenization of asymmetric information.

Propositions 4 and 5 also tell us that, for sufficiently large values of c , P will not induce information gathering and, thus, A will not have access to private information about her type. At first glance, this may suggest that here we are effectively in Benchmark 2 (see Section 3.2), where none of the parties is informed and there is full insurance. This is not the case, however, as long as the cost is not too large (namely, as long as $c \in (c^*, c'')$). Although A chooses not to incur the information acquisition cost, the fact that she has the opportunity to do so leads to a distortion of the insurance policy (in particular, to underinsurance). The reason for this is that the IG-high constraint binds and prevents P from providing full insurance. One can think of the situation as one of moral hazard, because the source of the problem is that A can take a non-observable action. To appropriately incentivize (not taking) this action, P chooses to underinsure A .

Proposition 5 further provides us with results concerning the welfare effects of an exogenous reduction in the cost of information gathering. These comparative statics results are interesting and potentially useful, as they can shed light on the desirability of public policies that help consumers in insurance markets to learn about their accident risks. Examples of policy interventions that fit our framework include free or subsidized tests for genetic disorders and the provision of information (through websites, free telephone lines with expert advice, or campaigns).

We should remember that Proposition 5 deals only with cost reductions where c is, both before and after the change, located either below or above c^* (the welfare effects of moving c across the threshold c^* will be discussed shortly). With that in mind, however, we can conclude that Proposition 5 lends mixed support to the naive intuition that a lower cost of information gathering should be welfare enhancing. While the results are consistent with that intuition for sufficiently low costs ($c < c^*$), for somewhat higher costs the results are either ambiguous or they point in the opposite direction. Indeed, it is striking that the only situation described in Proposition 5 for which a cost reduction is beneficial for both parties is *when the cost level at the outset is so low that A already gathers information*. It seems plausible to believe that this is precisely the kind of situation where a policymaker would not undertake the policy, as it would not be thought of as being needed. Similarly, for cost levels that are so high that A does not acquire information (say $c \in (c^*, c'')$), meaning that a policymaker is likely to think of a cost-reducing policy as indeed being required and desirable, the results in Proposition 5 suggest that the policy in fact would *not* be desirable (or at least it would make one of the parties worse off). In particular, if $c > c^*$ and $c \in (c', c'')$ expected utility is constant in c while profits are increasing. Hence, a reduction in c would lead to a Pareto inferior outcome.

These observations lead us to the following policy conclusion: If empirical data suggest that consumers indeed tend to gather information about their health risks, then facilitating gathering such information (further) is welfare enhancing;¹⁹ however, if the data suggest that consumers tend not to gather information, then facilitating learning has mixed welfare effects and may indeed be undesirable (using only the Pareto criterion).

What is the intuition for the welfare results reported in Proposition 5? The results concerning the effect of a change in c on P 's expected profits have already been discussed in Section 4. We concluded there that those results are driven by the fact that an increase in c makes the binding information gathering constraint tighter and looser, respectively, depending on whether $c < c^*$ or $c > c^*$. What about the consumer welfare results? First consider the case $c \in (c^*, c')$. Here, again, it is the binding IG-high constraint that explains why A 's expected utility is decreasing in c : A larger c reduces the “threat” of acquiring information, which relaxes IG-high and thus hurts A . For the case where $c \in (0, c^*)$, expected utility is decreasing in c for other reasons: The simple intuition is that A must incur the cost and therefore a larger c hurts. The more elaborate mechanism is that a larger c tightens the binding IG-low constraint and thereby leads to more

¹⁹Obviously, this is true only if the act of reducing the information acquisition cost is itself not prohibitively costly for the policymaker.

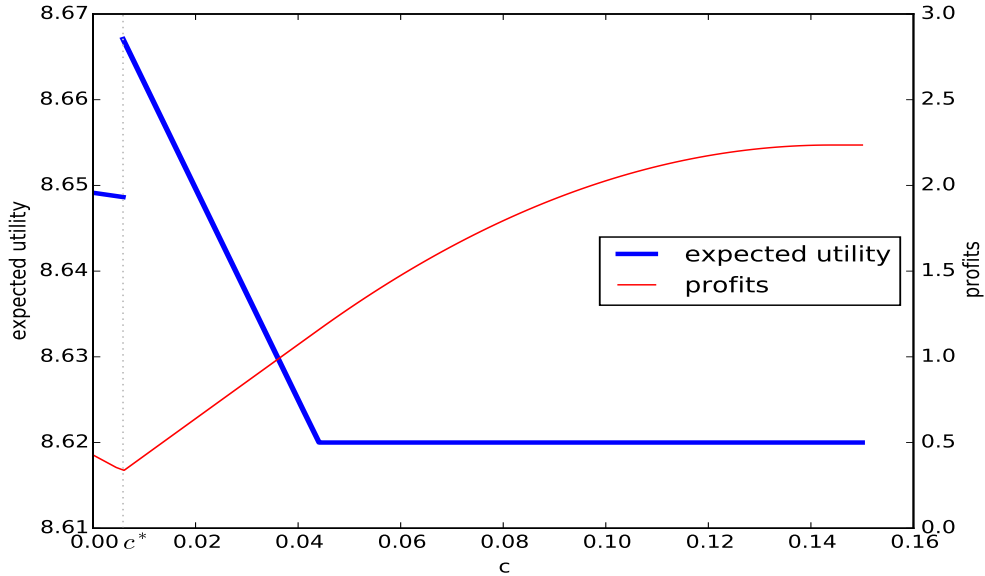


Figure 3: The consumer’s expected utility and the firm’s profits (numerical example)

distortion of the low-coverage contract. Because IR-low is binding, this higher distortion implies that the low-coverage contract becomes less attractive from an ex ante point of view. Because IG-low is binding, this ex ante point of view determines A ’s expected utility.

One way in which we can reconcile the results in Proposition 5 with our naive intuition (which says that lower costs should always be good) is to modify the welfare criterion, assuming a consumer surplus standard: The proposition says that A is never hurt by a local reduction in the information gathering cost.²⁰ However, as pointed out above, Proposition 5 concerns cost reductions only within the range $c < c^*$ and within the range $c > c^*$. If a public policy lowers c from some level above c^* to a level below it, then it turns out that A can indeed be worse off. We now turn to an investigation of cost reductions that move c over the threshold c^* .

