

Imperfect private information and the design of information-generating mechanisms*

Frank Rosar[†] Elisabeth Schulte
University of Bonn University of Mannheim

This version: March 24, 2010

Abstract

An agent who is only imperfectly informed can use a device which generates public information about his type. While the agent's decision to use the device may signal his private information, the device can reveal information that goes beyond what the agent knows. The device shall be designed to learn as much about the agent's type as possible. The agent wants to be perceived as good and is risk-averse with respect to this perception. The optimal device commits at most one type of error: It may be subject to false negatives, but not to false positives. Moreover, the optimal device is either imperfect or not always used such that the agent's type cannot always be perfectly inferred. Inducing full participation can be optimal only if the agent's private information is very informative. Otherwise it is optimal to induce partial participation which allows perfect inference of private information.

JEL codes: D82, D83.

Keywords: imperfect information, information generation, signaling, mechanism design, no monetary transfers.

1 Introduction

We show in this paper how two features which are present in many design problems, but which do not fit into the classic design framework, may be incorporated and analyzed. The following problem highlights the features in which we are interested: The safety of operating systems, software applications and data storage is a major concern, in particular in organizations which

*We would like to thank Helmut Bester, Hans Peter Grüner, Tymofiy Mylovanov, Sergei Severinov, Konrad Stahl, Roland Strausz, as well as participants at the ESEM 2008 in Milan, the Jahrestagung des Vereins für Socialpolitik 2008 in Graz, and at seminars in Berlin and Mannheim for comments. A previous version of the model circulated under the title "Better Information Generation Through Inaccurate Signaling Devices." The first author gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15.

[†]Corresponding author: Frank Rosar, e-mail: email@frankrosar.de

handle sensitive data like government agencies. However, thorough safety assessments are only possible with access to the source code, which is at the producer's discretion to grant or to deny. The producer, on the other side, may not be perfectly informed about his product's safety either, for instance because it contains code provided by a third party or because the producer cannot perfectly assess whether there are still bugs in his own code. As a consequence, he is uncertain about what will be found if he grants access to the code. Although the decision not to grant access will be interpreted as an unfavorable signal, it may be better for the producer to face a relatively stable demand rather than a lottery of a very high demand following a positive assessment and a very low demand following a negative assessment. Such risk-averse preferences may for instance be induced by convex adjustment costs. Facing the producer's discretion to deny access, how should a test of an operating system or a software application be designed if one wants to achieve an as precise as possible assessment of its safety? The present paper proposes a model which allows to cope with such design questions.

The above problem is a non-standard design problem because two crucial assumptions are not met. In a standard problem with uncertainty about the agent's type it is assumed that the agent either perfectly knows his type or does not have an informational advantage at all, and that in the presence of private information the designer can only indirectly learn about the agent's type through his actions or reports. In our model we capture the features of the situation described above as follows: First, the agent is only imperfectly informed about his type. His private information is a probabilistic assessment of his type. Second, besides the possibility of learning about the agent's type through his actions, the designer can construct a device which, if used, directly reveals information about the agent's type. I.e. the mechanism reacts to the true information, not to the agent's imperfect private information. We assume that the designer wants to learn the agent's true type, while the agent wants to be perceived as a good type but is risk-averse with respect to this perception. The designer can observe either that the device is not used or that it is used and, if so, then also the generated information. If the decision to use the device is contingent on the agent's private information, then the observed participation behavior is indirectly revealing about the agent's type as his private information is informative about the true type. In addition, if the device is used, the generated information is directly revealing about the agent's type. *Ceteris paribus*, the more precise the information generated by the device is, the less attractive is participation for a risk-averse agent. Thus, if the usage of the device is at the agent's discretion, there may arise a trade-off between the quality of the generated information and the participation level. This paper explores this trade-off, giving rise to the following questions: First, for any intended participation behavior, what is the optimal quality? Related to that, which errors, if any, shall the device be subject to? False positives, false negatives, or both? Second, what is the optimal participation behavior, i.e. how important is indirect learning through self-selection into the device?

Imperfect private information in conjunction with the possibility to generate information plays an important role in many applications. A used-car salesman typically has better information

about the quality of a car than a potential buyer, but a technical inspection authority is able to generate information that goes beyond what the salesman knows. A local authority has a better picture about the terror potential in the country than the rest of the world. Local investigations by an international organization nevertheless may disclose information that goes beyond what the authority knows. The medical history of a patient's ancestors allows an imperfect assessment of his risk of having a certain genetic defect. A genetical scan can improve upon that information. Likewise, a student has an idea about his ability, but a professor might get an even better picture from a test, an exam or a personal conversation. All these examples as well as our introductory example have in common that information generation might not be possible without the agent's consent. The software producer decides whether to grant access to the source code. Having a car inspected requires the salesman's agreement, conducting investigations on sovereign territory requires the agreement of the local authority, and conducting a genetic scan is not legally possible without the patient's consent. Moreover, even if explicit agreement seems not to be required, some information-generating mechanisms offer the implicit possibility to avoid the generation of meaningful information. While it might not be an option for a student not to answer a question at all, it is often possible to evade questions by making meaningless statements instead of giving possibly incorrect answers.

In all these cases subsequently designed contracts (sales contracts, international agreements, insurance contracts, employment contracts) can condition on the available information. With more precise information the less informed party may save information rents and/or realize efficiency gains, i.e. the less informed party generally benefits from learning about the type. For the better informed party private information gives rise to a signaling motive. However, conditional on being able to signal superior private information, the generation of additional information may be harmful. As an example, consider the software producer who assesses his products' safety as above average. Though he has an interest in signaling a high safety standard, he might shy away from additional information generation because this introduces additional uncertainty in his demand which may be harmful due to convex adjustment costs. Such preferences can be represented by (reduced form) utility functions which are monotonic in the less informed party's perception and which exhibit risk-aversion with respect to the realized perception. Risk-aversion effectively constrains information generation. We explore optimal information generation under that constraint.

The paper proceeds as follows: In Section 3 we introduce the model. We argue in Section 4 that the usefulness of generating imperfect information goes hand in hand with the imperfection of private information. With perfect private information optimal learning can already be achieved with a device which is not prone to errors, without any private information learning is impossible.¹

¹Similar incentive effects as in our model may arise in a model where the agent is endowed with perfect private information but where he cannot perfectly anticipate which public signal will be generated if he uses the device, for instance because the device is always imperfect for technical reasons. An interesting question in such a framework is the following: Should the device designer spent effort to reduce the probability of false positives, of false negatives, or should he rather refrain from trying to improve the quality of the device. The answer to

The main analysis of the model is conducted in Sections 5, 6 and 7. What makes the design problem complicated is that general equilibrium type effects arise. Equilibrium behavior and the utility associated to a certain behavior/to the signals that may be generated by the device are simultaneously determined for any given device. Small adjustments in the device cause ample direct and indirect effects which generally go into different directions making compound effects complex and hard to assess. Moreover, we have to cope with the fact that a certain device may be consistent with a number of different equilibrium participation behaviors and a certain equilibrium participation behavior can in general be induced through different devices. First, in Section 5, we derive properties of equilibrium participation behavior and beliefs for given devices. Then, in Section 6, we explain how the designer must construct the device in order to induce a certain participation behavior. We characterize the set of implementable device–equilibrium–combinations. Finally, in Section 7, we study the designer’s optimal device choice.

Our main findings are: The optimal device is subject to only one of two possible types of error. Generated information that is “high” allows perfect inference that the agent’s type is good (no false positives), while generated information that is “low” allows only for imperfect inference (false negatives are possible). I.e., if the agent’s true type is bad, the device always generates a low signal, while the signal issued when the agent’s true type is good may be either high or low. With binary private information, it is optimal to induce a participation behavior from which the agent’s private information can either be perfectly deduced or not deduced at all. The importance of indirect learning through self–selection into the device depends on the informativeness of the agent’s private information. The designer prefers perfect inference of the agent’s private information to full participation except possibly in the case where the agent’s private information is very informative about his type. In this case it is sometimes optimal that private information cannot be inferred at all in equilibrium. Moreover, in order to increase the incentive of an agent with low private information to use the device, it must be harder to obtain a favorable signal.

In Section 8 we extend our model in two directions. First, we allow for a larger device space. In the mainpart of the paper information–generating devices generate binary public information, we show that such devices are indeed optimal when the designer can construct devices which generate an arbitrary finite number of public signals. Second, we show that the structure of the optimal device is robust with respect to an alternative specification of private information. While the analysis in the main part of the paper is for binary private information, we show that the optimal device has the same properties for continuous private information. We conclude in

those questions is directly related to the analysis in the present paper.

As an example, in 2000, being accused of drug abuse, the designated coach of the German national football team voluntarily agreed on a drug test which turned out to be positive. Seemingly, he was hoping for a negative result of the imperfect test although he was perfectly informed about his guilt. In that particular case, the imperfectness of the generated information is triggered by technical restrictions. However, with a more informative device, the drug test would probably not have been carried out such that no information would have been generated. In the present paper, we show that imperfectness may arise endogenously in the absence of exogenous technical restrictions.

2 Related literature

We begin our literature review with contributions that study related design problems and go on with papers which share essential ingredients of our model framework.

Ostrovsky and Schwarz (2008) assume that schools maximize the average utility of their students. Noisy reports about students' abilities are optimal if the students' utilities are concave in their perceived ability. Noisy reports induce a redistribution from high ability to low ability agents, which increases average utility as the former lose less than the latter gain. There is neither private information nor a participation decision on the students's side in their model. Under the assumption that the cost of a (tax-financed) public signaling device increase in its accuracy, Stiglitz (1975) derives the individually preferred accuracy. Due to the pooling effect, agents with types above the mode want a higher accuracy than those below the mode. Stiglitz does not consider voluntary participation. However, he points out that risk averse individuals who are not perfectly informed about their types may prefer to be treated as averages. The problem, he observes, is the fact that individuals cannot insure against their type-risk.² Kamenica and Gentzkow (2009) consider optimal persuasion mechanisms where the designer, as in our paper, controls the conditional probabilities with which certain signals will be issued. The designer's motive is to influence the beliefs of a decision maker. There is no private information and participation is not an issue. Information is generated only if the designer's (reduced form) utility is convex in the induced belief. Whereas the designer in Kamenica and Gentzkow (2009) is interested in providing as little information as necessary in order to persuade the decision maker to take a certain action, the designer in control of the information-generating mechanism in our paper wants to generate as precise information as possible.

Important applications in which learning about an agent's type is essential are task assignments or partner matching.³ There, the learning technology is exogenous, there is symmetric

²This has also been pointed out by Hirshleifer (1971). He argues that the revelation of information harms those whose endowment is worthless in the revealed state. Risk averse individuals do not want a public announcement of information prior to trading. In our paper, it is not possible to insure the type-risk either. However, an imperfection of the generated information provides at least a partial insurance as it becomes harder to discriminate between the types such that their payoffs move closer together.

³In Prescott and Visscher (1980), a firm learns about the workers type through experimenting with task assignment, where a trade-off arises between generating more precise information and using the already generated information. In Eeckhout and Weng (2009), it is the agent who chooses a firm and therewith the learning process. In Waldman (1989), the firm has direct control over the precision of generated information. A cost of information generation and public observability constrain the precision. In MacDonald (1980) and in Burdett and Mortensen (1981), it is the agent who invests in information. There is no superior type in MacDonald (1980), i.e. better information impacts only on efficiency which yields an incentive to invest. In Burdett and Mortensen (1981), the agent decides whether to buy a perfectly revealing signal. It is assumed that the agent internalizes the efficiency gains from better allocation which yields an incentive to invest. In Anderson and Smith (2010), agents with uncertain types match and learning takes place through stochastic production.

information and the agent’s agreement to information generation is not an issue. In our paper instead, the designer is free to choose the precision of information but needs the (better-informed) agent’s agreement for information generation.

Provision of hard information is key in certification procedures, as analyzed e.g. in Lerner and Tirole (2006) or in Farhi, Lerner, and Tirole (2005).⁴ Again, in these contributions the information acquisition technology is exogenous, whereas it is the object of analysis in our paper. Grossman and Hart (1980), Milgrom (1981) and Okuno-Fujiwara, Postlewaite, and Suzumura (1990) point out an unraveling effect of the availability of hard information. However, in our model there is imperfect private information. Thus, complete unraveling is not possible. Risk aversion on the agent’s part attenuates his desire to signal superior information. Ex-post information necessarily remains imperfect, either because information is not always generated or because the generated information is imperfect.

Imperfect private information about a payoff-relevant state variable plays an essential role in the literature on committee decision making.⁵ There, information generation takes place by choosing the set of committee members, taking into account incentives for information acquisition and revelation. Committee members’ payoffs depend on the beliefs assigned to the state variable only indirectly via its influence on individual votes, and hence on the decision.

Geanakoplos and Pearce (1989) introduced a class of games, called psychological games, in which players’ payoffs do not only depend on actions, but also on a system of beliefs about actions. Therewith, it is possible to capture belief-based preferences such as surprise or guilt. Related to that, the (reduced form) utilities functions in the present paper depend (only) on beliefs.

3 The model

3.1 Information, devices and timing

There are two players, an agent (he) who is imperfectly informed about his type and the designer of an information-generating device (she). If the device is used by the agent, it generates information about his type.

The agent’s type is either *good* ($\omega = 1$) or *bad* ($\omega = 0$), but he is not completely sure about what his type actually is. He observes a private signal which is either *high* ($\theta = h$) or *low* ($\theta = l$) and which is imperfectly correlated with his true type. We denote the joint probability of type ω and private information θ by $p_{\theta\omega} := \text{Prob}(\theta, \omega)$ and the marginal distributions by $p_{\theta} := \sum_{\omega} p_{\theta\omega}$ and by $p_{\omega} := \sum_{\theta} p_{\theta\omega}$. Both types and both private signals occur with positive probability and there is no private signal from which the agent can perfectly infer his type, i.e. $p_{\theta\omega} \in (0, 1)$. Moreover, an agent with a high private signal is more likely to be good than an agent with a low private signal, i.e. $\delta := \text{Prob}(\omega = 1|\theta = h) - \text{Prob}(\omega = 1|\theta = l) = \frac{p_{h1}}{p_h} - \frac{p_{l1}}{p_l} > 0$. δ is our

⁴Information becomes “hard” because of a shared interest between the certifier and the receiver of information.

⁵For a survey, see Gerling, Grüner, Kiel, and Schulte (2005).

| | $\omega = 0$ | $\omega = 1$ |
|--------------|--------------------------------------|--------------------------------------|
| $\theta = l$ | $p_{l0} := p_l p_0 + \delta p_l p_h$ | $p_{l1} := p_l p_1 - \delta p_l p_h$ |
| $\theta = h$ | $p_{h0} := p_h p_0 - \delta p_l p_h$ | $p_{h1} := p_h p_1 + \delta p_l p_h$ |

Table 1: Joint distribution of private information and type

measure for how informative the agent’s private information is about his type. The higher δ , the better the agent can infer his type from his private information. If δ is close to zero, he can infer almost nothing.

The joint distribution of the private information and the type possesses three degrees of freedom. It will be convenient to choose $p_l \in (0, 1)$, $p_h \in (0, 1)$ and $\delta \in (0, \delta_{\max})$ with $\delta_{\max} := \min\{\frac{p_l}{p_h}, \frac{1-p_l}{1-p_h}\}$ as the primitives of our model instead of three probabilities of the joint distribution.⁶ The joint distribution that is induced by our primitives $[p_l, p_h, \delta]$ is as stated in Table 1.

The agent has the opportunity to use the designer’s device which generates information about his type. If such information is generated, it is public in the sense that it is also observed by the designer. As device space we consider the set of all probabilistic mappings from types into either a *high* public signal ($\sigma' = H$) or a *low* public signal ($\sigma' = L$). Any device can be described by two parameters $(s_0, s_1) \in [0, 1]^2$ where $s_\omega := \text{Prob}(\sigma' = H|\omega)$ is the probability that a high public signal is generated when the agent’s true type is ω . Without loss of generality we can assume $s_0 \leq s_1$. We call devices with $s_0 < s_1$ informative and devices with $s_0 = s_1$ uninformative. To simplify the statement of our results and the exposition of our proofs, we consider only informative devices, i.e. we assume $s_0 < s_1$.⁷ Hence, there is a positive correlation between obtaining a high public signal and being good.