We first make the straightforward observation that if the cost is lowered from some cost level satisfying $c > c^*$ and $c \in (c', c'')$ to some cost level $c < c^*$, then the consumer gains, as she here receives a positive rent only after the cost reduction. (Whether the insurer is better or worse off

²⁰Another way in which we can modify our welfare criterion is to disregard the profit and utility levels and instead focus on the extent to which the insurance policies in P ’s menu are distorted away from full coverage. Proposition 5 shows that if the cost parameter is such that A acquires information, then a reduction in c leads to less underinsurance and, therefore, to less distortion. However, if the cost parameter is such that A does not acquire information (but still not so high that the cost does not matter for the optimal insurance policy), a reduction in the cost leads to *more severe* underinsurance and, therefore, to more distortion. (A caveat to this statement is that in the case described under b it is guaranteed to hold only if the condition $-\frac{u'''}{u''} \geq -2\frac{u''}{u'}$ is satisfied.) This result is broadly in line with what we concluded above, namely that a public policy that facilitates consumer learning is welfare enhancing for low cost levels (so low that the consumer at the outset gathers information), but not necessarily otherwise.

depends on the particular parameter values—see Figure 2.) Now consider the more subtle case of a change in c from some $c \in (c^*, c')$ to a $c < c^*$, meaning that the consumer receives a rent also before the cost reduction. Here the analysis is harder, but we have explored the effects of this kind of cost reduction with the help of numerical methods.²¹ In particular, we have verified that there exist parameter configurations for which the consumer is indeed worse off from the cost reduction. This is shown in Figure 3, which graphs the consumer’s expected utility (as well as the firm’s profits) as a function of c . The calculations behind the figure assume a square-root utility function and the following parameter values: $w = 100$, $d = 51$, $\bar{\theta} = 0.7$, $\underline{\theta} = 0.4$, $\nu = 0.8$. Under these assumptions, the three endogenous cutoff values of the model are given by $c^* = 0.00585$, $c' = 0.044058$, and $c'' = 0.144$. As shown in the figure, and consistent with our analytical results, the expected utility is constant in the range $c' < c < c''$. Moreover, in each one of the ranges $0 < c < c^*$ and $c^* < c < c'$ the expected utility is (again consistent with the analytical results) decreasing in c . Interestingly, at the point $c = c^*$ the expected utility makes a jump, and it is clear from the figure that there are many possible values of the cost, before and after the change, that would make a cost reduction bad for the consumer. Indeed, the value of c that yields the highest expected utility for the consumer is strictly positive and sufficiently large to ensure that the insurer does not induce information gathering.

Our numerical analysis thus shows that *a public policy that “successfully” facilitates consumer learning* (in the sense that the consumers gather information after, but not prior to, the cost reduction) *can be counterproductive in that it lowers the consumer surplus*. What is the logic behind this result? There are, for the consumer, two potential advantages with a c that is high enough not to induce information gathering. One of them is that then the cost does not have to be incurred. The other potential advantage is that, loosely speaking, the distortions in the insurance contracts can be smaller when information acquisition is not induced. In our example in Figure 3, both these reasons play a role. The consumer’s expected utility immediately right and left of the cutoff c^* equals 8.66707 and 8.64864, respectively. This means that the vertical length of the jump at c^* (i.e., the consumer’s loss in expected utility when the cost is lowered) equals $8.66707 - 8.64864 = 0.01843$. However, the cost $c^* = 0.00585$ is only about one-third (31.7 percent) of this utility difference and it can therefore not account fully for the consumer’s utility loss. The remaining two-thirds can be attributed to the fact that, overall, the contracts are more severely distorted left of c^* .²² We should note that in this example the insurer interacts with

²¹The code that we have used for these simulations (written in Python) is available on our websites.

²²Of course, the high-risk consumer’s contract is not distorted at all left of c^* , so this is indeed “overall” or on

both consumer types when $c < c^*$. In an optimum in which the low-risk type is shut down, the consumer does not receive *any* rents left of the cutoff c^* . This reinforces the reasoning outlined above and we should thus expect, again, that the consumer’s ideal cost level is strictly positive.

6. Conclusions

This paper has investigated an insurance market in which consumers can learn about their risk—our leading example has been a market for health insurance. In particular, we studied the welfare effects of a public policy that facilitates consumer learning about the own (health) risk. We did this by endogenizing the information structure in the monopoly insurance model due to Stiglitz (1977). The previous literature has not carried out this kind of exercise in an insurance setting and in the presence of market power, nor has it focused on our comparative statics exercises.

We find that the Stiglitz model is robust and that, for a sufficiently small information acquisition cost, the consumer of the model is, endogenously, privately informed about her health risks. We also show that a public policy that facilitates consumer learning is welfare enhancing if the information acquisition cost is low enough to ensure that the consumer indeed incurs it. For higher cost levels, however, a cost-reducing policy can hurt welfare (using only the Pareto criterion) or otherwise it can benefit one party and hurt the other. Interestingly, a cost reduction that is “successful,” in the sense that it ensures that the consumer starts to gather information, can in fact be a failure in that it makes both the consumer and the insurer worse off.

A couple of caveats are in order. First, in our analysis we did not explicitly incorporate a direct cost of facilitating consumer learning (e.g., the costs of subsidizing testing for genetic disorders or providing public websites with information). Any such costs must be added to the indirect costs that we identified in our analysis. Similarly, we have not allowed for any direct benefits of learning one’s health risk, such as the possibility of taking some kind of preventive action (e.g., a woman having a mastectomy if learning that she has a high risk of developing breast cancer). For applications where such possibilities exist, the benefits of facilitating consumer learning would obviously be larger than what our analysis here suggests (still, the costs of facilitating learning that our analysis has identified would be there and they would have to be traded off against the benefits). However, our modeling assumption is the appropriate one for disorders such as Huntington’s disease, where there is no known treatment.

In our analysis, we made some further modeling choices that are not obvious. One of these

average.

is the assumption that the information acquisition cost enters the consumer’s payoff additively. This assumption has been made previously by Doherty and Thistle (1996); we also offer a couple of possible justifications for our assumption in footnote 9. Still, it is natural to wonder how our results would be affected if one relaxed the assumption. In the supplementary material to this paper (available on our websites), we investigate the implications of an alternative setting where the cost c is deducted from A ’s wealth (and thus shows up in the argument of the utility function u). We can show that with CARA utility, which ensures that wealth effects do not affect the degree of risk aversion, most of our qualitative results reported in the present paper go through.²³ When wealth effects affect risk aversion (i.e., if the utility function does not satisfy CARA), the algebra becomes intractable. However, at least for sufficiently small wealth effects, we expect most of our results to be robust with respect to the assumption that c enters as an additive term instead of as a deduction from A ’s wealth.