The device $(s_0, s_1) = (0, 1)$ is called perfect as it, if used, generates perfect information about the agent’s type, i.e. it generates high public signals if and only if the agent is good. Any other device is subject subject to at least one of two kinds of errors: For devices with $s_1 < 1$ a good agent may get a low public signal (false negative) and for devices with $s_0 > 0$ a bad agent may get a high public signal (false positive).

The device may be interpreted as a test which produces pass–fail–results. A false negative then means that a good agent fails the test while a false positive means that a bad agent passes the test.⁸ While considering only pass–fail–tests might appear restrictive at a first glance, we show

⁶This is possible because the following lemma holds: Let $\mathcal{A} := (0, 1) \times (0, 1) \times (0, \delta_{\max})$ and $\mathcal{B} := \{[p_{\theta\omega}] \in (0, 1)^4 \mid \frac{p_{h1}}{p_{h0}+p_{h1}} - \frac{p_{l1}}{p_{l0}+p_{l1}} > 0\}$. (i) If $[p_{\theta\omega}] \in \mathcal{B}$, then $(p_l, p_h, \delta) \in \mathcal{A}$. (ii) If $(p_l, p_h, \delta) \in \mathcal{A}$, then there exists a unique distribution $[p_{\theta\omega}]$ such that $p_l = p_{l1} + p_{h1}$, $p_h = p_{h0} + p_{h1}$ and $\delta = \frac{p_{h1}}{p_h} - \frac{p_{l1}}{p_l}$. Moreover, $[p_{\theta\omega}] \in \mathcal{B}$.

⁷It is never optimal for the designer to choose an uninformative device. If the designer chooses an uninformative device, there exists no equilibrium where she learns anything about the agent’s type, while there always exist informative devices for which learning is possible.

⁸In an “indirect implementation” of the test, s_0 and s_1 would be determined by elements of luck (e.g. multiple choice), the difficulty and scope of tasks, and the time given to solve the test.

in Subsection 8.1 that such tests are indeed optimal within the model framework we consider.

As the agent is not aware whether his true type is 0 or 1, he is only indirectly interested in s_0 and s_1 . He is interested in how likely it is for him to get a high public signal taking into account his private information. We denote this probability by $s_\theta := \text{Prob}(\sigma' = H|\theta)$ and obtain

$$\begin{aligned} s_l &= \frac{p_{l0}}{p_l}s_0 + \frac{p_{l1}}{p_l}s_1 = p_0s_0 + p_1s_1 - \delta p_h(s_1 - s_0) \\ s_h &= \frac{p_{h0}}{p_h}s_0 + \frac{p_{h1}}{p_h}s_1 = p_0s_0 + p_1s_1 + \delta p_l(s_1 - s_0). \end{aligned}$$

The agent decides between using the device ($d = Y$) and not using the device ($d = N$). We denote the agent's mixed participation behavior if he has private information θ by $\alpha_\theta := \text{Prob}(d = Y|\theta)$.

The timing is as follows: First, the designer chooses a device $(s_0, s_1) \in [0, 1]^2$ with $s_1 > s_0$. Then, the agent learns his private information $\theta \in \{l, h\}$ and chooses a mixed participation behavior $\alpha_\theta \in [0, 1]$. If the device is used, it generates a signal $\sigma' \in \{L, H\}$ which is publicly observed. Finally, the designer updates her belief about the agent's type according to Bayes' Law (as far as possible) and payoffs (as described below) realize.

The designer of the device can observe whether the device is used and, if it is used, she observes also the generated public signal. Thus, there are three different observations she can make: $(d = N)$, $(d = Y, \sigma' = L)$ and $(d = Y, \sigma' = H)$. To simplify notation, we say that the public signal that is generated is $\sigma \in \{N, L, H\}$ with $\sigma = N$ meaning $(d = N)$, $\sigma = L$ meaning $(d = Y, \sigma' = L)$ and $\sigma = H$ meaning $(d = Y, \sigma' = H)$. The probability with which the designer observes public signal σ is $p_\sigma := \text{Prob}(\sigma)$.⁹

3.2 Beliefs

Private information and types do not coincide in our framework. The designer can form two different kinds of beliefs after observing public signal σ : beliefs about the agent's private information and beliefs about the agent's type.

To describe these beliefs, we introduce the notation $p_{\sigma\theta} := \text{Prob}(\sigma, \theta)$ for the joint probability of public signal σ and private information θ , and $p_{\sigma\omega} := \text{Prob}(\sigma, \omega)$ for the joint probability of public signal σ and type ω .¹⁰ If signal σ is generated with positive probability, beliefs for this signal are determined by Bayes' Law and we obtain that after observing σ the designer believes that the agent's private information is high with probability $\eta_\sigma := \text{Prob}(\theta = h|\sigma) = \frac{p_{\sigma h}}{p_\sigma}$ and that his type is good with probability $\mu_\sigma := \text{Prob}(\omega = 1|\sigma) = \frac{p_{\sigma 1}}{p_\sigma}$. Moreover, the designer infers $\eta_Y := \text{Prob}(\theta = h|d = Y)$ and $\mu_Y := \text{Prob}(\omega = 1|d = Y)$ from the agent's participation (before the signal realizes).¹¹

⁹We have $p_N = p_l(1 - \alpha_l) + p_h(1 - \alpha_h)$, $p_L = [p_{l1}\alpha_l + p_{h1}\alpha_h](1 - s_1) + [p_{l0}\alpha_l + p_{h0}\alpha_h](1 - s_0)$ and $p_H = [p_{l1}\alpha_l + p_{h1}\alpha_h]s_1 + [p_{l0}\alpha_l + p_{h0}\alpha_h]s_0$.

¹⁰We have $p_{Nh} = p_h(1 - \alpha_h)$, $p_{Lh} = [p_{h1}\alpha_h](1 - s_1) + [p_{h0}\alpha_h](1 - s_0)$, $p_{Hh} = [p_{h1}\alpha_h]s_1 + [p_{h0}\alpha_h]s_0$, $p_{N1} = p_{l1}(1 - \alpha_l) + p_{h1}(1 - \alpha_h)$, $p_{L1} = [p_{l1}\alpha_l](1 - s_1) + [p_{h1}\alpha_h](1 - s_1)$ and $p_{H1} = [p_{l1}\alpha_l]s_1 + [p_{h1}\alpha_h]s_1$.

¹¹We have $\eta_Y = \frac{p_{h1}\alpha_h + p_{l1}\alpha_l}{p_h\alpha_h + p_l\alpha_l}$ and $\mu_Y = \frac{p_{h1}\alpha_h + p_{l1}\alpha_l}{p_h\alpha_h + p_l\alpha_l}$.

If non-participation occurs off the equilibrium path, the designer is nevertheless constrained with respect to her belief about the agent: Since the agent is himself imperfectly informed, the worst inference that the designer may make is that the agent's private signal is low and not that he is actually bad.¹²

Only the beliefs for $\sigma = N$ in case the device is always used and the beliefs for $\sigma = L$ and $\sigma = H$ in case the device is never used are not pinned down by Bayes' Law. We assume that if the designer observes that the device is not used although the agent is supposed to always use it, she believes that she faces an agent with low private information. Conversely, if the designer observes that the device is used although the agent is supposed to never use it, she believes that the agent has high private information. Since—as will be shown later—an agent with high private signal has a strictly higher incentive to use the device than an agent with a low private signal, these are just the beliefs which survive the Intuitive Criterion. Formally, we assume $\eta_N = 0$ if $\alpha_l = \alpha_h = 1$ and $\eta_Y = 1$ if $\alpha_l = \alpha_h = 0$. It follows directly $\mu_N = \frac{p_l}{p_l}$ if $\alpha_l = \alpha_h = 1$ and $\mu_L = \frac{p_{h1}(1-s_1)}{p_{h1}(1-s_1)+p_{h0}(1-s_0)}$ as well as $\mu_H = \frac{p_{h1}s_1}{p_{h1}s_1+p_{h0}s_0}$ if $\alpha_l = \alpha_h = 0$.

3.3 Reduced form utility functions

We are interested in the (reduced form) problem where the designer is interested in learning about the agent's type and the agent wants to be perceived as good by the designer and is risk averse with respect to this perception. Therefore the agent's and the designer's utility depend directly on the designer's *belief about the agent's type* to which we will refer just as "*belief*" or "*perception*" in the remainder of the paper.

If an agent with private information θ uses the device, the belief will be μ_L with probability $1 - s_\theta$ and μ_H with probability s_θ . Thus, the agent faces a lottery with expected value $\mathbb{E}_\theta := (1 - s_\theta)\mu_L + s_\theta\mu_H$ and variance $\mathbb{V}_\theta := (1 - s_\theta)\mu_L^2 + s_\theta\mu_H^2 - \mathbb{E}_\theta^2$. If the agent does not use the device, the belief will be μ_N for sure. The agent evaluates beliefs with a smooth utility index $u(\mu)$ which is strictly increasing and strictly concave on $(0, 1)$, and he chooses his participation behavior to maximize expected utility

$$U_\theta := \begin{cases} U_\theta^Y := (1 - s_\theta)u(\mu_L) + s_\theta u(\mu_H) & \text{if } d = Y \\ U^N := u(\mu_N) & \text{if } d = N \end{cases}.$$

To abbreviate notation in the proofs, we will occasionally write u_σ and u'_σ instead of $u(\mu_\sigma)$ and $u'(\mu_\sigma)$, respectively.

For the designer, the device choice induces a lottery over beliefs where belief μ_σ occurs with probability p_σ . She is interested in learning about the agent's type and thus prefers learning the realization of any lottery to learning its expected value. A utility function to represent such preferences is convex in beliefs. We assume that the designer evaluates beliefs with a strictly convex utility index $v(\mu)$ and that she chooses a device to maximize her expected utility $V := p_N v(\mu_N) + p_L v(\mu_L) + p_H v(\mu_H)$.

¹²Hence, "forcing" full participation with adverse beliefs in case of non-participation is not possible in our framework.

3.4 A non-reduced problem

Suppose that after observing public signal $\sigma \in \{N, L, H\}$, the designer assesses her opinion about the agent in a report. Her assessment $\hat{a} \in [0, 1]$ is the probability with which she believes the agent to be good. Her aim is to be as accurate as possible. Formally, she may strive for minimizing the quadratic difference between her assessment \hat{a} and the agent's actual type ω such that her non-reduced form utility index is $\tilde{v}(\hat{a}, \omega) = -(\hat{a} - \omega)^2$. Hence, she maximizes $\mathbf{E}[\tilde{v}(\hat{a}, \omega)|\sigma] = -\mu_\sigma(\hat{a} - 1)^2 - (1 - \mu_\sigma)(\hat{a} - 0)^2$ and it follows that the optimal assessment is $\hat{a} = \mu_\sigma$ inducing a reduced form utility index $v(\mu) = -\mu(1 - \mu)$. The agent is interested in getting a good assessment and he is risk-averse with respect to how he is assessed. For instance, he might minimize the quadratic difference between the assessment he actually obtains and the best possible assessment leading to a non-reduced form utility index $\tilde{u}(\hat{a}) = -(1 - \hat{a})^2$ and a reduced form utility index $u(\mu) = -(1 - \mu)^2$.

In the remainder of the paper we will restrict attention to the reduced form problem and we will extend our analysis to the non-quadratic case to make clear which properties of the utility indices drive our results. All our results hold for any quadratic utility indices, $u(\mu)$ and $v(\mu)$ with $u' > 0$, $u'' < 0$ and $v'' > 0$. A reader not interested in the technical details might therefore ignore the conditions imposed on $u(\cdot)$ and $v(\cdot)$ in the statement of our results and just think of the quadratic case.

3.5 Equilibrium concept

An equilibrium of the game induced by a device $(s_0, s_1) \in [0, 1]^2$ with $s_0 < s_1$ is any combination of beliefs $(\mu_N, \mu_L, \mu_H) \in [0, 1]^3$ and any mixed participation strategy $(\alpha_l, \alpha_h) \in [0, 1]^2$ such that

1. given a certain participation strategy, beliefs are consistent with Bayes' Law and our assumption on out-of-equilibrium beliefs,
2. given beliefs, the agent's participation behavior is individually rational, i.e. $\alpha_\theta = 1$ if $U^N < U_\theta^Y$, $\alpha_\theta \in [0, 1]$ if $U^N = U_\theta^Y$ and $\alpha_\theta = 0$ if $U^N > U_\theta^Y$.

As common in the mechanism design literature, we ignore the issue of equilibrium multiplicity and assume that the designer can pick her favorite equilibrium when multiple equilibria exist.

4 Relation to the classical design framework

There are two benchmark cases: The case where the agent has no private information (i.e., $\delta = 0$) and the case where he has perfect private information (i.e., $\delta = 1$).

If the agent does not have any private information, there exists no equilibrium where he is willing to use an informative device. This is so because the expected perception conditional on using and not using the device is the same. However, using the device involves a risk regarding which perception will realize. A risk-averse agent thus has a strict incentive to abstain from

using the device. Hence, with no informational advantage on the agent’s side, it is not possible for the designer to learn anything about his type.

By contrast, if private information is perfect, an unraveling argument applies. Since the agent’s private signal is perfectly correlated with the public signal generated by a perfect device, the agent can perfectly foresee which public signal will be generated by a perfect device and thus faces no risk regarding the perception that will realize. As a consequence, there are two ways for the designer to perfectly learn the agent’s type: First, there is an equilibrium where the agent uses the perfect device if and only if his private signal is high. When the device is not used, the designer can infer that the agent has a low private signal, i.e. that he is bad. Second, there is an equilibrium where the agent always uses the perfect device and the device reveals the agent’s type. While in the first case the designer learns perfectly via the agent’s private information through his participation behavior, he learns perfectly via the information generated by the device in the second case.

The problem with imperfect private information (i.e., $\delta \in (0, 1)$) differs in two respects from the benchmark cases. First, while there is never a use for an imperfect device in the benchmark cases, such devices might be optimal with imperfect information.¹³ We are therefore interested in the questions *Is the optimal device imperfect? If so, which kind of errors is it subject to?* Second, while with perfect private information it makes no difference for the designer to learn via the agent’s private information through his participation decision or via the information generated by the device, it does make a difference with imperfect private information. Hence, we study the question *Through which channels does the designer want to learn about the agent’s type?*

5 Properties of equilibrium candidates for a given device

In this section we take the device as given and show how candidates for the equilibrium participation behavior look like and which beliefs are associated with them. Afterwards, in Section 6, we analyze how the designer has to construct the device in order to induce an equilibrium with a certain participation behavior and in Section 7 we derive the designer’s optimal device choice.

Due to Bayes’ Law and our assumption on out-of-equilibrium beliefs, beliefs are for a given device (s_0, s_1) uniquely pinned down by the participation strategy (α_l, α_h) . Therefore, as long as it is clear which device we consider, we can write $U_\theta^Y(\alpha_l, \alpha_h)$ and $U^N(\alpha_l, \alpha_h)$ for the agent’s utility from using and from not using the device when the supposed participation strategy is (α_l, α_h) .

For any informative device, the agent has a strictly larger incentive to use the device when his private signal is high than when it is low.

Lemma 1 *For any device and for any (α_l, α_h) , $U_l^Y(\alpha_l, \alpha_h) < U_h^Y(\alpha_l, \alpha_h)$.*

¹³If δ is sufficiently small, the agent never uses the perfect device. Then, neither is the observed participation behavior informative, nor is information generated by the device. However, it is always possible to learn by implementing an informative imperfect device which an agent with a high private signal is willing to use.

type 0: $\alpha_l = 0$ and $\alpha_h = 0$; necessary and sufficient: $U_h^Y(0, 0) \leq U^N(0, 0)$

type I: Equilibria with $\alpha_l = 0$ and $\alpha_h \in (0, 1]$.

Ia: $\alpha_h \in (0, 1)$; necessary and sufficient: $U^N(0, \alpha_h) = U_h^Y(0, \alpha_h)$

Ib: $\alpha_h = 1$; necessary and sufficient: $U_l^Y(0, 1) \leq U^N(0, 1) \leq U_h^Y(0, 1)$

type II: Equilibria with $\alpha_l \in (0, 1]$ and $\alpha_h = 1$.