Another specific assumption that we made was that acquiring information is a binary decision. We do not think of this assumption as merely simplifying. We believe that, for many real world situations, the information gathering decision *is* essentially binary—a natural example of this is a consumer’s decision whether to take a test that can detect a genetic disorder. However, one could also think of situations where consumers decide how much effort to exert when gathering information. More effort implies better, i.e. less noisy, information. In a companion paper (Lagerlöf and Schottmüller, 2015), we study such a model with a continuous effort choice. We show there that the logic of the continuous-effort model is fundamentally different from the logic of the model here (and therefore also many of the results are different). In the continuous setting there are, because of the many effort levels, a larger number of information gathering constraints, several of which are binding at the optimum. In fact, in that model it is in the interest of the principal that the agent is badly informed. This is due to something we call the *missorting effect*, which does not show up when the information gathering decision is binary.

A broad lesson coming out of the present paper is that a public policy that facilitates consumer learning in insurance markets can affect welfare not only directly through the consumer’s cost savings, but also via the market’s optimal response. More surprisingly, our analysis shows that a lower information gathering cost can lead to smaller consumer surplus and lower welfare.

²³The main exception to this concerns the welfare results for the consumer when $c < c^*$. But that is the case where we, in Proposition 5, could report results that are in line with naive intuition. Hence, if the alternative assumption were to alter those results, then the new results would reinforce our point that facilitating consumer learning is not always welfare enhancing.

Appendix A: Inducing information gathering

Lemma A1. *Suppose that IR-low and IC-high are satisfied. Then IR-high is satisfied with a strict inequality.*

Proof. First note that by $\underline{p} \geq 0$,²⁴ IR-low implies the following inequality:

$$(1 - \bar{\theta}) \underline{u}_N + \bar{\theta} \underline{u}_A \geq \bar{U}^*. \quad (18)$$

To see this, rewrite IR-low as follows:

$$\begin{aligned} (1 - \underline{\theta}) u(w - \underline{p}) + \underline{\theta} u(w - d + \underline{a}) &\geq (1 - \underline{\theta}) u(w) + \underline{\theta} u(w - d) \Leftrightarrow \\ \underline{\theta} [u(w) - u(w - \underline{p}) + u(w - d + \underline{a}) - u(w - d)] &\geq u(w) - u(w - \underline{p}). \end{aligned}$$

Given that $\underline{p} \geq 0$, the right-hand side of the last inequality is (weakly) positive; hence, so is the expression in square brackets on the left-hand side. This means that the inequality must still hold if we substitute $\bar{\theta}$ for $\underline{\theta}$:

$$\bar{\theta} [u(w) - u(w - \underline{p}) + u(w - d + \underline{a}) - u(w - d)] \geq u(w) - u(w - \underline{p}),$$

which is equivalent to (18). The claim in the lemma now follows immediately from IC-high and the inequality in (18). \square

Lemma A2. *IR-ante is implied by IR-low and IG-low.*

Proof. Rewrite the left-hand side of IR-ante as follows:

$$\begin{aligned} v [(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A] + (1 - v) [(1 - \bar{\theta}) \bar{u}_N + \bar{\theta} \bar{u}_A] - c \\ \geq v \underline{U}^* + (1 - v) [(1 - \bar{\theta}) \underline{u}_N + \bar{\theta} \underline{u}_A] \geq v \underline{U}^* + (1 - v) \bar{U}^*. \end{aligned}$$

Here the first inequality uses IR-low and IG-low, and the second inequality uses (18). Since the last line equals the right-hand side of IR-ante, the claim follows. \square

Proof of Lemma 1. The claims that IG-low implies IC-high, and that IG-high implies IC-low, follow from inspection. The remaining claims follow from Lemmas A1 and A2. \square

Lemma A3. *IG-low and IR-low are both binding at the optimum.*

²⁴If $\underline{p} < 0$, P would be better off not offering the contract.

Proof. The Lagrangian of the principal's maximization problem can be written as²⁵

$$\begin{aligned}\mathcal{L} &= \widehat{w} - v [(1 - \underline{\theta}) h(\underline{u}_N) + \underline{\theta} h(\underline{u}_A)] - (1 - v) [(1 - \bar{\theta}) h(\bar{u}_N) + \bar{\theta} h(\bar{u}_A)] \\ &\quad + \lambda [(1 - \underline{\theta}) \underline{u}_N + \underline{\theta} \underline{u}_A - \underline{U}^*] - \bar{\mu} \{v [(1 - \underline{\theta}) (\bar{u}_N - \underline{u}_N) + \underline{\theta} (\bar{u}_A - \underline{u}_A)] + c\} \\ &\quad + \underline{\mu} \{(1 - v) [(1 - \bar{\theta}) (\bar{u}_N - \underline{u}_N) + \bar{\theta} (\bar{u}_A - \underline{u}_A)] - c\},\end{aligned}$$

where $\lambda \geq 0$ is the shadow price associated with IR-low, $\bar{\mu} \geq 0$ is the shadow price associated with IG-high, and $\underline{\mu} \geq 0$ is the shadow price associated with IG-low.

First consider the first-order condition with respect to \bar{u}_N :

$$\frac{\partial \mathcal{L}}{\partial \bar{u}_N} = 0 \Leftrightarrow (1 - v) (1 - \bar{\theta}) h'(\bar{u}_N) = \underline{\mu} (1 - v) (1 - \bar{\theta}) - \bar{\mu} v (1 - \underline{\theta}). \quad (19)$$

Given $h' > 0$ and $\bar{\mu} \geq 0$, the first claim follows immediately from (19). The second claim is obvious if P interacts only with the high-risk type (while the low-risk type remains uninsured). Thus suppose P interacts with both types. In this case the first-order condition with respect to \underline{u}_N is

$$\frac{\partial \mathcal{L}}{\partial \underline{u}_N} = 0 \Leftrightarrow v (1 - \underline{\theta}) h'(\underline{u}_N) = \lambda (1 - \underline{\theta}) + \bar{\mu} v (1 - \underline{\theta}) - \underline{\mu} (1 - v) (1 - \bar{\theta}). \quad (20)$$

Now, add (19) and (20):

$$(1 - v) (1 - \bar{\theta}) h'(\bar{u}_N) + v (1 - \underline{\theta}) h'(\underline{u}_N) = \lambda (1 - \underline{\theta}). \quad (21)$$

This equality (using $h' > 0$) implies that $\lambda > 0$. \square

Lemma A4. *The low-risk type is underinsured ($\underline{u}_N > \underline{u}_A$) at the optimum.*

Proof. The claim is obvious if P chooses to interact only with the high-risk type. Hence, let us concentrate on the case where P interacts with both types. The first-order condition with respect to \underline{u}_A is

$$\frac{\partial \mathcal{L}}{\partial \underline{u}_A} = 0 \Leftrightarrow v \underline{\theta} h'(\underline{u}_A) = \lambda \underline{\theta} + \bar{\mu} v \underline{\theta} - \underline{\mu} (1 - v) \bar{\theta}. \quad (22)$$

To prove the claim in the lemma, multiply (20) by $\underline{\theta}$ and multiply (22) by $(1 - \underline{\theta})$:

$$\begin{aligned}v \underline{\theta} (1 - \underline{\theta}) h'(\underline{u}_N) &= \lambda \underline{\theta} (1 - \underline{\theta}) + \bar{\mu} v \underline{\theta} (1 - \underline{\theta}) - \underline{\mu} (1 - v) \underline{\theta} (1 - \bar{\theta}), \\ v \underline{\theta} (1 - \underline{\theta}) h'(\underline{u}_A) &= \lambda \underline{\theta} (1 - \underline{\theta}) + \bar{\mu} v \underline{\theta} (1 - \underline{\theta}) - \underline{\mu} (1 - v) \bar{\theta} (1 - \underline{\theta}).\end{aligned}$$

Next subtract the second one of those equalities from the first one and simplify:

$$v \underline{\theta} (1 - \underline{\theta}) [h'(\underline{u}_N) - h'(\underline{u}_A)] = \underline{\mu} (1 - v) (\bar{\theta} - \underline{\theta}). \quad (23)$$

The claim now follows from the strict convexity of h and the result in Lemma A3 that $\underline{\mu} > 0$. \square

²⁵If P prefers not to interact with the low-risk type, then the Lagrangian is a function of only \bar{u}_N and \bar{u}_A , as the other two utility levels are given by the outside option.

Lemma A5. *Suppose that it is optimal for P to induce information gathering. Then, at the optimum, IG-high is lax (i.e., then $\bar{\mu} = 0$) and the high-risk type is fully insured (i.e. $\bar{u}_N = \bar{u}_A$).*

Proof. We first state the remaining one of the four first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \bar{u}_A} = 0 \Leftrightarrow (1 - v) \bar{\theta} h'(\bar{u}_A) = \underline{\mu} (1 - v) \bar{\theta} - \bar{\mu} v \underline{\theta}, \quad (24)$$

We now obtain the following result from the first-order conditions (19) and (24): *If IG-high is lax (binding), the high-risk type is fully insured (overinsured).* To prove this claim, multiply (19) by $\bar{\theta}$ and multiply (24) by $(1 - \bar{\theta})$:

$$\begin{aligned} (1 - v) \bar{\theta} (1 - \bar{\theta}) h'(\bar{u}_N) &= \underline{\mu} (1 - v) \bar{\theta} (1 - \bar{\theta}) - \bar{\mu} v \bar{\theta} (1 - \underline{\theta}), \\ (1 - v) \bar{\theta} (1 - \bar{\theta}) h'(\bar{u}_A) &= \underline{\mu} (1 - v) \bar{\theta} (1 - \bar{\theta}) - \bar{\mu} v \underline{\theta} (1 - \bar{\theta}). \end{aligned}$$

Subtracting the second one of those equalities from the first one and then simplifying yield

$$(1 - v) \bar{\theta} (1 - \bar{\theta}) [h'(\bar{u}_N) - h'(\bar{u}_A)] = -\bar{\mu} v (\bar{\theta} - \underline{\theta}).$$

The claim now follows from the strict convexity of h .

Now suppose, per contra, that IG-high is binding in the kind of situation described in the lemma. Then, by the result of the previous paragraph, $\bar{u}_A > \bar{u}_N$. We will now show that P can earn higher profits by *not* inducing information gathering ($x = 0$) and offering a pooled full-coverage insurance policy (instead of inducing $x = 1$ and offering the menu $\{(\underline{u}_N, \underline{u}_A), (\bar{u}_N, \bar{u}_A)\}$)—thus contradicting the assumption that it is optimal to induce information gathering. Let the ex ante expected utility level of this pooled full-coverage policy be denoted by u_p , where by construction $u_p = EU_{x=1}$. We can write:

$$\begin{aligned} \pi &= \hat{w} - v [(1 - \underline{\theta}) h(\underline{u}_N) + \underline{\theta} h(\underline{u}_A)] - (1 - v) [(1 - \bar{\theta}) h(\bar{u}_N) + \bar{\theta} h(\bar{u}_A)] \\ &< \hat{w} - h[v(1 - \underline{\theta}) \underline{u}_N + v \underline{\theta} \underline{u}_A + (1 - v)(1 - \bar{\theta}) \bar{u}_N + (1 - v) \bar{\theta} \bar{u}_A] \\ &= \hat{w} - h(u_p + c) < \hat{w} - h(u_p) = \pi_p, \end{aligned}$$

where the first inequality follows from the strict convexity of h and the last one from $c > 0$ and $h' > 0$. It remains to check that A indeed has an incentive to refrain from gathering information and to purchase the pooled full-coverage policy, thus receiving the payoff u_p . To this end, first note that (by construction) A 's ex ante expected utility remains unchanged. Hence, IR-ante is satisfied. Second, we must check that it is not optimal for A to gather information and then purchase the insurance policy only if learning that $\theta = \bar{\theta}$ (and remain uninsured if $\theta = \underline{\theta}$). Because IG-high is

binding, we know that $u_p = [v\underline{\theta} + (1-v)\bar{\theta}] \bar{u}_A + [v(1-\underline{\theta}) + (1-v)(1-\bar{\theta})] \bar{u}_N$. Since $\bar{u}_A > \bar{u}_N$ and $\bar{\theta} > \underline{\theta}$, this implies $u_p < \bar{\theta}\bar{u}_A + (1-\bar{\theta})\bar{u}_N$. Therefore,

$$\begin{aligned} u_p &= v[(1-\underline{\theta})\underline{u}_N + \underline{\theta}\underline{u}_A] + (1-v)[(1-\bar{\theta})\bar{u}_N + \bar{\theta}\bar{u}_A] - c \\ &> v[(1-\underline{\theta})\underline{u}_N + \underline{\theta}\underline{u}_A] + (1-v)u_p - c \\ &\geq v\underline{U}^* + (1-v)u_p - c, \end{aligned}$$

where the last step uses IR-low. But this inequality implies that buying the pooling insurance policy without gathering information leads to a higher utility than gathering information and buying insurance only if learning that the accident risk is high. \square

Proof of Proposition 2. For the case where P interacts with both types, all the claims except for equation (8) follow from Lemmas 1, A3, A4, A5, and from the arguments in the text. To derive equation (8), add (22) and (24), using $\bar{\mu} = 0$:

$$(1-v)\bar{\theta}h'(\bar{u}_A) + v\underline{\theta}h'(\underline{u}_A) = \lambda\underline{\theta}. \quad (25)$$

Notice that the right-hand side of (21) multiplied by $\underline{\theta}$ equals the right-hand side of (25) multiplied by $(1-\underline{\theta})$. We thus have:

$$\underline{\theta}[(1-v)(1-\bar{\theta})h'(\bar{u}_N) + v(1-\underline{\theta})h'(\underline{u}_N)] = (1-\underline{\theta})[(1-v)\bar{\theta}h'(\bar{u}_A) + v\underline{\theta}h'(\underline{u}_A)],$$

which rewritten (using $\bar{u}_A = \bar{u}_N = \bar{u}$) yields equation (8).