IIa: $\alpha_l \in (0, 1)$; necessary and sufficient: $U^N(\alpha_l, 1) = U_l^Y(\alpha_l, 1)$

IIb: $\alpha_l = 1$; necessary and sufficient: $U^N(1, 1) \leq U_l^Y(1, 1)$

Table 2: Classification of equilibrium types

It follows directly that it cannot happen that the agent uses the device with positive probability when his private signal is low, while he uses it with probability less than one when his private signal is high.

Lemma 2 *For any device and in any equilibrium, $\alpha_l = 0$ or $\alpha_h = 1$.*

Three different types of equilibrium may occur. The definitions of these types and the associated necessary and sufficient equilibrium conditions are stated in Table 2. In an equilibrium of type I the agent never uses the device when his private signal is low, while in an equilibrium of type II an agent with a low private signal uses the device with positive probability. An equilibrium where the device is never used is of type 0. In the course of the analysis, equilibria in which the agent uses the device if and only if his private signal is high (type Ib) and equilibria in which he always uses the device (type IIb) will be of special importance. In an equilibrium of type Ib the designer can perfectly infer the agent's private information from his participation behavior.¹⁴ We call such an equilibrium an *equilibrium with perfect separation of private information*. By contrast, in an equilibrium of type IIb, the designer can infer nothing about the agent's private information from his participation behavior.¹⁵ We call such an equilibrium an *equilibrium with perfect pooling of private information*.

Figure 1 displays how beliefs typically depend on the agent's participation behavior. On the left part of the graph beliefs are displayed for $\alpha_l = 0$ and increasing α_h (type 0/I), on the right part beliefs are displayed for $\alpha_h = 1$ and increasing α_l (type II). Note that the beliefs at the very left and the very right of the graph are not completely pinned down by Bayes' Law. The beliefs which we specified for the out-of-equilibrium events have the effect that the belief mapping becomes continuous at the very left and the very right of the graph.

If $\alpha_l = 0$ and $\alpha_h \in [0, 1]$, beliefs are

$$\mu_N^I := \frac{p_{l1} + p_{h1}(1 - \alpha_h)}{p_l + p_h(1 - \alpha_h)}, \quad \mu_L^I := \frac{p_{h1}(1 - s_1)}{p_{h1}(1 - s_1) + p_{h0}(1 - s_0)} \quad \text{and} \quad \mu_H^I := \frac{p_{h1}s_1}{p_{h1}s_1 + p_{h0}s_0}.$$

¹⁴I.e., $\eta_Y = 1$ and $\eta_N = 0$.

¹⁵I.e., $\eta_Y = p_h$ and non-participation does not occur with positive probability.

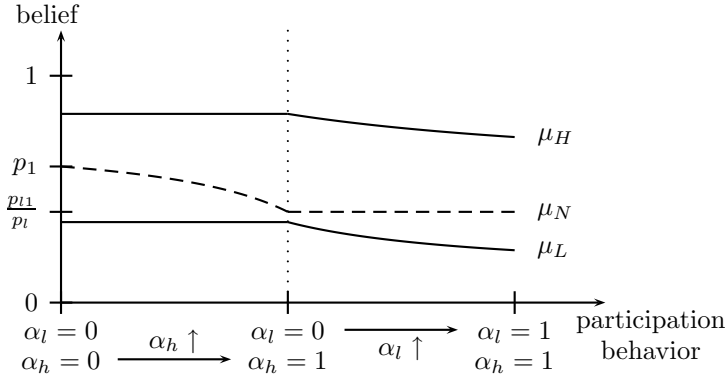


Figure 1: Beliefs ($p_1 = \frac{3}{5}, p_h = \frac{3}{5}, \delta = \frac{1}{3}, s_0 = \frac{1}{2}, s_1 = \frac{9}{10}$)

If $\alpha_h = 1$ and $\alpha_l \in [0, 1]$, beliefs are

$$\mu_N^{II} := \frac{p_{l1}}{p_l}, \mu_L^{II} := \frac{\mu_Y(1-s_1)}{p_{L|Y}} \text{ and } \mu_H^{II} := \frac{\mu_Y s_1}{p_{H|Y}}$$

where $p_{L|Y} := \mu_Y(1-s_1) + (1-\mu_Y)(1-s_0)$ and $p_{H|Y} := \mu_Y s_1 + (1-\mu_Y)s_0$. $p_{L|Y}$ and $p_{H|Y}$ are the probabilities that a high and a low public signal are generated conditional on that the device is used, respectively. Note that μ_Y depends on who is using the device, i.e. on α_l .

The structure of beliefs as displayed in Figure 1 is the general structure except for two aspects: First, in Figure 1 $\mu_N > \mu_L$ is true for any participation behavior. However, the device may also be constructed such that $\mu_N < \mu_L$ is true for some participation behavior. I.e., it is endogenous whether not using the device is a worse signal than obtaining a low public signal or vice versa. Second, for devices with $s_0 = 0$ or $s_1 = 1$ the graph looks somewhat differently. For $s_0 = 0$ a bad agent never obtains a high public signal such that from a high public signal the designer can perfectly infer that the agent is good (i.e., there are no false positives). Hence, $\mu_H = 1$ for any participation behavior. Similarly, for $s_1 = 1$ a good agent never obtains a low public signal such that from a low public signal the designer can perfectly infer that the agent is bad (i.e., there are no false negatives). Hence, $\mu_L = 0$ for any participation behavior.

6 Characterization of device–equilibrium–combinations

In this section we take a participation behavior as given and show how the designer has to construct the device in order to induce the respective participation behavior. We consider the case with $\alpha_h = 1$ and $\alpha_l \in [0, 1]$ in Subsection 6.1 and with $\alpha_l = 0$ and $\alpha_h \in [0, 1]$ in Subsection 6.2. Therewith we characterize all device–equilibrium–combinations which might be optimal for the designer.¹⁶ The same participation behavior can in general be induced by different devices.

¹⁶In Subsection 6.1 we characterize all the device–equilibrium–combinations for which the participation constraint is binding. This is without loss of generality as we show in Section 7 that it cannot be optimal for the designer to have the participation constraint slack. In Subsection 6.2 we distinguish two cases. For parameter val-

We also analyze how these devices differ. This will be helpful in the analysis of the designer’s optimal device choice in Section 7.

6.1 Inducing participation of an agent with a low private signal

Suppose $\alpha_h = 1$ and $\alpha_l \in [0, 1]$. The problem to incentivize an agent with a low private signal to use the device is the following: If an agent with low private information does not use the device, he is perceived as having low private information. Only an agent with low private information is supposed to abstain from using the device, i.e. $\text{Prob}(\omega = 1|d = N, \theta = l) = \frac{p_l}{p_l}$. If an agent with low private information uses the device instead, he faces a lottery over beliefs, i.e. he incurs a risk regarding how he is perceived. It can only be optimal for him to use the device if the lottery yields an expected perception that is higher than $\frac{p_l}{p_l}$. However, if the device is perfect, a fair lottery is induced and the resulting expected belief is just $\text{Prob}(\omega = 1|d = Y, \theta = l) = \frac{p_l}{p_l}$. Hence, an agent with a low private signal has a strict incentive not to use the perfect device. To incentivize him, the designer has to “redistribute” expected perception from the case where his private signal is high to that where it is low.¹⁷

How does redistribution work? When the device is perfect, a good agent always obtains a high public signal and a bad agent always obtains a low public signal. I.e., the expected belief obtained by a good and by a bad agent are 1 and 0, respectively. When the device is imperfect, a bad agent can obtain signals that also a good agent can obtain such that the type cannot be perfectly inferred anymore. As a consequence, the expected belief obtained by a good agent is strictly below 1 and that of a bad agent is strictly above 0. I.e., expected belief is redistributed from the case where the agent is good to that where he is bad and, by consequence, expected perception is also redistributed from the case where it is more likely that the agent is good (= high private signal) to that where it is less likely that he is good (= low private signal). Hence, the designer can incentivize an agent with a low private signal to use the device by making the device susceptible to errors.

The device can be made more susceptible to errors by increasing the probability of false positives (i.e., $s_0 \uparrow$) or false negatives (i.e., $s_1 \downarrow$). By increasing s_0 , it becomes relatively more likely for the bad agent than for the good agent to obtain a high public signal and relatively less

ues for which the agent is not willing to use the perfect device, we characterize all device–equilibrium–combinations for which the participation constraint is binding. Again, as we show in Section 7, that this is without loss of generality. For all other parameter values, the agent is willing to use the perfect device with certainty if his private information is high. Since this must already specify the for the designer optimal device which induces the considered type of participation behavior, we do not characterize any other device–equilibrium–combinations.

¹⁷The term “redistribution of expected perception” is appropriate since Bayes’ Law dictates that the expected perception from using the device is not affected by the device choice for a given participation behavior, i.e. $p_l \text{Prob}(\omega = 1|d = Y, \theta = l) + p_h \text{Prob}(\omega = 1|d = Y, \theta = h) = \mu_Y$. This condition can be interpreted as a “budget constraint” where $\text{Prob}(\omega = 1|d = Y, \theta = l)$ and $\text{Prob}(\omega = 1|d = Y, \theta = h)$ is expected perception that is “distributed” to the agent when his private information is low and high, respectively. If a change of the device leads to an increase of $\text{Prob}(\omega = 1|d = Y, \theta = l)$ by one unit, it causes a decrease of $\text{Prob}(\omega = 1|d = Y, \theta = h)$ by $\frac{p_l}{p_h}$ units.

likely to obtain a low public signal. As a consequence, μ_H^{II} decreases and μ_L^{II} increases:¹⁸

$$\frac{\partial \mu_H^{II}}{\partial s_0} = -\frac{(1-\mu_Y)\mu_Y}{p_{H|Y}^2} s_1 = -\frac{(1-\mu_Y)}{p_{H|Y}} \mu_H^{II} < 0 \quad (1)$$

$$\frac{\partial \mu_L^{II}}{\partial s_0} = \frac{(1-\mu_Y)\mu_Y}{p_{L|Y}^2} (1-s_1) = \frac{(1-\mu_Y)}{p_{L|Y}} \mu_L^{II} \geq 0 \quad (2)$$

By decreasing s_1 , it becomes relatively more likely for the good agent than for a bad agent to obtain a low public signal and relatively less likely to obtain a high public signal. Thus, again, μ_H^{II} decreases and μ_L^{II} increases:

$$\frac{\partial \mu_H^{II}}{\partial s_1} = \frac{(1-\mu_Y)\mu_Y}{p_{H|Y}^2} s_0 = \frac{\mu_Y}{p_{H|Y}} (1-\mu_H^{II}) \geq 0 \quad (3)$$

$$\frac{\partial \mu_L^{II}}{\partial s_1} = -\frac{(1-\mu_Y)\mu_Y}{p_{L|Y}^2} (1-s_0) = -\frac{\mu_Y}{p_{L|Y}} (1-\mu_L^{II}) < 0 \quad (4)$$

If the designer makes the device less informative by decreasing s_1 and/or by increasing s_0 , the incentive of an agent with a low private signal to use the device increases strictly.

Lemma 3 For any $\alpha_l \in [0, 1]$, $U_l^Y(\alpha_l, 1)$ is differentiable in s_0 and s_1 with $\frac{\partial}{\partial s_0} U_l^Y(\alpha_l, 1) > 0$ and $\frac{\partial}{\partial s_1} U_l^Y(\alpha_l, 1) < 0$.

Consider some fixed participation behavior $(\alpha_l, 1)$. The designer can find a device which induces this participation behavior as follows: The relevant participation constraint is $U_l^Y(\alpha_l, 1) = u(\frac{p_H}{p_l})$. If the agent's private signal is low, he has a strict incentive not to use the perfect device, i.e. $U_l^Y(\alpha_l, 1) < u(\frac{p_H}{p_l})$ for $(s_0, s_1) = (0, 1)$. However, he has a strict incentive to use a completely uninformative device as this allows him pooling with high private information without incurring a risk, i.e. $U_l^Y(\alpha_l, 1) > u(\frac{p_H}{p_l})$ for $(s_0, s_1) = (0, 0)$. From this and a continuity property we obtain by an Intermediate Value Theorem that there always exists a device $(0, s'_1)$ with $s'_1 \in (0, 1)$ which implements an equilibrium with participation behavior $(\alpha_l, 1)$. But besides device $(0, s'_1)$, there exist also other devices which implement this participation behavior. For instance, when the designer increases s_0 and s_1 slightly in the right proportion, the same behavior is implemented. Moreover, it follows from Lemma 3 that for any s_0 there exists at most one s_1 for which the participation constraint is binding. This allows us to describe the set of all devices for which the participation constraint is binding by a domain of s_0 -values and a function $s_1(s_0)$ on this domain.¹⁹

¹⁸We present two different ways to describe the partial derivatives of the beliefs. Either of them will be helpful to simplify different steps in the subsequent proofs. To refer to the first or the second way specifically, we append an a or a b behind the number of the formula, respectively. I.e., (1a) refers to $\frac{\partial \mu_H^{II}}{\partial s_0} = -\frac{(1-\mu_Y)\mu_Y}{p_{H|Y}^2} s_1$, while (1b)

refers to $\frac{\partial \mu_H^{II}}{\partial s_0} = -\frac{(1-\mu_Y)}{p_{H|Y}} \mu_H^{II}$.

¹⁹As the set of devices for which the participation constraint is binding depends on the participation behavior considered, a more concise notation would be $s_1^{(\alpha_l, 1)}(s_0)$. For the sake of better readability, we omit $(\alpha_l, 1)$ and make clear through the context for which participation behavior the function $s_1(s_0)$ is defined.

Lemma 4 *For any participation behavior $(\alpha_l, 1)$ with $\alpha_l \in [0, 1]$ there exists a continuous, strictly increasing and onto function $s_1 : [0, \bar{s}_0] \rightarrow [\underline{s}_1, 1]$ with $\underline{s}_1 \in (0, 1)$, $\bar{s}_0 \in (0, 1)$ and $s_1'(s_0) = -\frac{\partial U_l^Y}{\partial s_0} / \frac{\partial U_l^Y}{\partial s_1}$ such that the participation constraint is binding for device (s_0, s_1) if and only if $s_1 = s_1(s_0)$.*

Since all devices $(s_0, s_1(s_0))$ implement the same participation behavior, the question arises how these devices differ. In the remainder of this subsection we study how the induced beliefs, the induced ad interim expected belief, the induced ad interim variance of belief and the induced ad interim expected utility of the agent depend on s_0 . Keep in mind that a change in s_0 is accompanied by a change in s_1 according to Lemma 4. In Figure 2 we illustrate the comparison results for a numerical example. Although the graphs are for a specific utility index and specific parameter values, the depicted properties hold fairly general. Figure 2(a) shows devices $(s_0, s_1(s_0))$ which implement equilibria with perfect pooling of private information ($\alpha_l = \alpha_h = 1$).

Consider first the effect of the device choice on beliefs μ_H^{II} and μ_L^{II} . By increasing s_0 and s_1 , it becomes more likely that the higher belief μ_H^{II} is generated conditional on the device being used, i.e. $p_{H|Y}$ increases. Moreover, high public signals become more likely for both, a good and a bad agent, and low public signals become less likely for both of them. Therefore, the effect on μ_H^{II} and μ_L^{II} is not obvious. We know that for all devices which induce the same participation behavior Bayes' Law dictates that the expected perception of an agent who uses the device is be the same, i.e. $(1 - p_{H|Y})\mu_L^{II} + p_{H|Y}\mu_H^{II} = \mu_Y = \text{constant}$ for all devices $(s_0, s_1(s_0))$. Since the probability with which the higher of the two beliefs is generated, $p_{H|Y}$, increases in s_0 , we can conclude that either μ_L^{II} or μ_H^{II} must decrease in s_0 . It turns out that both do (see Figure 2(b)). Note that for increasing s_0 the stigma of failure becomes more severe, i.e. it becomes more harmful to obtain a low public signal.

Lemma 5 *Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in [0, 1]$ and consider devices $(s_0, s_1(s_0))$. Then, $\frac{d\mu_L^{II}}{ds_0} < 0$ and $\frac{d\mu_H^{II}}{ds_0} < 0$.*

Consider now the effect of the device choice on ad interim expected beliefs \mathbb{E}_l and \mathbb{E}_h : From the point of view of an agent who already knows his private signal, the device induces a lottery with an expected belief which is (weakly) higher than μ_Y for an agent with a high private signal, $\mathbb{E}_h \in [\mu_Y, \frac{p_{h1}}{p_h}]$, and (strictly) lower than μ_Y for an agent with a low private signal, $\mathbb{E}_l \in [\frac{p_{l1}}{p_l}, \mu_Y]$. If the device is not perfect, $\mathbb{E}_l > \frac{p_{l1}}{p_l}$ and $\mathbb{E}_h < \frac{p_{h1}}{p_h}$. The size of the redistributive effect depends on which types of errors the device is susceptible to. For instance, in the example displayed in Figure 2(c), there is more redistribution when the probability of false positives increases, i.e. when s_0 is higher. The result that redistribution is most effective for devices with the largest possible s_0 generalizes to the case with arbitrary primitives $[p_1, p_h, \delta]$, any participation level $\alpha_l \in (0, 1]$ and any utility index satisfying

Assumption 1 $u'''(\mu) \geq 0$ for all μ .²⁰

²⁰This assumption holds, for instance, for quadratic utility and for CARA-preferences.

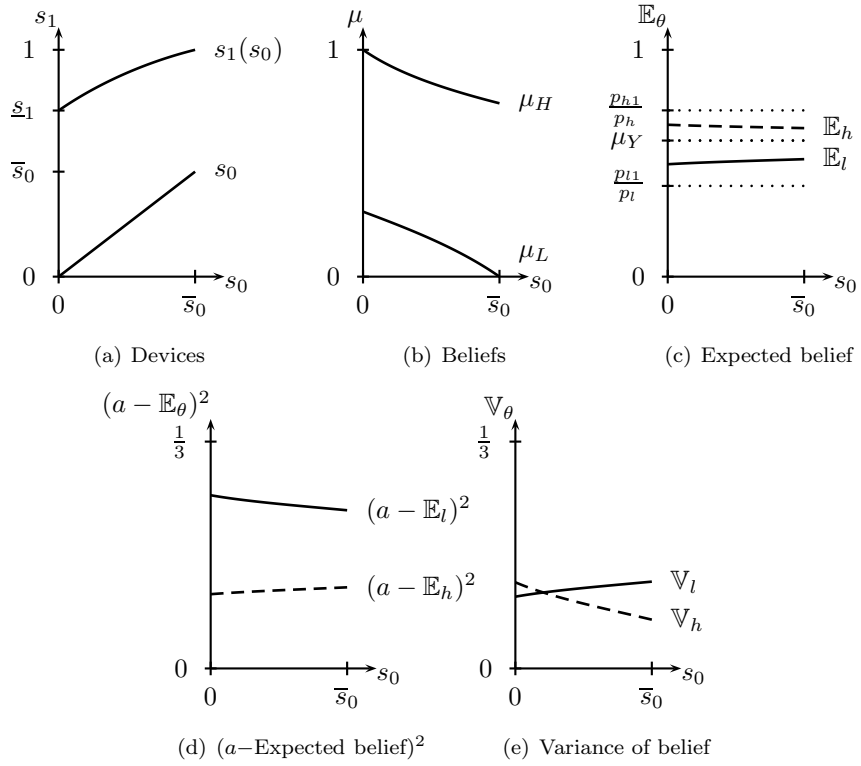


Figure 2: Devices for which the participation constraint is binding and properties thereof ($p_l = \frac{3}{5}$, $p_h = \frac{3}{5}$, $\delta = \frac{1}{3}$, $\alpha_l = \alpha_h = 1$, $u(\mu) := -(a - \mu)^2$, $a = 1$)

Lemma 6 Fix some participation behavior $(\alpha_l, 1)$ and consider devices $(s_0, s_1(s_0))$. If Assumption 1 holds and $\alpha_l \in (0, 1]$, $\frac{d\mathbb{E}_l}{ds_0} > 0$ and $\frac{d\mathbb{E}_h}{ds_0} < 0$. If Assumption 1 holds and $\alpha_l = 0$, $\frac{d\mathbb{E}_l}{ds_0} > 0$ and $\frac{d\mathbb{E}_h}{ds_0} = 0$.

Next, consider the effect of the device choice on the risk the agent incurs by using the device as measured by the variance of belief and consider for the moment the case where the agent's utility index is quadratic.^{21,22} Then his expected utility from using the device depends only on the induced expected belief (+) and the induced variance of belief (-):²³

$$U_\theta^Y = -(a - \mathbb{E}_\theta)^2 - \mathbb{V}_\theta \text{ with } a \geq 1. \quad (5)$$

How the variance incurred by an agent with a low private signal depends on the device choice is straightforward: By construction, $U_l^Y(\alpha_l, 1)$ is constant for all devices $(s_0, s_1(s_0))$. This is only possible if an increase in the expected belief, \mathbb{E}_l , comes along with an increase in the variance of belief, \mathbb{V}_l . Hence, $\frac{d\mathbb{E}_l}{ds_0} > 0$ implies $\frac{d\mathbb{V}_l}{ds_0} > 0$. Figure 2(e) shows that \mathbb{V}_l is increasing in our example. Figure 2(d) shows that changes in \mathbb{V}_l are accompanied by changes in $(a - \mathbb{E}_l)^2$ of equal size and different direction.

How the variance incurred by an agent with a high private signal depends on the device choice is more involved since there are two countervailing effects. To see these effects, it is helpful to make a decomposition of \mathbb{V}_h :²⁴

$$\mathbb{V}_h = \mathbb{V}_l + ([\mu_H^{II} + \mu_L^{II}] - [\mathbb{E}_l + \mathbb{E}_h])(\mathbb{E}_h - \mathbb{E}_l) \quad (6)$$

The first effect is an *informativeness effect*: By choosing a device $(s_0, s_1(s_0))$ with a higher s_0 , \mathbb{E}_l increases (Lemma 6). This makes the participation constraint less binding and allows the designer to make the device more informative which increases not only the variance incurred by an agent with a low private signal, but also that incurred by an agent with a high private signal suggesting that \mathbb{V}_l and \mathbb{V}_h move into the same direction. Hence, the informativeness effect is positive (see the first term on the right-hand side of (6)).

The second effect is a *distributional effect*. The lotteries over beliefs induced by devices $(s_0, s_1(s_0))$ may differ substantially for different values of s_0 . An example for this is displayed in Figure 3. Figure 3(a) is for the device which implements an equilibrium with perfect pooling of private information and which is least effective in redistributing expected belief, Figure 3(b) is for the device which implements the same participation behavior, but which is most effective

²¹We study the variance of belief because it helps us getting a better understanding for how the agent's expected utility depends on the device choice. Focussing on the quadratic case here is without loss of generality for our main results because we do not make use of the variance result later on.

²²Any quadratic utility index $u(\mu)$ which is strictly increasing and strictly convex on $(0, 1)$ is a positive linear transformation of $u(\mu) := -(a - \mu)^2$ with $a \geq 1$.

²³With $u(\mu) := -(a - \mu)^2$ we obtain $U_\theta^Y = -(1 - s_\theta)(a - \mu_L)^2 - s_\theta(a - \mu_H)^2 = -(1 - s_\theta)[(a - \mathbb{E}_\theta)^2 + 2(a - \mathbb{E}_\theta)(\mathbb{E}_\theta - \mu_L) + (\mathbb{E}_\theta - \mu_L)^2] - s_\theta[(a - \mathbb{E}_\theta)^2 + 2(a - \mathbb{E}_\theta)(\mathbb{E}_\theta - \mu_H) + (\mathbb{E}_\theta - \mu_H)^2] = -\mathbb{V}_\theta - (a - \mathbb{E}_\theta)^2$.

²⁴We have $\mathbb{V}_h = (1 - s_h)\mu_L^2 + s_h\mu_H^2 - \mathbb{E}_h^2 = (1 - s_l - [s_h - s_l])\mu_L^2 + (s_l + [s_h - s_l])\mu_H^2 - \mathbb{E}_l^2 + (\mathbb{E}_l^2 - \mathbb{E}_h^2) = \mathbb{V}_l + (s_h - s_l)(\mu_H^2 - \mu_L^2) + (\mathbb{E}_l^2 - \mathbb{E}_h^2)$. Using $\mathbb{E}_h - \mathbb{E}_l = [s_h - s_l](\mu_H - \mu_L)$, we obtain $\mathbb{V}_h = \mathbb{V}_l + ([\mu_H + \mu_L] - [\mathbb{E}_l + \mathbb{E}_h])(\mathbb{E}_h - \mathbb{E}_l)$.

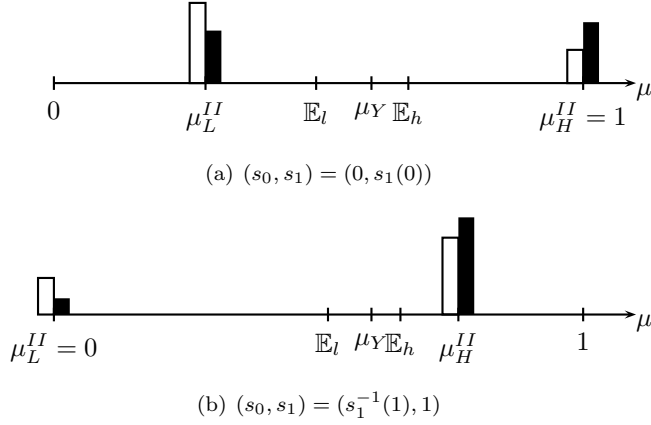


Figure 3: Probabilities and beliefs induced by two devices which both implement perfect pooling of private information ($p_l = \frac{3}{5}$, $p_h = \frac{3}{5}$, $\delta = \frac{1}{3}$, $\alpha_l = \alpha_h = 1$, $u(\mu) := -(a - \mu)^2$, $a = 1$)

in redistributing expected belief. The graphs show the two beliefs that may realize and the probabilities with which these beliefs occur for an agent with a low private signal (white bars) and an agent with a high private signal (black bars). It can be seen that while the ad interim expected beliefs \mathbb{E}_l and \mathbb{E}_h differ only slightly in the two cases, there are large changes in the beliefs that may actually realize, μ_L^{II} and μ_H^{II} (compare also Figures 2(b) and 2(c)). More importantly, the change in the belief lottery may affect the variance incurred by an agent with a high and with a low private signal differently. Note that \mathbb{V}_θ is a weighted sum of the two squared deviations $(\mu_H^{II} - \mathbb{E}_\theta)^2$ and $(\mu_L^{II} - \mathbb{E}_\theta)^2$, and note that the former deviation is relatively more likely in the case where the agent's private information is high, while the latter is relatively more likely in the case where the agent's private information is low. By increasing s_0 , both beliefs decrease (Lemma 5) such that the deviation which is relatively more likely when the agent's private signal is high, $(\mu_H^{II} - \mathbb{E}_\theta)^2$, decreases and the deviation which is relatively more likely in the case when the agent's signal is low, $(\mu_L^{II} - \mathbb{E}_\theta)^2$, increases. Hence, there might be a countervailing force moving \mathbb{V}_l and \mathbb{V}_h into different directions. To be a little bit more precise, look at the second term on the right-hand side of (6): $([\mu_H^{II} + \mu_L^{II}] - [\mathbb{E}_l + \mathbb{E}_h])(\mathbb{E}_h - \mathbb{E}_l)$. Since the change in beliefs is large relative to the change in expected beliefs (see Figure 3 again), how this term changes is driven by the change in the beliefs. As both beliefs decrease with s_0 , the distributional effect is negative.

Since there are two effects which move into different directions, the question arises which effect is stronger. Heuristically, \mathbb{E}_l differs only slightly for different values of s_0 such that the designer can make the device only slightly more informative when s_0 increases,²⁵ while how the

²⁵Somewhat more formally, we get that the informational effect is small relative to the distributional effect by noting that $U_l^Y = -(a - \mathbb{E}_l)^2 - \mathbb{V}_l = \text{constant}$ implies $\mathbb{V}_l = -(a - \mathbb{E}_l)^2 - \text{constant}$ for all devices $(s_0, s_1(s_0))$ such that the informational effect is small when the change in \mathbb{E}_l is small.

beliefs vary around the expected belief differs substantially for different s_0 . Hence, the negative distributional effect dominates the positive informativeness effect such that $\frac{d\mathbb{V}_h}{ds_0} < 0$ (see Figure 2(e) for a graphical illustration of \mathbb{V}_l and \mathbb{V}_h in our numerical example). For any quadratic utility index, any primitives $[p_1, p_h, \delta]$ and any participation level $\alpha_l \in [0, 1]$ we obtain the following result:

Lemma 7 *Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in [0, 1]$ and consider devices $(s_0, s_1(s_0))$. If the agent's utility is quadratic, $\frac{d\mathbb{V}_l}{ds_0} > 0$ and $\frac{d\mathbb{V}_h}{ds_0} < 0$.*

If the agent's utility is quadratic, the utility of an agent with high private information depends only on \mathbb{E}_h (+) and \mathbb{V}_h (-). We already know that for a higher s_0 the expected belief, \mathbb{E}_h , decreases because there is more redistribution of expected perception (Lemma 6), but that there is a countervailing effect because the risk he faces, \mathbb{V}_h , decreases as well (Lemma 7). By a similar reasoning as above, the effect of the expected belief is small relative to the effect of the variance such that his utility increases in s_0 . Consult again Figures 2(d) and 2(e) to see that the change in the variance indeed dominates in the example.

If the agent's utility is not quadratic, additional effects arise such that how the variance of an agent with a high private signal changes with s_0 becomes more involved. Interestingly, under relatively mild assumptions we obtain clear effects regarding how his expected utility changes with s_0 . For any primitives $[p_1, p_h, \delta]$, any participation level $\alpha_l \in [0, 1]$ and for any utility index satisfying Assumption 1 and

Assumption 2 $\left(\frac{u''(\mu)}{u'(\mu)}\right)^2 - \frac{u'''(\mu)}{u'(\mu)} > 0$ and $u''''(\mu) = 0$ for all μ

the utility of an agent with high private information increases with s_0 although he is on average perceived as worse. The expected utility of an agent with low private information is not affected by the device choice by construction. Hence, from an ex ante perspective, the agent is clearly better off for higher values of s_0 .

Lemma 8 *Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in [0, 1]$ and consider devices $(s_0, s_1(s_0))$. If Assumptions 1 and 2 hold, $\frac{dU_h^Y}{ds_0} > 0$ and $\frac{dU_l^Y}{ds_0} = 0$.*

The role of Assumptions 1 and 2 is to limit the importance of higher order moments relative to the expected value and the variance.²⁶ This is the case which we consider particularly interesting. If higher order moments have a strong impact on the agent's utility, other effects than those described in Lemma 8 may arise.²⁷ Note further that Assumptions 1 and 2 are sufficient for the result in Lemma 8 but not necessary. For instance, for CARA-preferences with arbitrary risk parameter, we have $\left(\frac{u''(\mu)}{u'(\mu)}\right)^2 - \frac{u'''(\mu)}{u'(\mu)} = 0$ and higher order derivatives do not vanish. Still, we obtain:

²⁶Assumption 1 defines a lower bound on $\frac{u'''(\mu)}{u'(\mu)}$ while Assumption 2 defines an upper bound on $\frac{u'''(\mu)}{u'(\mu)}$ for the cubic case.

²⁷For instance, if we assume that $\left(\frac{u''(\mu)}{u'(\mu)}\right)^2 - \frac{2}{3} \frac{u'''(\mu)}{u'(\mu)} < 0$ and $u''''(\mu) = 0$, we obtain $\frac{dU_h^Y}{ds_0} < 0$.

Lemma 9 Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in [0, 1]$ and consider devices $(s_0, s_1(s_0))$. If the agent has CARA-preferences with an arbitrary risk parameter, we obtain $\frac{dU_h^Y}{ds_0} > 0$.

Before we continue our analysis with the second type of equilibrium, we would like to stress that the results obtained so far do neither depend on the primitives $[p_1, p_h, \delta]$ nor on which particular participation behavior (with $\alpha_h = 1$) we consider.