For the case where P interacts only with the high-risk type, note that the binding IG-low constraint becomes IR-ante. Using the fact that the offered contract has full coverage, IR-ante implies

$$u^{SD} = \bar{U}^* + \frac{c}{1-v}. \quad (26)$$

\square

Proof of Lemma 2. For the case where P interacts only with the high-risk type, the claim about $\partial\pi^{SB}/\partial c$ is obvious from (10). For the other case, the (general) envelope theorem,²⁶ implies $\partial\pi^{SB}/\partial c = -\underline{\mu}^{SB} < 0$ (for we know from Lemma A3 that $\underline{\mu}^{SB} > 0$). Next turn to the proof of the limit result. By Proposition 2, \underline{u}_A^{SB} , \underline{u}_N^{SB} , and \bar{u}^{SB} are defined by the binding IR-low and IG-low and by (8). By Proposition 1, the utility levels at the optimum of the Stiglitz model (\underline{u}_A^{SZ} , \underline{u}_N^{SZ} , and \bar{u}^{SZ}) are defined by the binding IR-low and IC-high and by (4). But (4) and (8) are identical, and in the limit where $c \rightarrow 0$, IG-low and IC-high also are identical. \square

²⁶See, for example, Sydsæter and Hammond (1995, p. 678 ff.).

Appendix B: Not inducing information gathering

Lemma A6. *The third constraint discussed in Section 4.2 (IG-low) is implied by the constraint IR-ante stated in the same subsection.*

Proof. To see this, write this third constraint as

$$EU_{x=0} \geq v [(1 - \underline{\theta}) u_N + \underline{\theta} u_A] + (1 - v) \bar{U}^* - c, \quad (\text{IG-low})$$

and note that this constraint is implied by IR-ante if $(1 - \underline{\theta})u_N + \underline{\theta}u_A \leq \underline{U}^*$. If $(1 - \underline{\theta})u_N + \underline{\theta}u_A > \underline{U}^*$, then $(1 - \bar{\theta})u_N + \bar{\theta}u_A \geq \bar{U}^*$ must hold as $u(w) - u(w - d) \geq u_N - u_A$ by $p \geq 0$ (cf. the proof of Lemma A1). This implies

$$\begin{aligned} v [(1 - \underline{\theta}) u_N + \underline{\theta} u_A] + (1 - v) \bar{U}^* - c \\ < v [(1 - \underline{\theta}) u_N + \underline{\theta} u_A] + (1 - v) ((1 - \bar{\theta}) u_N + \bar{\theta} u_A) = EU_{x=0}, \end{aligned}$$

which again implies IG-low. □

Proof of Proposition 3. Almost all of the claims in the proposition are proven in the preceding text. The only thing that remains is to verify that the optimal insurance policy in part c) of the proposition must satisfy equation (16). But this equality is simply the condition that, at the optimum (u_A^{NI}, u_N^{NI}) , the slope of the iso-profit curve equals the slope of IG-high. By using (13) and the equations in footnote 15, we obtain

$$\varphi'_{ante}(u_A^{NI}) \frac{h'(u_A^{NI})}{h'(u_N^{NI})} = \varphi'_{high}(u_A^{NI}) \Leftrightarrow \frac{v\underline{\theta} + (1 - v)\bar{\theta}}{v(1 - \underline{\theta}) + (1 - v)(1 - \bar{\theta})} \frac{h'(u_A)}{h'(u_N)} = \frac{\underline{\theta}}{1 - \underline{\theta}},$$

which simplifies to the equation in the proposition. □

Proof of Lemma 3. The Lagrangian of P 's problem in Section 4.2 can be written as

$$\begin{aligned} \mathcal{L}_{x=0} = & \hat{w} - v [(1 - \underline{\theta}) h(u_N) + \underline{\theta} h(u_A)] - (1 - v) [(1 - \bar{\theta}) h(u_N) + \bar{\theta} h(u_A)] \\ & + \lambda \left[EU_{x=0} - v \underline{U}^* - (1 - v) \bar{U}^* \right] + \bar{\mu} \left\{ EU_{x=0} - v \underline{U}^* - (1 - v) [(1 - \bar{\theta}) u_N + \bar{\theta} u_A] + c \right\} \\ & + \underline{\mu} \left\{ EU_{x=0} - v [(1 - \underline{\theta}) u_N + \underline{\theta} u_A] - (1 - v) \bar{U}^* + c \right\}, \end{aligned}$$

where $\lambda \geq 0$ is the shadow price associated with IR-ante, $\bar{\mu} \geq 0$ is the shadow price associated with IG-high, and $\underline{\mu} \geq 0$ is the shadow price associated with IG-low. By the (general) envelope theorem, we have

$$\frac{\partial \pi_{x=0}^*}{\partial c} = \frac{\partial \mathcal{L}_{x=0}}{\partial c} = \bar{\mu} + \underline{\mu}.$$

We know from Proposition 3 that, for $c < c''$, IG-high is always binding ($\bar{\mu} > 0$) and IG-low is not ($\underline{\mu} = 0$), which means that then $\partial \pi_{x=0}^* / \partial c > 0$. We also know that for $c > c''$, both IG-high and

IG-low are lax ($\bar{\mu} = \underline{\mu} = 0$), which means that then $\partial\pi_{x=0}^*/\partial c = 0$. It now only remains to prove part c) of the proposition. But this is to a large extent done in the text after the proposition. The standard result referred to there, that a pooled contract is always dominated by a separating one, is proven, for example, in Stiglitz (1977). We too implicitly showed the result when we solved Benchmark 1 (i.e., the Stiglitz model) and found that P never at the optimum offers the same contract to the two types, although the formulation of P 's problem allows for this possibility. \square