6.2 Deterring participation of an agent with a low private signal

Suppose $\alpha_l = 0$ and $\alpha_h \in [0, 1]$. The problem to incentivize an agent with a high private signal to use the device is the following: If he does not use the device, the designer infers that he has a low private signal with a positive probability and thus perceives him as worse than he actually is in expected terms, i.e. $\text{Prob}(\omega = 1 | d = N, \theta = h) = \mu_N^I < \frac{p_h 1}{p_h}$. If he uses the device, the designer infers correctly that he has a high private signal, i.e. $\text{Prob}(\omega = 1 | d = Y, \theta = h) = \frac{p_h 1}{p_h}$. However the agent incurs a risk as the device generates additional information about his type. If this risk is too large, he is better off not using the device. By making the device susceptible to errors, the information generated by the device becomes less precise, insuring the agent (partially) against his type-risk. Since $\alpha_l = 0$ remains true, the agent's expected perception is not affected. As a consequence, the designer can increase s_0 and/or decrease s_1 to incentivize the agent to use the device. We obtain the following result which is analogous to Lemma 3.

Lemma 10 For any $\alpha_h \in [0, 1]$, $U_h^Y(0, \alpha_h)$ is differentiable in the device (s_0, s_1) with $\frac{\partial}{\partial s_0} U_h^Y(0, \alpha_h) > 0$ and $\frac{\partial}{\partial s_1} U_h^Y(0, \alpha_h) < 0$.

Whether an agent with a high private signal is very eager to use the device or not depends on how much worse he is perceived in expected terms when he does not use the device. An indicator for this incentive is $\frac{p_h 1}{p_h} - \mu_N^I = \delta \frac{1 - p_h}{1 - \alpha_h p_h}$. For given parameters p_1 and p_h , he is very eager to separate when his private signal is very informative about his type, i.e. when δ is large. There exists a threshold $\hat{\delta} \in (0, \delta_{\max})$ such that for values of δ above this threshold there exists an equilibrium where an agent with a high private signal is even willing to use the perfect device (with probability one).²⁸ For values of δ below this threshold, the agent is not willing to use the perfect device with a positive probability. In this case there exists a unique equilibrium in which the perfect device is never used. This is stated in the following lemma which holds for any primitives $[p_1, p_h, \delta]$ and any utility index $u(\cdot)$.

Lemma 11 Consider the perfect device $(s_0, s_1) = (0, 1)$. For any $p_1, p_h \in (0, 1)$ there exists a $\hat{\delta}(p_1, p_h) \in (0, \delta_{\max})$ such that the following two properties hold: (i) If $\delta < \hat{\delta}(p_1, p_h)$, then there exists no equilibrium of type I. (ii) If $\delta \geq \hat{\delta}(p_1, p_h)$, then there exists an equilibrium of type Ib, i.e. one in which the agent uses the perfect device if and only if his private signal is high.

Consider $\delta < \hat{\delta}(p_1, p_h)$. In this case the agent is not willing to use the perfect device even if his private signal is high. As $\alpha_l = 0$, he has a strict incentive to use an uninformative device

²⁸This equilibrium may coexist with equilibria where the perfect device is used with a smaller probability.

as this allows him to signal superior private information without incurring a risk regarding how he will be perceived. From this and a continuity property we obtain by an Intermediate Value Theorem that any participation behavior $(0, \alpha_h)$ with $\alpha_h \in [0, 1]$ can be induced by some device $(s_0, s_1) = (0, s'_1)$. However, there are, again, multiple devices which induce any specific participation behavior. We obtain a characterization of the set of all devices which induce a certain participation behavior which is analogous to that in Lemma 4.

Lemma 12 *Let $\delta \in (0, \widehat{\delta}(p_1, p_h))$. For any participation behavior $(0, \alpha_h)$ with $\alpha_h \in [0, 1]$ there exists a continuous, strictly increasing and onto function $s_1 : [0, \bar{s}_0] \rightarrow [\underline{s}_1, 1]$ with $\underline{s}_1 := s_1(0) \in (0, 1)$, $\bar{s}_0 := s_1^{-1}(1) \in (0, 1)$ and $s'_1(s_0) = -\frac{\partial U_l^x}{\partial s_0} / \frac{\partial U_h^x}{\partial s_1}$ such that the participation constraint is binding for device (s_0, s_1) if and only if $s_1 = s_1(s_0)$.*

Consider $\delta < \widehat{\delta}(p_1, p_h)$ and consider any participation behavior $(0, \alpha_h)$ with $\alpha_h \in [0, 1]$. Then the designer is able to make the participation constraint binding implying that an agent with a high private signal obtains the same utility from using and from not using the device and thus also the same utility as an agent with a low private signal. Consider now $\delta > \widehat{\delta}(p_1, p_h)$ and participation behavior $(0, 1)$. Then the designer is not able to make the participation constraint binding as the agent has a strict incentive to use even the most informative device when his private signal is high. It follows that when the agent's private signal is high, he must be strictly better off when he uses the device than when he does not use it.

Lemma 13 *(i) Let $\delta \in (0, \widehat{\delta}(p_1, p_h))$, fix some participation behavior $(0, \alpha_h)$ with $\alpha_h \in [0, 1]$ and consider devices $(s_0, s_1(s_0))$. Then $U_l(0, \alpha_h) = U_h(0, \alpha_h) = U^N(0, \alpha_h)$ for any α_h . (ii) Let $\delta \in (\widehat{\delta}(p_1, p_h), \delta_{\max})$ and consider participation behavior $(0, 1)$ and device $(s_0, s_1) = (0, 1)$. Then $U_l(0, 1) = U^N(0, 1)$ and $U_h(0, 1) > U^N(0, 1)$.*

7 The optimal device

In this section we analyze the design of the optimal device. We start with a discussion of the designer's objective in Subsection 7.1. In Subsection 7.2 we assume that the designer wants to implement an equilibrium where the agent also uses the device when his private signal is low and in Subsection 7.3 we assume that she wants to implement one where he only uses it when his private signal is high. In both subsections we first take the participation behavior as given and look for the optimal device among those inducing this specific participation behavior. Next, we identify the optimal device inducing participation that belongs to the respective class. Since we then already know the optimal device for a given participation behavior, this is equivalent to identifying the optimal participation behavior in the respective class. Finally, we compare the two remaining devices to obtain the generally optimal device/participation behavior.

7.1 The designer’s objective

Assume for a moment that the designer’s and the agent’s utility index are both quadratic. Then, from an ex ante perspective, the designer’s and the agent’s expected utility depend only on the ex ante expected belief and the ex ante variance of belief. Due to Bayes’ Law, the ex ante expected belief is always equal to the prior p_1 such that the device choice affects only the ex ante variance of belief. Roughly speaking, a change in the device which increases the ex ante variance means better learning about the agent’s type. Hence, it increases the designer’s ex ante expected utility while it decreases the ex ante expected utility of the risk-averse agent. It follows that the designer’s optimal device choice minimizes the ex ante expected utility of the agent.

Lemma 14 *If the designer’s and the agent’s utility index are both quadratic, maximizing $\mathbf{E}[v(\mu)]$ is equivalent to minimizing $\mathbf{E}[u(\mu)]$.*

We are mainly interested in the optimal device choice for the case where decisions depend on the first two moments of the belief distribution. For this case we do not need to impose more structure on the utility indices than we imposed in the model section: the agent’s index is any quadratic function which is increasing and concave and the designer’s index is any quadratic function which is convex. All our subsequent results apply to the quadratic case and thus also to the non-reduced problem introduced in Subsection 3.4. However, we prove the subsequent results for a slightly more general case. This allows us, for instance, to extend our results to the CARA case when the designer’s preferences have a similar shape as the agent’s preferences. We adopt the assumption

Assumption 3 $v(\mu) := -C_2u(\mu) + C_1\mu + C_0$ with $C_0, C_1 \in \mathbb{R}$ and $C_2 > 0$

which allows that higher moments of the belief distribution do matter, but which restricts the “magnitude” and the “direction” of their effect on the designer’s utility. It directly follows

$$V = -C_2\mathbf{E}[u(\mu)] + C_1p_1 + C_0, \tag{7}$$

implying that the designer’s ex ante expected utility is maximized when the agent’s ex ante expected utility is minimized.

7.2 The optimal device inducing participation of an agent with a low private signal

Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in (0, 1]$. Because of Assumption 3, the designer learns most when she minimizes the agent’s ex ante expected utility. Since it cannot be optimal for her to have the participation constraint slack and since the agent’s outside option does not depend on the device choice, it follows that when the agent’s private signal is low, he obtains the same utility from using and from not using the device, and that this utility is the same for all devices $(s_0, s_1(s_0))$. Hence, the designer learns most when she minimizes the utility of the agent

in the case where his private signal is high. Her optimization problem becomes

$$\begin{aligned} \min_{s_0 \in [0, \bar{s}_0]} \quad & U_h^Y(\alpha_l, 1) \\ \text{s.t.} \quad & s_1 = s_1(s_0) \text{ (as defined in Lemma 4).} \end{aligned}$$

Under Assumptions 1 and 2, $U_h^Y(\alpha_l, 1)$ is minimized by device $(0, s_1(0))$ (Lemma 8), i.e. by a device which is subject to false negatives but not to false positives. Among all the devices for which the participation constraint is binding, this is the device which redistributes least expected belief from the case where the agent's private signal is high to the case where it is low (Lemma 6), but for which the stigma of failure is least severe (Lemma 5). The intuition that redistribution makes the participation constraint less binding and thus allows for better learning about the agent's type is wrong. While redistribution makes the participation constraint indeed slightly less binding, it is accompanied by a second, stronger effect which makes the inference that can be drawn from the generated signals less accurate. This second effect is basically the distributional effect motivated in Subsection 6.1.

Proposition 1 *Fix some participation behavior $(\alpha_l, 1)$ with $\alpha_l \in (0, 1]$. (i) It is optimal for the designer to have the participation constraint binding. (ii) If Assumptions 1, 2 and 3 hold, $s_0 = 0$ is optimal for the designer.*

Consider now the choice of the participation behavior α_l given the optimal device, $(s_0, s_1) = (0, s_1(0))$. Since the participation constraint is binding and since the utility from not using the device does not depend on α_l for the considered type of participation behavior, the agent's ex ante expected utility is again minimized by minimizing the utility of the agent in the case where his private signal is high:

$$\begin{aligned} \min_{\alpha_l \in (0, 1]} \quad & U_h^X(\alpha_l, 1) \\ \text{s.t.} \quad & s_0 = 0, s_1 = s_1(0) \text{ (as defined in Lemma 4).} \end{aligned}$$

Implementing an equilibrium with a higher α_l causes three effects: First, to incentivize the agent to use the device, the designer has to choose a less informative device, i.e. $s_1(0)$ decreases in α_l (-).²⁹ Second, the designer learns the agent's private information less accurately (-).³⁰ Third, a public signal is generated with a higher probability (+). We can show that under Assumption 3 the third (positive) effect dominates the first two (negative) effects. This means that if the designer wants to implement an equilibrium which is used with positive probability when the agent's private signal is low, then it is optimal for her to choose a device which is always used.

²⁹Suppose that the same device is associated with a higher participation level. As it becomes then relatively more likely that the agent's private signal is low conditional on the device being used, it becomes also relatively more likely that the agent's type is bad conditional on the device being used. For a given device, this does not affect the probability of getting a certain public signal but it is detrimental for the associated beliefs such that participation becomes less attractive. Hence, to induce a higher participation level, the designer has to make the device less informative by moving s_1 closer to $s_0 = 0$ (see Lemma 3).

³⁰From non-participation she can still infer that the agent's private signal is low, but what she can infer from participation becomes less accurate.

The reason for this result is the following: A higher level of participation, α_l , can only be induced by reducing the stigma of failure.³¹ Reducing the stigma of failure requires that the agent gets a low public signal more often when he is actually good. Since this hurts the agent the more the more likely it is that he is good, decreasing s_1 in a way that makes the agent indifferent when his private signal is low makes him worse off when his private signal turns out to be high. Under Assumption 3 this means better learning about the agent's type and thus dominance of the third effect over the first two effects.

Proposition 2 *If Assumptions 1, 2 and 3 hold and the designer wants to implement an equilibrium where the agent uses the device with positive probability when his private signal is low, then it is optimal for her to implement an equilibrium where the agent always uses the device, i.e. an equilibrium with perfect pooling of private information.*

Consider a test which produces pass–fail–results as an implementation of the optimal device. The property $s_0 = 0$ of the optimal test means that the bad type never passes the test, $s_1 < 1$ means that the good type fails with a positive probability. The lower s_1 , the harder the test is to pass. Interestingly, higher participation is optimally induced by making the test harder to pass.

7.3 The optimal device deterring participation of an agent with a low private signal

The best that the designer can possibly achieve within the class of equilibria where the device is not used by an agent with a low private signal, is learning the agent's private signal perfectly, and, if the agent's private signal turns out to be high, learning the agent's type perfectly. If the agent's private information is sufficiently informative, his interest in signaling superior information is strong enough such that this is actually feasible.

Proposition 3 *If $\delta \geq \widehat{\delta}(p_l, p_h)$, Assumption 3 holds and the designer wants to implement an equilibrium in which the device is not used when the agent's private information is low, then it is optimal for the designer to implement an equilibrium where perfect separation of private information is induced by the perfect device.*

Consider now $\delta < \widehat{\delta}(p_l, p_h)$ and fix some participation behavior $(0, \alpha_h)$ with $\alpha_h \in (0, 1]$. Again, because of Assumption 3, the designer strives for minimizing the agent's (ex ante) expected utility. In the considered case it is possible (by construction of $\widehat{\delta}(p_l, p_h)$) and optimal for the designer to make the agent's participation constraint binding. If the agent's private signal is low, he does not use the device and thus obtains utility $U^N(0, \alpha_h)$. If the agent's private utility is high, he uses the device with positive probability, but he is indifferent between using and not

³¹We argued in Footnote 29 that s_1 must decrease in response to an increase in α_l . It follows that the probability of getting a high public signal decreases for an agent with either kind of private signal ($\frac{ds_h}{d\alpha_l} < 0$). Since $\mu_H = 1$ is true for any participation level, the participation constraint can only remain to be satisfied when the stigma of failure becomes less severe ($\frac{d\mu_L^I}{d\alpha_l} > 0$).

using it. Hence, he also obtains $U^N(0, \alpha_h)$. From this it follows that the designer's problem simplifies as follows:

$$\begin{aligned} \min_{s_0 \in [0, \bar{s}_0]} \quad & U^N(0, \alpha_h) \\ \text{s.t.} \quad & s_1 = s_1(s_0) \text{ (as defined in Lemma 12)}. \end{aligned}$$

Since $U^N(0, \alpha_h)$ is not affected by the device choice, any device for which the participation constraint is binding is optimal.

Proposition 4 *Let $\delta < \widehat{\delta}(p_1, p_h)$ and fix some participation behavior $(0, \alpha_h)$ with $\alpha_h \in [0, 1]$. If Assumption 3 holds, then it is optimal for the designer to have the participation constraint binding and any device which induces participation behavior $(0, \alpha_h)$ with a binding participation constraint is optimal.*

Note that here $s_0 = 0$ is only weakly optimal. However, weak optimality is a peculiarity of the participation constraint being binding for the highest possible private signal. We show in Subsection 8.2 that with continuous private information $s_0 = 0$ is strictly optimal for any participation behavior for which the device is used with positive probability. Therefore the case where a device which is subject to false negatives but not to false positives ($s_0 = 0$ and $s_1 \leq 1$) is strictly optimal can be seen as the “general case”.

Consider now the optimal participation behavior α_h given an optimal device which induces this participation behavior, for instance, given device $(s_0, s_1) = (0, s_1(0))$. Since the participation constraint is binding, the designer's problem simplifies to

$$\begin{aligned} \min_{\alpha_h \in (0, 1]} \quad & U^N(0, \alpha_h) \\ \text{s.t.} \quad & s_0 = 0, s_1 = s_1(0) \text{ (as defined in Lemma 12)}. \end{aligned}$$

$U^N(0, \alpha_h)$ does not depend on the device choice. Hence, we can ignore the constraint and have to consider only how α_h affects $U^N(0, \alpha_h)$. Since the utility from not using the device becomes worse when the agent uses the device with a higher probability when his private signal is high, $\alpha_h = 1$ is optimal.

Looking at the non-simplified problem again, we obtain the following heuristics for the result: Implementing an equilibrium with a higher α_h causes three effects. The designer better learns the agent's private information (+), he observes a public signal with a higher probability (+), and he can make the device more informative as a higher participation level makes the agent's outside option worse and thus participation more attractive (+). Since all three effects are positive, $\alpha_h = 1$ must be optimal.

Proposition 5 *Let $\delta < \widehat{\delta}(p_1, p_h)$. If Assumption 3 holds and the designer wants to implement an equilibrium where the agent does not use the device when his private signal is low, then it is optimal for her to implement an equilibrium where the agent uses the device if and only if his private signal is high, i.e. to implement an equilibrium with perfect separation of private information.*

7.4 The optimal device: perfect pooling versus perfect separation of private information

To find the optimal device for the designer, we need to compare only two devices: the device $(0, s_1^p(0))$ which implements an equilibrium with perfect pooling of private information and the device $(0, s_1^s(0))$ which implements an equilibrium with perfect separation of private information.^{32,33} The designer faces a trade-off between learning more “newly generated” information through the device and learning also “already existing” private information through self-selection into the device. In this subsection we investigate this trade-off.

Our main result can be summarized as follows: If the agent’s private signal is not very informative about his type, then the designer wants to learn the agent’s private information, whereas if the private signal is very informative, then it might be optimal for the designer to forego learning this information and to induce full participation instead. At a first glance, this result appears paradoxical: For more informative private information the designer should have a stronger interest in learning the agent’s private information and not vice versa. However, the informativeness of the private information also determines how informative the device can be. Because of this additional effect, it is not obvious how more accurate private information affects the designer’s interest in inducing an equilibrium in which she can infer private information. To see what drives the result, it is helpful to have a closer look at the agent’s utility in both types of equilibrium.

Remember that with respect to information generation, the designer’s interest is opposed to the agent’s. Hence, the worse off the agent is, the better off is the designer, i.e. the better is information generation. In any case, the optimal device leaves the agent with low private information with expected utility $u(\frac{p_l}{p_l})$. Pooling allows the agent with high private information to earn a “rent” which can be extracted by the designer in the case of perfect separation. For $\delta < \widehat{\delta}(p_l, p_h)$, his incentive to signal superior information can be fully exploited by the designer. Hence, perfect separation is better than perfect pooling for low δ . When δ reaches $\widehat{\delta}(p_l, p_h)$, the agent is even willing to use a perfect device to signal high private information. For even higher δ , the designer cannot fully exploit the agent’s incentive to signal superior information anymore because the informativeness of the device cannot be increased beyond perfection. Thus, in both types of equilibria, perfect pooling and perfect separation of information, the agent obtains a rent if he has high private information. For δ slightly above $\widehat{\delta}(p_l, p_h)$ the rent earned in the pooling case is still larger than that in the separating case, but if δ moves closer to δ_{\max} , the rent in the separating case may actually become larger than the rent in the pooling case. Hence, for very informative private information, perfect pooling may eventually dominate perfect separation. If

³²We have $s_1^p(0) := s_1(0)$ with $s_1(0)$ as defined in Lemma 4, $s_1^s(0) := s_1(0)$ for $\delta < \widehat{\delta}(p_l, p_h)$ with $s_1(0)$ as defined in Lemma 12 and $s_1^s(0) := 1$ for $\delta \geq \widehat{\delta}(p_l, p_h)$.

³³Remember that if $\delta < \widehat{\delta}(p_l, p_h)$, then the designer is indifferent between all devices which implement perfect separation of private information. Hence, if $\delta < \widehat{\delta}(p_l, p_h)$ and the device $(0, s_1^s(0))$ is optimal, then it is not uniquely optimal. Otherwise, the optimal device is unique.

this happens or not depends on the primitives of the model.³⁴ We conclude this section by stating our main result formally:

Proposition 6 *Let Assumptions 1, 2 and 3 be satisfied. (i) If $\delta < \widehat{\delta}(p_1, p_h)$, then it is optimal for the designer to induce perfect separation of private information. (ii) If $\delta \geq \widehat{\delta}(p_1, p_h)$, it is either optimal for the designer to induce perfect separation or perfect pooling of private information.*

8 Extentions

One may wonder in how far our previous results hinge upon our binary structure. In this section we show that our main results prevail when we allow for more general devices and/or continuous private information.

8.1 More than two public signals

In this subsection we explain why it suffices to restrict attention to devices which, if used, generate only two different public signals. We show that a device which induces three different beliefs cannot be optimal as the designer always has an incentive to change the device in a way such that only two different beliefs do arise. This proves that there is no need for devices which are capable of generating three different public signals. The same argument can be applied to the case with any finite number of public signals showing that a two-signal-device is optimal in the class of devices which allows for any finite number of public signals.

The model with three public signals. The device generates a public signal $\sigma' \in \{L, M, H\}$ and is characterized by a vector $(s_{L|0}, s_{M|0}, s_{H|0}; s_{L|1}, s_{M|1}, s_{H|1})$ with $s_{L|0} + s_{M|0} + s_{H|0} = 1$ and $s_{L|1} + s_{M|1} + s_{H|1} = 1$. $s_{\sigma'|\omega}$ denotes the probability with which public signal σ' is generated when the agent's type is ω . If an agent with private information θ uses the device, he obtains public signal σ' with probability $s_{\sigma'|\theta} := \frac{p_{\theta 0}}{p_\theta} s_{\sigma'|0} + \frac{p_{\theta 1}}{p_\theta} s_{\sigma'|1}$. The belief assigned to public signal σ' is $\mu_{\sigma'} := \frac{\mu_Y s_{\sigma'|1}}{p_{\sigma'|Y}}$ with $p_{\sigma'|Y} := \mu_Y s_{\sigma'|1} + (1 - \mu_Y) s_{\sigma'|0}$. $p_{\sigma'|Y}$ is the probability that public signal σ' arises conditional on the device being used.

Sketch of proof. Consider a fixed participation behavior with $\alpha_l > 0$ and $\alpha_h = 1$. Under Assumption 3, this is the only case where a device which generates three public signals may strictly outperform any device which generates only two public signals.³⁵ Moreover, recall that

³⁴Consider $u(\mu) = -(1 - \mu)^2$ and any $v(\mu)$ which satisfies Assumption 3. If $p_1 = \frac{3}{4}$, $p_h = \frac{1}{4}$ and δ is sufficiently close to $\delta_{\max} = \frac{1}{3}$, then the designer prefers the equilibrium with perfect pooling of private information to that with perfect separation of private information. If $p_1 = \frac{1}{4}$, $p_h = \frac{3}{4}$, then the designer prefers for any $\delta \in (0, \delta_{\max})$ the equilibrium with perfect separation of private information.

³⁵For any participation behavior with $\alpha_l = 0$, there are only two possibilities: First, a perfect device is optimal. This requires only the generation of two different public signals. Second, any device which satisfies the participation constraint is optimal. Since the participation constraint can always be satisfied with a device that generates only two public signals when it can be satisfied with a device which generates more than two public signals, there exists a device which is optimal and which generates only two public signals.

under Assumption 3 the designer strives for minimizing U_h^Y and it cannot be optimal for her to have the participation constraint of an agent with a low private signal slack.

Assume now to the contrary that (i) the device $(s_{L|0}, s_{M|0}, s_{H|0}; s_{L|1}, s_{M|1}, s_{H|1})$ is optimal, and (ii) no device which generates only two public signals is optimal.

Because of (ii), we can assume without loss of generality that $\mu_L < \mu_M < \mu_H$ and $p_{\sigma'|Y} > 0$.³⁶

Fix now the parts of the device which are responsible for the high signal $\sigma' = H$, i.e. $s_{H|0}$ and $s_{H|1}$. Then the designer has only two degrees of freedom left, say $s_0 := s_{M|0}$ and $s_1 := s_{M|1}$. Using notation $\beta_0 := 1 - s_{H|0}$ and $\beta_1 := 1 - s_{H|1}$, we have $s_{L|0} = \beta_0 - s_0$ and $s_{L|1} = \beta_1 - s_1$. This implies $s_{H|\theta} = \frac{p_{\theta 0}}{p_\theta}(1 - \beta_0) + \frac{p_{\theta 1}}{p_\theta}(1 - \beta_1)$, $s_{M|\theta} = \frac{p_{\theta 0}}{p_\theta}s_0 + \frac{p_{\theta 1}}{p_\theta}s_1$ and $s_{L|\theta} = \frac{p_{\theta 0}}{p_\theta}(\beta_0 - s_0) + \frac{p_{\theta 1}}{p_\theta}(\beta_1 - s_1)$ and $U_\theta^Y := s_{L|\theta}u(\mu_L) + s_{M|\theta}u(\mu_M) + s_{H|\theta}u(\mu_H)$. Note that neither $s_{H|\theta}$ nor μ_H depend on s_0 or s_1 such that $s_{H|\theta}u(\mu_H)$ is constant with respect to s_0 and s_1 .

Condition (i) requires that s_0 and s_1 solve a problem with the following structure:

$$\begin{aligned} \min_{\substack{s_0 \in [0, \beta_0] \\ s_1 \in [0, \beta_1]}} \quad & s_{L|h}u(\mu_L) + s_{M|h}u(\mu_M) + \text{constant} \\ \text{s.t.} \quad & s_{L|l}u(\mu_L) + s_{M|l}u(\mu_M) + \text{constant} = \text{constant} \end{aligned}$$

The structure of this problem is very similar to the structure of the optimal device problem in the case where only devices which generate two different public signals are feasible. The main difference is that we do not have $\beta_0 = \beta_1 = 1$, but $\beta_1 < 1$, $\beta_0 \leq 1$. I.e., there are stricter and possibly asymmetric upper bounds on s_0 and s_1 . It follows that if $s_0 = \beta_0$ or $s_1 = \beta_1$, it is not feasible to change the device in a way which increases s_0 and s_1 . However, using similar arguments as those used in the proof to Lemma 8 and Proposition 1, we obtain that for any $s_0 > 0$ the designer profits from decreasing s_0 and s_1 in a way such that the participation constraint continues to hold. Since such a modification is always feasible, only devices with $s_0 = 0$ may be consistent with (i). However, since $s_0 = 0$ implies $\mu_M = 1$, this can not be consistent with the ordering of beliefs induced by (ii). Contradiction.

More generally, let any device which generates a finite number $X \in \mathbb{N}$ of different public signals and which induces X different beliefs with $\mu_1 < \mu_2 < \dots < \mu_X$ be given. Then, for any $i, j \in \mathbb{N}$ with $i < j < X$ there exists a device for which μ_i and μ_j become slightly larger and for which all other beliefs remain unchanged, which is strictly preferred by the designer. Hence, a device can only be optimal if it induces less than two public signals that are smaller than μ_X . As nothing can be learned from a device which generates only one public signal, a device which generates two public signals is optimal.

Using the above strategy of proof, we can establish the following proposition:³⁷

Proposition 7 *Let Assumptions 1, 2 and 3 be satisfied and suppose devices may generate any finite number of public signals. Then the optimal device induces only two different beliefs and*

³⁶If two of the induced beliefs were equal or if one belief did not occur with positive probability, the same lottery over beliefs could be induced by a device which generates only two different public signals. If the ordering of beliefs was different, it could be restored by relabeling.

³⁷A rigorous proof can be obtained from the authors upon request.

can thus be implemented by a device which generates two different public signals.

8.2 Continuous private information

In this subsection we assume that the agent’s private information θ is continuously distributed on $[0, 1]$, while we maintain the assumption that his true type is binary, i.e. $\omega \in \{0, 1\}$ and that a higher θ is associated with a higher probability that $\omega = 1$. We intend to show that our result that the optimal device is not subject to false positives is indeed general.

While equilibria in the discrete setting are characterized by a mixed participation behavior, equilibria in the continuous setting are characterized by a threshold $\tilde{\theta}$ that separates the signals for which the agent participates and those for which he does not.³⁸ We are interested in the construction of the optimal device for a given participation behavior. Hence, we can consider $\tilde{\theta}$ as a fixed parameter.

In both, the binary and the continuous version of the model, only a single participation constraint is relevant for the design problem: the constraint for the lowest private signal with which the agent is supposed to use the device. The result that a device with $s_0 = 0$ is strictly optimal to induce a certain participation behavior $(\alpha_l, 1)$ is driven by the fact that the average participant is better than the marginal participant. In the continuous version of the model this is true for any $\tilde{\theta} < 1$. In the proof given below we show that this property does indeed suffice to show that the optimal device problem in the continuous case can be transformed into a problem that can be treated as the problem in the binary case. Hence, strict optimality of $s_0 = 0$ carries over to the continuous version of the model for any participation threshold $\tilde{\theta} < 1$. The designer’s indifference between all the devices which induce a certain participation behavior $(0, \alpha_h)$ stems from the fact that the average participant coincides with the marginal participant. In the continuous version this is only the case for $\tilde{\theta} = 1$. Since for $\tilde{\theta} = 1$ the device is used with probability zero, inducing such a participation behavior cannot be optimal for the designer. Hence, the possibility that a device with $s_0 > 0$ might be optimal is a peculiarity of the binary version of the model. Therefore $s_0 = 0$ being strictly optimal can be seen as the “general” result. In the remainder of this subsection we present the proof for the case with $\tilde{\theta} \in (0, 1)$.

The problem with continuous private information. We have to adapt some of the notation introduced in Section 3 to the continuous distribution of private information. In particular, we have to redefine the joint distribution of information and type. It will be convenient to choose a notation which is similar to that in the discrete version of the model, but to indicate parameters and variables which are affected by the changes in the distribution by hats.³⁹

The distribution is completely specified by the unconditional probability that the agent’s type is good, $\text{Prob}(\omega = 1) =: \hat{p}_1$, and the density functions of the agent’s private information

³⁸Since—as in the binary setting—an agent has a strictly higher incentive to use an informative device when his private information is higher, threshold strategies are the only strategies which can be incentive compatible.

³⁹For instance, we will use the hat-notation to describe the distribution or the beliefs as they are derived from the distribution, but we can use the old notation to describe the device, the public signals or the utility indices.

conditional on the agent's type which we denote by $\widehat{p}(\theta|\omega)$. We assume that $\widehat{p}(\theta|\omega)$ is continuously differentiable and strictly positive on its support. From these definitions it follows that the joint density of θ and ω is $\widehat{p}_\omega \widehat{p}(\theta|\omega) =: \widehat{p}(\theta, \omega)$ and that the marginal density of θ is $\widehat{p}_0 \widehat{p}(\theta|\omega = 0) + \widehat{p}_1 \widehat{p}(\theta|\omega = 1) =: \widehat{p}(\theta)$ with $\widehat{p}_0 := 1 - \widehat{p}_1$. We assume that $\text{Prob}(\omega = 1|\theta) = \frac{\widehat{p}(\theta, \omega=1)}{\widehat{p}(\theta)}$ is increasing in θ . This is analogous to assuming that there is a positive correlation between the agent's private information and his type in the version of the model with binary private information.

Now we have everything at hand to describe the probability that an agent with private information $\theta \in [0, 1]$ obtains a high public signal from a given device (s_0, s_1) ,

$$\widehat{s}_\theta := \left(1 - \frac{\widehat{p}(\theta, \omega = 1)}{\widehat{p}(\theta)}\right) s_0 + \frac{\widehat{p}(\theta, \omega = 1)}{\widehat{p}(\theta)} s_1,$$

and the beliefs which may realize:

$$\begin{aligned} \widehat{\mu}_N &:= \frac{\text{Prob}(\theta < \widetilde{\theta}, \omega = 1)}{\text{Prob}(\theta < \widetilde{\theta})} = \frac{\int_0^{\widetilde{\theta}} \widehat{p}(\theta, \omega = 1) d\theta}{\int_0^{\widetilde{\theta}} \widehat{p}(\theta) d\theta} \\ \widehat{\mu}_Y &:= \frac{\text{Prob}(\theta \geq \widetilde{\theta}, \omega = 1)}{\text{Prob}(\theta \geq \widetilde{\theta})} = \frac{\int_{\widetilde{\theta}}^1 \widehat{p}(\theta, \omega = 1) d\theta}{\int_{\widetilde{\theta}}^1 \widehat{p}(\theta) d\theta} \\ \widehat{\mu}_L &:= \frac{\text{Prob}(\theta \geq \widetilde{\theta}, \omega = 1, \sigma = L)}{\text{Prob}(\theta \geq \widetilde{\theta}, \sigma = L)} = \frac{\widehat{\mu}_Y (1 - s_1)}{\widehat{\mu}_Y (1 - s_1) + (1 - \widehat{\mu}_Y) (1 - s_0)} \\ \widehat{\mu}_H &:= \frac{\text{Prob}(\theta \geq \widetilde{\theta}, \omega = 1, \sigma = H)}{\text{Prob}(\theta \geq \widetilde{\theta}, \sigma = H)} = \frac{\widehat{\mu}_Y s_1}{\widehat{\mu}_Y s_1 + (1 - \widehat{\mu}_Y) s_0} \end{aligned}$$

The utility that an agent with private information θ obtains from using the device is $\widehat{U}_\theta^Y(\widetilde{\theta}) := (1 - \widehat{s}_\theta)u(\widehat{\mu}_L) + \widehat{s}_\theta u(\widehat{\mu}_H)$ and the utility that he obtains from not using the device is $\widehat{U}^N(\widetilde{\theta}) := u(\widehat{\mu}_N)$. Using a reasoning similar to that in the proof of Lemma 1, there exists an equilibrium characterized by some threshold $\widetilde{\theta} \in (0, 1)$ if and only if $\widehat{U}_\theta^Y(\widetilde{\theta}) = \widehat{U}^N(\widetilde{\theta})$.