Appendix C: The overall optimum

Proof of Proposition 5, part a. First consider part a) of the proposition and the comparative statics results concerning the ex post utility levels. We know that these utility levels are defined by the two binding constraints (IG-low and IR-low) and by equation (8). Multiplying IR-low by $\bar{\theta}$ and IG-low by $\underline{\theta}$, and then combining, yield:

$$(\bar{\theta} - \underline{\theta}) \underline{u}_N^* = \underline{U}^* \bar{\theta} - \left(\bar{u}^* - \frac{c}{1-v} \right) \underline{\theta}.$$

Multiplying IR-low by $1 - \bar{\theta}$ and IG-low by $1 - \underline{\theta}$, and then combining, yield:

$$(\bar{\theta} - \underline{\theta}) \underline{u}_A^* = (1 - \underline{\theta}) \left(\bar{u}^* - \frac{c}{1-v} \right) - (1 - \bar{\theta}) \underline{U}^*.$$

Equation (8) is:

$$v \underline{\theta} (1 - \underline{\theta}) [h'(\underline{u}_N^*) - h'(\underline{u}_A^*)] = (1 - v) (\bar{\theta} - \underline{\theta}) h'(\bar{u}^*).$$

Now differentiate these three equations with respect to c and then write the resulting expressions on matrix form:

$$M \begin{bmatrix} \frac{\partial \underline{u}_N^*}{\partial c} \\ \frac{\partial \underline{u}_A^*}{\partial c} \\ \frac{\partial \bar{u}^*}{\partial c} \end{bmatrix} = \begin{bmatrix} \frac{\underline{\theta}}{1-v} \\ -\frac{1-\underline{\theta}}{1-v} \\ 0 \end{bmatrix},$$

where

$$M \equiv \begin{bmatrix} \bar{\theta} - \underline{\theta} & 0 & \underline{\theta} \\ 0 & \bar{\theta} - \underline{\theta} & -(1 - \underline{\theta}) \\ v \underline{\theta} (1 - \underline{\theta}) h''(\underline{u}_N^*) & -v \underline{\theta} (1 - \underline{\theta}) h''(\underline{u}_A^*) & -(1 - v) (\bar{\theta} - \underline{\theta}) h''(\bar{u}^*) \end{bmatrix}.$$

We denote the determinant of M by D , which is straightforward to calculate as:

$$D = -(\bar{\theta} - \underline{\theta}) \left[(1 - v) (\bar{\theta} - \underline{\theta})^2 h''(\bar{u}^*) + v \underline{\theta} (1 - \underline{\theta})^2 h''(\underline{u}_A^*) + v \underline{\theta}^2 (1 - \underline{\theta}) h''(\underline{u}_N^*) \right] < 0.$$

Cramer's rule thus gives us

$$\begin{aligned} D \frac{\partial \underline{u}_N^*}{\partial c} &= -\frac{\underline{\theta}}{1-v} \left[(\bar{\theta} - \underline{\theta})^2 (1 - v) h''(\bar{u}^*) + v \underline{\theta} (1 - \underline{\theta})^2 h''(\underline{u}_A^*) \right] + \frac{1 - \underline{\theta}}{1 - v} v \underline{\theta}^2 (1 - \underline{\theta}) h''(\underline{u}_N^*) \\ &= -\frac{1}{1-v} \left[(1 - v) \underline{\theta} (\bar{\theta} - \underline{\theta})^2 h''(\bar{u}^*) \right] \Leftrightarrow \end{aligned}$$

$$\frac{\partial \underline{u}_N^*}{\partial c} = \frac{\underline{\theta} (\bar{\theta} - \underline{\theta}) h''(\bar{u}^*)}{(1-v)(\bar{\theta} - \underline{\theta})^2 h''(\bar{u}^*) + v\underline{\theta}(1-\underline{\theta})^2 h''(\underline{u}_A^*) + v\underline{\theta}^2(1-\underline{\theta}) h''(\underline{u}_N^*)} > 0.$$

And

$$\begin{aligned} D \frac{\partial \underline{u}_A^*}{\partial c} &= -v\underline{\theta}(1-\underline{\theta}) h''(\underline{u}_N^*) \left[\frac{(1-\underline{\theta})\underline{\theta}}{1-v} - \frac{\underline{\theta}(1-\underline{\theta})}{1-v} \right] - (1-v)(\bar{\theta} - \underline{\theta}) h''(\bar{u}^*) \times \\ &\quad \times \left[-\frac{(\bar{\theta} - \underline{\theta})(1-\underline{\theta})}{1-v} \right] = (\bar{\theta} - \underline{\theta})^2 (1-\underline{\theta}) h''(\bar{u}^*) \Leftrightarrow \end{aligned}$$

$$\frac{\partial \underline{u}_A^*}{\partial c} = -\frac{(\bar{\theta} - \underline{\theta})(1-\underline{\theta}) h''(\bar{u}^*)}{(1-v)(\bar{\theta} - \underline{\theta})^2 h''(\bar{u}^*) + v\underline{\theta}(1-\underline{\theta})^2 h''(\underline{u}_A^*) + v\underline{\theta}^2(1-\underline{\theta}) h''(\underline{u}_N^*)} < 0.$$

And

$$\begin{aligned} D \frac{\partial \bar{u}^*}{\partial c} &= v\underline{\theta}(1-\underline{\theta}) h''(\underline{u}_N^*) \left[-(\bar{\theta} - \underline{\theta}) \frac{\underline{\theta}}{1-v} \right] + v\underline{\theta}(1-\underline{\theta}) h''(\underline{u}_A^*) \left[-(\bar{\theta} - \underline{\theta}) \frac{1-\underline{\theta}}{1-v} \right] \\ &= -\frac{v\underline{\theta}(1-\underline{\theta})(\bar{\theta} - \underline{\theta})}{1-v} [\underline{\theta} h''(\underline{u}_N^*) + (1-\underline{\theta}) h''(\underline{u}_A^*)] \Leftrightarrow \\ \frac{\partial \bar{u}^*}{\partial c} &= \frac{v\underline{\theta}(1-\underline{\theta}) [\underline{\theta} h''(\underline{u}_N^*) + (1-\underline{\theta}) h''(\underline{u}_A^*)]}{(1-v) \left[(1-v)(\bar{\theta} - \underline{\theta})^2 h''(\bar{u}^*) + v\underline{\theta}(1-\underline{\theta})^2 h''(\underline{u}_A^*) + v\underline{\theta}^2(1-\underline{\theta}) h''(\underline{u}_N^*) \right]} > 0. \end{aligned}$$

The claim about P 's ex ante expected profit follows from Lemma 2 and Proposition 4.