From an ex ante perspective, the agent's expected utility is

$$\begin{aligned} \mathbf{E}[u(\widehat{\mu})] &= \int_0^{\widetilde{\theta}} \widehat{U}^N(\widetilde{\theta}) \widehat{p}(\theta) d\theta + \int_{\widetilde{\theta}}^1 \widehat{U}_\theta^Y(\widetilde{\theta}) \widehat{p}(\theta) d\theta \\ &= \left[\int_0^{\widetilde{\theta}} \widehat{p}(\theta) d\theta \right] u(\widehat{\mu}_N) + \left[\int_{\widetilde{\theta}}^1 \widehat{p}(\theta) d\theta \right] [(1 - \widehat{\mu}_Y)(1 - s_0) + \widehat{\mu}_Y (1 - s_1)] u(\widehat{\mu}_L) \\ &\quad + \left[\int_{\widetilde{\theta}}^1 \widehat{p}(\theta) d\theta \right] [(1 - \widehat{\mu}_Y) s_0 + \widehat{\mu}_Y s_1] u(\widehat{\mu}_H). \end{aligned}$$

Hence, under Assumption 3 the designer's optimization problem is

$$\min_{s_0, s_1} \mathbf{E}[u(\widehat{\mu})] \text{ s.t. } \widehat{U}_\theta^Y(\widetilde{\theta}) = \widehat{U}^N(\widetilde{\theta}).$$

An auxiliary problem with binary private information. Consider now for a moment the setting with discrete private information characterized by

$$p_l = \int_0^{\widetilde{\theta}} \widehat{p}(\theta) d\theta, \quad p_h = \int_{\widetilde{\theta}}^1 \widehat{p}(\theta) d\theta, \quad p_{l\omega} = \frac{\widehat{p}(\widetilde{\theta}, \omega)}{p(\widetilde{\theta})} p_l, \quad p_{h\omega} = \int_{\widetilde{\theta}}^1 \widehat{p}_\omega \widehat{p}(\theta, \omega) d\theta$$

and note that for $\tilde{\theta} \in (0, 1)$ we obtain $p_{l1}, p_{l0}, p_{h1}, p_{h0} \in (0, 1)$ and $\delta = \frac{p_{h1}}{p_h} - \frac{p_{l1}}{p_l} > 0$ such that the probability distribution satisfies the basic assumptions introduced in Section 3. p_l , p_h and $p_{h\omega}$ can be seen as the probabilities that arise in a discretized version of our continuous model where “participants” and “non-participants” are pooled. p_l is the mass of non-participants, p_h the mass of participants, and $\frac{p_{h1}}{p_h}$ is the “average” probability with which a participant believes to be good. However, $\frac{p_{l1}}{p_l}$ is not the “average” probability with which a non-participant believes to be good, but the probability with which the “best” non-participant believes to be good. This is to take account of the fact that the participation constraint has to be binding for the best non-participant in the continuous version of the model. As described below, the solution to this auxiliary binary problem allows us to infer the solution to the original continuous problem.

Consider participation behavior with $\alpha_l = 0$ and a binding participation constraint as analyzed in Subsection 6.1. The belief assigned to the event that the device is not used is

$$\mu_N = \frac{p_{l1}}{p_l} = \frac{\hat{p}(\tilde{\theta}, \omega = 1)}{\hat{p}(\tilde{\theta})}$$

and the belief assigned to the event that the device is used is

$$\mu_Y = \frac{p_{h1}}{p_h} = \frac{\int_{\tilde{\theta}}^1 \hat{p}_1 \hat{p}(\theta, \omega = 1) d\theta}{\int_{\tilde{\theta}}^1 \hat{p}(\theta) d\theta}.$$

Since we have $\mu_Y = \hat{\mu}_Y$ (by construction), it follows $\mu_L = \hat{\mu}_L$ and $\mu_H = \hat{\mu}_H$. Moreover, the probability with which a high public signal is generated when an agent with a low private signal uses the device is

$$s_l = \left(1 - \frac{\hat{p}(\tilde{\theta}, \omega = 1)}{\hat{p}(\tilde{\theta})}\right) s_0 + \frac{\hat{p}(\tilde{\theta}, \omega = 1)}{\hat{p}(\tilde{\theta})} s_1$$

such that $s_l = \hat{s}_{\tilde{\theta}}$ and $U_l^Y(0, 1) = \hat{U}_{\tilde{\theta}}^Y(\tilde{\theta})$. However, for the utility the agent obtains from not using the device we obtain $U^N(0, 1) = u(\mu_N)$ which differs from $\hat{U}^N(\tilde{\theta}) = u(\hat{\mu}_N)$. Taking this difference into account, the agent’s ex ante expected utility in the auxiliary problem can be expressed as his ex ante expected utility in the original problem adjusted for the difference in utility from not using the device: $\mathbf{E}[u(\mu)] = \mathbf{E}[u(\hat{\mu})] + p_l[u(\mu_N) - u(\hat{\mu}_N)]$. Under Assumption 3 the designer’s optimization problem becomes

$$\min_{s_0, s_1} \mathbf{E}[u(\hat{\mu})] + p_l[u(\mu_N) - u(\hat{\mu}_N)] \text{ s.t. } \hat{U}_{\tilde{\theta}}^Y(\tilde{\theta}) = U^N(0, 1).$$

Since $p_l[u(\hat{\mu}_N) - u(\mu_N)]$ is not affected by the device choice, this problem is equivalent to

$$\min_{s_0, s_1} \mathbf{E}[u(\hat{\mu})] \text{ s.t. } \hat{U}_{\tilde{\theta}}^Y(\tilde{\theta}) = U^N(0, 1)$$

which is the original problem in the continuous version of the model, except for the fact that the constant on the right-hand side of the participation constraint differs.

The optimal device in the continuous version of the model. Lemma 8 implies that under Assumptions 1, 2 and 3 $s_0 = 0$ is the solution to the auxiliary problem. Recall that the

only difference between the optimization problem in the continuous and in the discrete version of the model is the size of the constant on the right-hand side of the participation constraint. However, the proof to Lemma 8 does not directly make use of the size of this constant, it makes only use of the existence of a function $s_1(s_0)$ (with the properties specified in Lemma 4) for which the participation constraint is binding. By reconsidering the proof to Lemma 4, we obtain that if a function $s_1(s_0)$ exists also in the case with the altered participation constraint, then it has the same properties as in the discrete version of the model such that Lemma 8 applies also to the continuous version of the model. Hence, if this is the case, we can conclude that $s_0 = 0$ is optimal also with continuous private information. However, with continuous private information it might happen that a function $s_1(s_0)$ for which the participation constraint is binding fails to exist. The reason for this is the following: Since an agent with private information $\tilde{\theta}$ is on average perceived as worse than he actually is when he does not use the device, he might have a strict incentive to use a perfect device. If this is the case, there exists no device (s_0, s_1) for which the participation constraint is binding. A consequence of this is that an agent with private information $\tilde{\theta}$ has a strict incentive to use any device. By a continuity property, the agent has also a strict incentive to use the device when his private information is just below $\tilde{\theta}$ contradicting that there exists an equilibrium with participation threshold $\tilde{\theta}$. This establishes the following result:

Proposition 8 *If Assumptions 1, 2 and 3 hold and if there exists an equilibrium of the continuous version of the model with threshold $\tilde{\theta}$, then there exists also an equilibrium in which this threshold is induced by a device with $s_0 = 0$ and the equilibrium induced by the device with $s_0 = 0$ is optimal for the designer.*

9 Conclusion

The present paper explores a framework with imperfect private information and voluntary participation to study information generation from a mechanism design perspective. In our model a mechanism is a device which is capable of generating information that goes beyond what the agent knows. Exploiting the agent's incentive to signal superior private information, the designer can incentivize the agent to use a mechanism which generates additional information. On the other hand, with a mechanism which does not generate any additional information, it is also not possible to learn anything about the agent's private information. However, voluntary participation in conjunction with risk aversion effectively constrain information generation.

When the designer is interested in a binary feature (e.g., whether the agent is good/bad, he has/does not have a certain genetic defect or a disease, his product satisfies/does not satisfy certain quality standards, ...), the optimal device generates binary information and can thus be interpreted as a test which produces a pass-fail-result. In order to induce participation, it may be necessary that the device is susceptible to errors. However, the optimal device commits at most one type of error. While it may be subject to false negatives, it is not subject to false positives. That is, the optimal test is hard to pass and if the test is passed, then it can be perfectly inferred

that the agent's true type is good. If the designer wants the agent to participate not only if he holds the most optimistic beliefs about his true type, then the test becomes even harder to pass when higher participation shall be induced. The reason is that the stigma of failure is lower with a harder test such that participation becomes more attractive for the agent in case he has pessimistic beliefs about his true type. In many design problems, the designer is not entirely free to choose the precision of generated information (e.g. due to technical restrictions). With voluntary participation, such restrictions need not be binding because the generated information may need to be distorted in order to induce participation anyway.

A Proofs

Proofs of Section 5

Proof (Lemma 1). Note that $U_h^Y(\alpha_l, \alpha_h) - U_l^Y(\alpha_l, \alpha_h) = (s_h - s_l)[u(\mu_H) - u(\mu_L)] = \delta(s_1 - s_0)[u(\mu_H) - u(\mu_L)]$. We have $\delta > 0$, $s_1 > s_0$ and $u(\mu)$ strictly increasing. Thus, if we can show that $\mu_H > \mu_L$, we are done. If $\alpha_l > 0$ or $\alpha_h > 0$, Bayes' Law yields $\mu_H - \mu_L = \frac{[p_{l1}\alpha_l + p_{h1}\alpha_h][p_{l0}\alpha_l + p_{h0}\alpha_h]}{p_L p_H}(s_1 - s_0)$. This expression is strictly positive. If $\alpha_l = \alpha_h = 0$, it follows from our assumptions on beliefs that $\mu_H - \mu_L = \frac{p_{h0}p_{h1}}{[p_{h1}s_1 + p_{h0}s_0][p_{h1}(1-s_1) + p_{h0}(1-s_0)]}(s_1 - s_0)$ which is strictly positive as well. q.e.d.

Proof (Lemma 2). Assume to the contrary that there exists a device which induces an equilibrium with $\alpha_l \in (0, 1]$ and $\alpha_h \in [0, 1)$. Then individual rationality implies $U_h^Y(\alpha_l, \alpha_h) \leq U^N(\alpha_l, \alpha_h)$ and $U^N(\alpha_l, \alpha_h) \leq U_l^Y(\alpha_l, \alpha_h)$ contradicting Lemma 1. q.e.d.

Proofs of Section 6

Proof (Lemma 3). It can easily be checked that μ_L^{II} , μ_H^{II} and s_l are differentiable in (s_0, s_1) . This and differentiability of $u(\mu)$ imply that also $U_l^Y(\alpha_l, 1)$ is differentiable in s_0 and s_1 . We get

$$\frac{\partial}{\partial s_\omega} U_l^Y(\alpha_l, 1) = \frac{\partial s_l}{\partial s_\omega} [u_H - u_L] + (1 - s_l) \frac{\partial \mu_L^{II}}{\partial s_\omega} u'_L + s_l \frac{\partial \mu_H^{II}}{\partial s_\omega} u'_H. \quad (8)$$

Consider first $\omega = 0$: Inserting (1b) and (2b) in (8) and using

$$\frac{1 - s_l}{p_{L|Y}} = 1 + \frac{[\mu_Y - p_1 + \delta p_h]}{p_{L|Y}}(s_1 - s_0) \quad (9)$$

as well as

$$\frac{s_l}{p_{H|Y}} = 1 - \frac{[\mu_Y - p_1 + \delta p_h]}{p_{H|Y}}(s_1 - s_0) \quad (10)$$

we obtain after rearranging:

$$\begin{aligned} \frac{\partial U_l^Y}{\partial s_0} &= [(p_0 + \delta p_h)u_H - (1 - \mu_Y)\mu_H^{II}u'_H] - [(p_0 + \delta p_h)u_L - (1 - \mu_Y)\mu_L^{II}u'_L] \\ &\quad + (1 - \mu_Y)(\mu_Y - p_1 + \delta p_h)(s_1 - s_0) \cdot \left\{ \frac{\mu_L^{II}u'_L}{p_{L|Y}} + \frac{\mu_H^{II}u'_H}{p_{H|Y}} \right\}. \end{aligned}$$

Note that the derivative of $(p_0 + \delta p_h)u(\mu) - (1 - \mu_Y)\mu u'(\mu)$ with respect to μ is $(\mu_Y - p_1 + \delta p_h)u'(\mu) - (1 - \mu_Y)\mu u''(\mu)$ which is strictly positive since $\mu_Y \geq p_1$ and since $u(\mu)$ is strictly increasing and strictly concave. Thus, the expression in the first line is strictly positive. Moreover, the expression in the last line is also strictly positive such that $\frac{\partial U_l^Y}{\partial s_0} > 0$.

Consider now $\omega = 1$. Inserting (3b) and (4b) in (8) and using (9) and (10) we obtain:

$$\begin{aligned} \frac{\partial U_l^Y}{\partial s_1} &= [(p_1 - \delta p_h)u_H + \mu_Y(1 - \mu_H^{II})u'_H] - [(p_1 - \delta p_h)u_L + \mu_Y(1 - \mu_L^{II})u'_L] \\ &\quad - \mu_Y(\mu_Y - p_1 + \delta p_h)(s_1 - s_0) \cdot \left\{ \frac{(1 - \mu_L^{II})u'_L}{p_{L|Y}} + \frac{(1 - \mu_H^{II})u'_H}{p_{H|Y}} \right\}. \end{aligned}$$

Note that the derivative of $(p_1 - \delta p_h)u(\mu) + \mu_Y(1 - \mu)u'(\mu)$ with respect to μ is $-(\mu_Y - p_1 + \delta p_h)u'(\mu) + \mu_Y(1 - \mu)u''(\mu)$ which is strictly negative. Thus, the expression in the first line is strictly negative. Moreover, the expression in the last line is also strictly negative such that $\frac{\partial U_l^Y}{\partial s_1} < 0$. q.e.d.

Proof (Lemma 4). Note that $U^N(\alpha_l, 1) = u(\frac{p_l}{p_l})$ for any device. $(s_0, s_1) \rightarrow (0, 0)$ implies $U_l^Y(\alpha_l, 1) \rightarrow u(\mu_Y) > u(\frac{p_l}{p_l})$. Thus, the agent has a strict incentive to use a sufficiently uninformative device when his private information is low. By strict concavity of $u(\mu)$, the agent has a strict incentive not to use the perfect device $(s_0, s_1) = (0, 1)$ when his private information is low: $U_l^Y(\alpha_l, 1) = (1 - s_l)u(0) + s_l u(1) < u(s_l) = U^N(\alpha_l, 1)$. Moreover, by Lemma 3, $U_l^Y(\alpha_l, 1)$ is continuous in s_1 for $s_0 = 0$. Hence, we can apply an Intermediate Value Theorem to obtain that there exists a $\underline{s}_1 \in (0, 1)$ such that $U^N(\alpha_l, 1) = u(\frac{p_l}{p_l})$ for device $(0, \underline{s}_1)$. Since $U_l^Y(\alpha_l, 1)$ is strictly increasing in s_1 by Lemma 3, \underline{s}_1 is unique.