Let us finally prove the claim about A 's ex ante expected utility. First, if P optimally interacts with only the high-risk type, then A does not get any rent and, thus, her ex ante expected utility is independent of c . Next suppose P optimally interacts with both types. We then know that IR-low binds: $(1-\underline{\theta})\underline{u}_N^* + \underline{\theta}\underline{u}_A^* = \underline{U}^*$. Differentiating this identity yields

$$(1-\underline{\theta}) \frac{\partial \underline{u}_N^*}{\partial c} + \underline{\theta} \frac{\partial \underline{u}_A^*}{\partial c} = 0 \Leftrightarrow \frac{\partial \underline{u}_N^*}{\partial c} = -\frac{\underline{\theta}}{1-\underline{\theta}} \frac{\partial \underline{u}_A^*}{\partial c}. \quad (27)$$

A 's ex ante expected utility can be written as

$$EU^* = v[(1-\underline{\theta})\underline{u}_N^* + \underline{\theta}\underline{u}_A^*] + (1-v)\bar{u}^* - c = v\underline{U}^* + (1-v)[(1-\bar{\theta})\underline{u}_N^* + \bar{\theta}\underline{u}_A^*], \quad (28)$$

where the second line makes use of the binding IR-low (see above) and the binding IG-low $((1-v)\bar{u}^* - c = (1-v)[(1-\bar{\theta})\underline{u}_N^* + \bar{\theta}\underline{u}_A^*])$. Now differentiate the expression for A 's ex ante expected utility in (28) with respect to c :

$$\begin{aligned} \frac{\partial EU^*}{\partial c} &= (1-v) \left[(1-\bar{\theta}) \frac{\partial \underline{u}_N^*}{\partial c} + \bar{\theta} \frac{\partial \underline{u}_A^*}{\partial c} \right] = (1-v) \left[(1-\bar{\theta}) \left(-\frac{\underline{\theta}}{1-\underline{\theta}} \frac{\partial \underline{u}_A^*}{\partial c} \right) + \bar{\theta} \frac{\partial \underline{u}_A^*}{\partial c} \right] \\ &= (1-v) \frac{\partial \underline{u}_A^*}{\partial c} \left[\frac{\bar{\theta} - \underline{\theta}}{1-\underline{\theta}} \right], \end{aligned}$$

where the second equality makes use of (27). The last expression is strictly negative as $\partial \underline{u}_A^* / \partial c < 0$. \square

Proof of Proposition 5, part b. Here the ex post utility levels are defined by the binding IG-high and equality (16):

$$v(1-\underline{\theta})u_N^* + v\underline{\theta}u_A^* = vU^* - c, \quad (29)$$

$$\Phi h'(u_A^*) = h'(u_N^*), \quad \text{where} \quad \Phi \equiv \frac{v\underline{\theta}(1-\underline{\theta}) + (1-v)\bar{\theta}(1-\underline{\theta})}{v\underline{\theta}(1-\underline{\theta}) + (1-v)\underline{\theta}(1-\bar{\theta})} > 1.$$

Differentiate the two identities with respect to c :

$$v(1-\underline{\theta})\frac{\partial u_N^*}{\partial c} + v\underline{\theta}\frac{\partial u_A^*}{\partial c} = -1, \quad \Phi h''(u_A^*)\frac{\partial u_A^*}{\partial c} = h''(u_N^*)\frac{\partial u_N^*}{\partial c}$$

Write these two equations on matrix form:

$$\begin{bmatrix} v(1-\underline{\theta}) & v\underline{\theta} \\ h''(u_N^*) & -\Phi h''(u_A^*) \end{bmatrix} \begin{bmatrix} \frac{\partial u_N^*}{\partial c} \\ \frac{\partial u_A^*}{\partial c} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Using Cramer's rule we have

$$\frac{\partial u_N^*}{\partial c} = -\frac{\Phi h''(u_A^*)}{v(1-\underline{\theta})\Phi h''(u_A^*) + v\underline{\theta}h''(u_N^*)} < 0, \quad (30)$$

$$\frac{\partial u_A^*}{\partial c} = -\frac{h''(u_N^*)}{v(1-\underline{\theta})\Phi h''(u_A^*) + v\underline{\theta}h''(u_N^*)} < 0. \quad (31)$$

Thus, a lower c leads to more underinsurance ($\partial u_N^*/\partial c < \partial u_A^*/\partial c$) if and only if

$$\begin{aligned} \frac{\partial u_N^*}{\partial c} - \frac{\partial u_A^*}{\partial c} &= \frac{-\Phi h''(u_A^*) + h''(u_N^*)}{v(1-\underline{\theta})\Phi h''(u_A^*) + v\underline{\theta}h''(u_N^*)} < 0 \Leftrightarrow \Phi > \frac{h''(u_N^*)}{h''(u_A^*)} \Leftrightarrow \\ &\frac{h'(u_N^*)}{h'(u_A^*)} > \frac{h''(u_N^*)}{h''(u_A^*)} \Leftrightarrow 1 > \frac{h''(u_N^*)/h'(u_N^*)}{h''(u_A^*)/h'(u_A^*)}. \end{aligned} \quad (32)$$

We know that $u_N^* > u_A^*$. Therefore, a sufficient condition for (32) to hold is that $h''(z)/h'(z)$ is weakly decreasing in z . We have $\frac{\partial(h''(z)/h'(z))}{\partial z} \leq 0 \Leftrightarrow h'''(z)h'(z) \leq [h''(z)]^2$. It is straightforward (but a bit cumbersome) to verify that $h' = 1/u'$,

$$h'' = -\frac{u''}{(u')^3}, \quad \text{and} \quad h''' = \left(\frac{3u''}{u'} - \frac{u'''}{u''}\right) \frac{u''}{(u')^4}.$$

Therefore, a sufficient condition for (32) to hold is that

$$h'''(z)h'(z) \leq [h''(z)]^2 \Leftrightarrow \left(\frac{3u''}{u'} - \frac{u'''}{u''}\right) \frac{u''}{(u')^5} \leq \frac{(u'')^2}{(u')^6} \Leftrightarrow \frac{3u''}{u'} - \frac{u'''}{u''} \geq \frac{u''}{u'} \Leftrightarrow -\frac{u'''}{u''} \geq -2\frac{u''}{u'}.$$

Finally, the claim about A 's ex ante expected utility follows immediately from the result above in (30) and (31), and the claim about P 's ex ante expected profits follows from Lemma 3a and Proposition 4. \square

Proof of Proposition 5, part c. Here the ex post utility levels are defined by the binding IG-high in equation (29) and the binding IR-ante, or by equation (12). By inspection, we have

$\partial u_N^*/\partial c < 0$ and $\partial u_A^*/\partial c > 0$. The claim about P 's ex ante expected profits follows from Lemma 3a and Proposition 4. Finally, the claim about A 's ex ante expected utility follows from the fact that IR-ante is binding. \square

Proof of Proposition 5, part d. Here the single ex post utility level, u^* , is defined by the binding IR-ante and involves full insurance. Therefore it is independent of c and we thus have $\partial u^*/\partial c = 0$. The claim about P 's ex ante expected profits follows from Lemma 3b and Proposition 4. Finally, the claim about A 's ex ante expected utility follows from the fact that IR-ante is binding. \square

References

- Baron, D. P. and R. B. Myerson (1982). Regulating a Monopolist with Unknown Costs. *Econometrica* 50(4), 911–930.
- Barzel, Y. (1977). Some Fallacies in the Interpretation of Information Costs. *Journal of Law and Economics* 20(2), 291–307.
- Chade, H. and E. Schlee (2012). Optimal Insurance with Adverse Selection. *Theoretical Economics* 7(3), 571–607.
- Craswell, R. (1988). Precontractual Investigation as an Optimal Precaution Problem. *The Journal of Legal Studies* 17(2), 401–436.
- Crémer, J. and F. Khalil (1992). Gathering Information Before Signing a Contract. *The American Economic Review* 82(3), 566–578.
- Crémer, J. and F. Khalil (1994). Gathering Information Before the Contract is Offered: The Case with Two States of Nature. *European Economic Review* 38(3–4), 675–682.
- Crémer, J., F. Khalil, and J.-C. Rochet (1998a). Contracts and Productive Information Gathering. *Games and Economic Behavior* 25(2), 174–193.
- Crémer, J., F. Khalil, and J.-C. Rochet (1998b). Strategic Information Gathering Before a Contract is Offered. *Journal of Economic Theory* 81(1), 163–200.
- Dafny, L. S. (2010). Are Health Insurance Markets Competitive? *American Economic Review* 100(4), 1399–1431.

- Demski, J. S. and D. E. M. Sappington (1987). Delegated Expertise. *Journal of Accounting Research* 25(1), pp. 68–89.
- Dionne, G., N. Fombaron, and N. Doherty (2013). Adverse Selection in Insurance Contracting. In G. Dionne (Ed.), *Handbook of Insurance* (Second ed.), Chapter 10, pp. 231–280. Springer.
- Doherty, N. A. and P. D. Thistle (1996). Adverse Selection with Endogenous Information in Insurance Markets. *Journal of Public Economics* 63(1), 83–102.
- Drèze, J. H. (1960). Le paradoxe de l'information. *Economie appliquée* 13, 71–80. English translation in: J. H. Drèze, *Essays on Economic Decisions under Uncertainty*, Chapter 4, 1987, Cambridge University Press.
- Hirshleifer, J. (1971). The Private and Social Value of Information and the Reward to Inventive Activity. *The American Economic Review* 61(4), 561–574.
- Hoel, M., T. Iversen, T. Nilssen, and J. Vislie (2006). Genetic Testing in Competitive Insurance Markets with Repulsion from Chance: A Welfare Analysis. *Journal of Health Economics* 25(5), 847–860.
- Hoy, M., R. Peter, and A. Richter (2014). Take-Up for Genetic Tests and Ambiguity. *Journal of Risk and Uncertainty* 48(2), 111–133.
- Hoy, M. and M. Polborn (2000). The Value of Genetic Information in the Life Insurance Market. *Journal of Public Economics* 78(3), 235–252.
- Ikram, M. A., S. Seshadri, J. C. Bis, M. Fornage, A. L. DeStefano, Y. S. Aulchenko, S. Debette, T. Lumley, A. R. Folsom, E. G. van den Herik, et al. (2009). Genomewide Association Studies of Stroke. *New England Journal of Medicine* 360(17), 1718–1728.
- Iossa, E. and D. Martimort (2015). Pessimistic Information Gathering. *Games and Economic Behavior* 91, 75–96.
- Lagerlöf, J. N. M. and C. Schottmüller (2015, August). Monopoly Insurance with Endogenous Information. Mimeo.
- Lee, T. K. (1982). Resource Information Policy and Federal Resource Leasing. *The Bell Journal of Economics* 13(2), 561–568.

- Lerman, C., S. Narod, K. Schulman, C. Hughes, A. Gomez-Caminero, G. Bonney, K. Gold, B. Trock, D. Main, J. Lynch, et al. (1996). BRCA1 Testing in Families with Hereditary Breast-Ovarian Cancer: A Prospective Study of Patient Decision Making and Outcomes. *Journal of the American Medical Association (JAMA)* 275(24), 1885–1892.
- Levy, D. E., S. D. Byfield, C. B. Comstock, J. E. Garber, S. Syngal, W. H. Crown, and A. E. Shields (2011). Underutilization of BRCA1/2 Testing to Guide Breast Cancer Treatment: Black and Hispanic Women Particularly at Risk. *Genetics in Medicine* 13(4), 349–355.
- Ligon, J. A. and P. D. Thistle (1996). Information Asymmetries and Informational Incentives in Monopolistic Insurance Markets. *Journal of Risk and Insurance* 63(3), 434–459.
- Marshall, J. M. (1974). Private Incentives and Public Information. *American Economic Review* 64, 373–390.
- Meiser, B. and S. Dunn (2000). Psychological Impact of Genetic Testing for Huntington’s Disease: An Update of the Literature. *Journal of Neurology, Neurosurgery & Psychiatry* 69(5), 574–578.
- Rothschild, M. and J. Stiglitz (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics* 90(4), 629–649.
- Schlee, E. (2001). The Value of Information in Efficient Risk Sharing Arrangements. *American Economic Review* 91(3), 509–524.
- Snow, A. (2015). Monopoly Insurance and the Value of Information. *Risks* 3, 277–289.
- Stiglitz, J. E. (1977). Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market. *The Review of Economic Studies* 44(3), 407–430.
- Sydsæter, K. and P. Hammond (1995). *Mathematics for Economic Analysis*. Prentice Hall International.
- Szalay, D. (2009). Contracts with Endogenous Information. *Games and Economic Behavior* 65(2), 586–625.
- Wilson, C. (1977). A Model of Insurance Markets with Incomplete Information. *Journal of Economic Theory* 16(2), 167–207.