It follows directly from Lemma 3 that there exists a connected and closed domain $[0, \bar{s}_0]$ and a continuous and increasing function $s_1(s_0)$ on this domain such that $U_l^Y(\alpha_l, 1) = U^N(\alpha_l, 1)$ if and only if $s_0 \in [0, \bar{s}_0]$ and $s_1 = s_1(s_0)$. By construction, $\underline{s}_1 = s_1(0) \in (0, 1)$. For $s_0 = 1$ there exists no device for which the participation constraint is binding, since an agent with a low private signal has a strict incentive to use the uninformative device $(1, 1)$. Hence, $\bar{s}_0 < 1$.

Since $s_1(s_0)$ is monotonous, it is differentiable almost everywhere. We obtain $s_1'(s_0) = -\frac{\partial U_l^Y}{\partial s_0} / \frac{\partial U_l^Y}{\partial s_1} > 0$ for any s_0 in the domain. q.e.d.

Proof (Lemma 5). Lemma 4 implies $\frac{d\mu_H^{II}}{ds_0} = \frac{\partial \mu_H^{II}}{\partial s_0} + s_1'(s_0) \frac{\partial \mu_H^{II}}{\partial s_1} = [-\frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] / [-\frac{\partial U_l^Y}{\partial s_1}]$. It follows from Lemma 3 that $\frac{d\mu_H^{II}}{ds_0} < 0$ if and only if $[-\frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] < 0$.

$$\begin{aligned} &-\frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0} \\ &= [u_H - u_L] \left[\frac{\partial s_l}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} \right] + (1 - s_l) \left[\frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} \right] u'_L. \end{aligned} \quad (11)$$

Using (1a) and (3a) we get $\left[\frac{\partial s_l}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} \right] = \frac{\mu_Y(1 - \mu_Y)}{p_{H|Y}^2} s_l$ and using (1b), (2b), (3b) and

(4b) we get $\left[\frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} \right] = -\frac{\mu_Y(1-\mu_Y)}{p_{L|Y}p_{H|Y}}(\mu_H^{II} - \mu_L^{II})$. Inserting this into (11), we obtain

$$\dots = \frac{\mu_Y(1-\mu_Y)}{p_{H|Y}^2} s_l (\mu_H^{II} - \mu_L^{II}) \left\{ \frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}} - \frac{p_{H|Y}}{s_l} \frac{1-s_l}{p_{L|Y}} u'_L \right\}. \quad (12)$$

Since $\frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}} < u'_L$ by strict concavity of $u(\mu)$ and since $\frac{p_{H|Y}}{s_l} \frac{1-s_l}{p_{L|Y}} > 1$, we obtain that (12) is strictly negative. Hence, $\frac{d\mu_H^{II}}{ds_0} < 0$.

Lemma 4 implies $\frac{d\mu_L^{II}}{ds_0} = \frac{\partial \mu_L^{II}}{\partial s_0} + s'_1(s_0) \frac{\partial \mu_L^{II}}{\partial s_1} = [-\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] / [-\frac{\partial U_l^Y}{\partial s_1}]$. It follows from Lemma 3 that $\frac{d\mu_L^{II}}{ds_0} < 0$ if and only if $[-\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] < 0$.

$$\begin{aligned} & -\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0} \\ &= [u_H - u_L] \left[\frac{\partial s_l}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} \right] - s_l \left[\frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} \right] u'_H \end{aligned} \quad (13)$$

Using (2a) and (4a) we get $\left[\frac{\partial s_l}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} \right] = -\frac{\mu_Y(1-\mu_Y)}{p_{L|Y}^2} (1-s_l)$ and using (1b), (2b), (3b) and (4b) we get $\left[\frac{\partial \mu_H^{II}}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial \mu_H^{II}}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} \right] = -\frac{\mu_Y(1-\mu_Y)}{p_{L|Y}p_{H|Y}} (\mu_H^{II} - \mu_L^{II})$. Inserting this into (13), we obtain

$$\dots = \frac{\mu_Y(1-\mu_Y)}{p_{L|Y}^2} (1-s_l) (\mu_H^{II} - \mu_L^{II}) \left\{ -\frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}} + \frac{s_l}{p_{H|Y}} \frac{p_{L|Y}}{1-s_l} u'_H \right\}. \quad (14)$$

Since $\frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}} > u'_H$ by strict concavity of $u(\mu)$ and since $\frac{s_l}{p_{H|Y}} \frac{p_{L|Y}}{1-s_l} < 1$, we obtain that (14) is strictly negative. Hence, $\frac{d\mu_L^{II}}{ds_0} < 0$. q.e.d.

Proof (Lemma 6). Lemma 4 implies $\frac{d\mathbb{E}_l}{ds_0} = \frac{\partial \mathbb{E}_l}{\partial s_0} + s'_1(s_0) \frac{\partial \mathbb{E}_l}{\partial s_1} = [-\frac{\partial \mathbb{E}_l}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mathbb{E}_l}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] / [-\frac{\partial U_l^Y}{\partial s_1}]$. It follows from Lemma 3 that $\frac{d\mathbb{E}_l}{ds_0} > 0$ if and only if $[-\frac{\partial \mathbb{E}_l}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mathbb{E}_l}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0}] > 0$.

$$\begin{aligned} & -\frac{\partial \mathbb{E}_l}{\partial s_0} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial \mathbb{E}_l}{\partial s_1} \frac{\partial U_l^Y}{\partial s_0} \\ &= -\left\{ \frac{\partial s_l}{\partial s_0} [\mu_H^{II} - \mu_L^{II}] + (1-s_l) \frac{\partial \mu_L^{II}}{\partial s_0} + s_l \frac{\partial \mu_H^{II}}{\partial s_0} \right\} \\ & \quad \cdot \left\{ \frac{\partial s_l}{\partial s_1} [u_H - u_L] + (1-s_l) \frac{\partial \mu_L^{II}}{\partial s_1} u'_L + s_l \frac{\partial \mu_H^{II}}{\partial s_1} u'_H \right\} \\ & \quad + \left\{ \frac{\partial s_l}{\partial s_1} [\mu_H^{II} - \mu_L^{II}] + (1-s_l) \frac{\partial \mu_L^{II}}{\partial s_1} + s_l \frac{\partial \mu_H^{II}}{\partial s_1} \right\} \\ & \quad \cdot \left\{ \frac{\partial s_l}{\partial s_0} [u_H - u_L] + (1-s_l) \frac{\partial \mu_L^{II}}{\partial s_0} u'_L + s_l \frac{\partial \mu_H^{II}}{\partial s_0} u'_H \right\} \\ &= (1-s_l) \{ [\mu_H^{II} - \mu_L^{II}] u'_L - [u_H - u_L] \} \left[\frac{\partial s_l}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial s_l}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1} \right] \\ & \quad + s_l \{ [\mu_H^{II} - \mu_L^{II}] u'_H - [u_H - u_L] \} \left[\frac{\partial s_l}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} - \frac{\partial s_l}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} \right] \\ & \quad + (u'_L - u'_H) s_l (1-s_l) \left[\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} \right]. \end{aligned} \quad (15)$$

Using (2a) and (4a) we get $\left[\frac{\partial s_l}{\partial s_1} \frac{\partial \mu_L^{II}}{\partial s_0} - \frac{\partial s_l}{\partial s_0} \frac{\partial \mu_L^{II}}{\partial s_1}\right] = \frac{\mu_Y(1-\mu_Y)}{p_{L|Y}^2}(1-s_l)$, using (1a) and (3a) we get $\left[\frac{\partial s_l}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} - \frac{\partial s_l}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1}\right] = -\frac{\mu_Y(1-\mu_Y)}{p_{H|Y}^2}s_l$ and using (1b), (2b), (3b) and (4b) we get $\left[\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0}\right] = -\frac{\mu_Y(1-\mu_Y)}{p_{L|Y}p_{H|Y}}(\mu_H^{II} - \mu_L^{II})$. Inserting this into (15), we obtain

$$\begin{aligned}
\dots &= \mu_Y(1-\mu_Y)\frac{(1-s_l)^2}{p_{L|Y}^2} \{[\mu_H^{II} - \mu_L^{II}]u'_L - [u_H - u_L]\} \\
&\quad - \mu_Y(1-\mu_Y)\frac{s_l^2}{p_{H|Y}^2} \{[\mu_H^{II} - \mu_L^{II}]u'_H - [u_H - u_L]\} \\
&\quad - \mu_Y(1-\mu_Y)[\mu_H^{II} - \mu_L^{II}](u'_L - u'_H)\frac{s_l}{p_{H|Y}}\frac{(1-s_l)}{p_{L|Y}} \\
&= \mu_Y(1-\mu_Y)\left(\frac{1-s_l}{p_{L|Y}} - \frac{s_l}{p_{H|Y}}\right)[\mu_H^{II} - \mu_L^{II}] \\
&\quad \cdot \left\{\frac{1-s_l}{p_{L|Y}}[u'_L - \frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}}] + \frac{s_l}{p_{H|Y}}[u'_H - \frac{u_H - u_L}{\mu_H^{II} - \mu_L^{II}}]\right\}. \tag{16}
\end{aligned}$$

We have $(1-s_l)/p_{L|Y} > 1$ and $s_l/p_{H|Y} < 1$. Since also $u'_L - (u_H - u_L)/(\mu_H^{II} - \mu_L^{II}) > 0$ by strict concavity of $u(\mu)$, $(u'_L + u'_H)[\mu_H^{II} - \mu_L^{II}] - 2(u_H - u_L) \geq 0$ is sufficient for (16) > 0 .

Define $Z_\mu(\epsilon) := [u'(\mu + \epsilon) + u'(\mu)]\epsilon - 2[u(\mu + \epsilon) - u(\mu)]$ with $\epsilon > 0$. Then $Z'_\mu(\epsilon) = u''(\mu + \epsilon)\epsilon + [u'(\mu + \epsilon) + u'(\mu)] - 2u'(\mu + \epsilon)$, $Z''_\mu(\epsilon) = u'''(\mu + \epsilon)\epsilon$, $Z_\mu(0) = 0$ and $Z'_\mu(0) = 0$. From this it follows

$$\begin{aligned}
Z_\mu(\epsilon) \geq 0 &\Leftrightarrow \int_0^\epsilon Z'_\mu(e_1)de_1 + Z_\mu(0) \geq 0 \\
&\Leftrightarrow \int_0^\epsilon \left[\int_0^{e_1} Z''_\mu(e_2)de_2 + Z'_\mu(0) \right] de_1 \geq 0 \\
&\Leftrightarrow \int_0^\epsilon \int_0^{e_1} u'''(\mu + e_2)e_2de_2de_1 \geq 0.
\end{aligned}$$

This is true if $u'''(\mu) \geq 0$. Thus, $u'''(\mu) \geq 0$ implies $\frac{d\mathbb{E}_l}{ds_0} > 0$. This together with $\frac{d\mu_Y}{ds_0} = 0$ (which holds by construction of $s_1(s_0)$) implies $\frac{d\mathbb{E}_h}{ds_0} < 0$ if $\alpha_l > 0$. If $\alpha_l = 0$, we have $\mu_Y = 0 \cdot \mathbb{E}_l + 1 \cdot \mathbb{E}_h$ such that a change in \mathbb{E}_l does not affect \mathbb{E}_h . q.e.d.

Proof (Lemma 7). This follows directly from Lemma 6 and Lemma 8 below. \mathbb{E}_l increases with s_0 , while U_l is constant in s_0 . By (5), this is only possible when \mathbb{V}_l increases with s_0 . Moreover, \mathbb{E}_h decreases (weakly) with s_0 , while U_h increases with s_0 . Again, by (5), this is only possible when \mathbb{V}_h decreases with s_0 . q.e.d.

Proof (Lemma 8). $\frac{dU^Y}{ds_0} = 0$ is true by construction of $s_1(s_0)$.

Lemma 4 implies $\frac{dU^Y_h}{ds_0} = \frac{\partial U^Y_h}{\partial s_0} + s'_1(s_0)\frac{\partial U^Y_h}{\partial s_1} = [-\frac{\partial U^Y_h}{\partial s_0}\frac{\partial U^Y_l}{\partial s_1} + \frac{\partial U^Y_h}{\partial s_1}\frac{\partial U^Y_l}{\partial s_0}]/[-\frac{\partial U^Y_l}{\partial s_1}]$. It follows from Lemma 3 that $\frac{dU^Y_h}{ds_0} > 0$ if and only if $[-\frac{\partial U^Y_h}{\partial s_0}\frac{\partial U^Y_l}{\partial s_1} + \frac{\partial U^Y_h}{\partial s_1}\frac{\partial U^Y_l}{\partial s_0}] > 0$.

$$-\frac{\partial U^Y_h}{\partial s_0}\frac{\partial U^Y_l}{\partial s_1} + \frac{\partial U^Y_h}{\partial s_1}\frac{\partial U^Y_l}{\partial s_0}$$

$$\begin{aligned}
&= - \left\{ \frac{\partial s_h}{\partial s_0} [u_H - u_L] + (1 - s_h) \frac{\partial \mu_L^{II}}{\partial s_0} u'_L + s_h \frac{\partial \mu_H^{II}}{\partial s_0} u'_H \right\} \\
&\quad \cdot \left\{ \frac{\partial s_l}{\partial s_1} [u_H - u_L] + (1 - s_l) \frac{\partial \mu_L^{II}}{\partial s_1} u'_L + s_l \frac{\partial \mu_H^{II}}{\partial s_1} u'_H \right\} \\
&+ \left\{ \frac{\partial s_h}{\partial s_1} [u_H - u_L] + (1 - s_h) \frac{\partial \mu_L^{II}}{\partial s_1} u'_L + s_h \frac{\partial \mu_H^{II}}{\partial s_1} u'_H \right\} \\
&\quad \cdot \left\{ \frac{\partial s_l}{\partial s_0} [u_H - u_L] + (1 - s_l) \frac{\partial \mu_L^{II}}{\partial s_0} u'_L + s_l \frac{\partial \mu_H^{II}}{\partial s_0} u'_H \right\} \\
&= [u_H - u_L]^2 \left[\frac{\partial s_l}{\partial s_0} \frac{\partial s_h}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial s_h}{\partial s_0} \right] \\
&+ [u_H - u_L] u'_L \left\{ \left[(1 - s_l) \frac{\partial s_h}{\partial s_1} - (1 - s_h) \frac{\partial s_l}{\partial s_1} \right] \frac{\partial \mu_L^{II}}{\partial s_0} \right. \\
&\quad \left. + \left[(1 - s_h) \frac{\partial s_l}{\partial s_0} - (1 - s_l) \frac{\partial s_h}{\partial s_0} \right] \frac{\partial \mu_L^{II}}{\partial s_1} \right\} \\
&+ [u_H - u_L] u'_H \left\{ \left[s_l \frac{\partial s_h}{\partial s_1} - s_h \frac{\partial s_l}{\partial s_1} \right] \frac{\partial \mu_H^{II}}{\partial s_0} + \left[s_h \frac{\partial s_l}{\partial s_0} - s_l \frac{\partial s_h}{\partial s_0} \right] \frac{\partial \mu_H^{II}}{\partial s_1} \right\} \\
&+ u'_L u'_H [(1 - s_l) s_h - (1 - s_h) s_l] \left[\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0} \right] \tag{17}
\end{aligned}$$

Note that $[\frac{\partial s_l}{\partial s_0} \frac{\partial s_h}{\partial s_1} - \frac{\partial s_l}{\partial s_1} \frac{\partial s_h}{\partial s_0}] = \delta$, $[(1 - s_l) \frac{\partial s_h}{\partial s_1} - (1 - s_h) \frac{\partial s_l}{\partial s_1}] = \delta(1 - s_0)$, $[(1 - s_h) \frac{\partial s_l}{\partial s_0} - (1 - s_l) \frac{\partial s_h}{\partial s_0}] = \delta(1 - s_1)$, $[s_l \frac{\partial s_h}{\partial s_1} - s_h \frac{\partial s_l}{\partial s_1}] = \delta s_0$, $[s_h \frac{\partial s_l}{\partial s_0} - s_l \frac{\partial s_h}{\partial s_0}] = \delta s_1$ and $[(1 - s_l) s_h - (1 - s_h) s_l] = \delta(s_1 - s_0)$. Moreover, using (1b), (2b), (3b) and (4b) we get $[\frac{\partial \mu_L^{II}}{\partial s_0} \frac{\partial \mu_H^{II}}{\partial s_1} - \frac{\partial \mu_L^{II}}{\partial s_1} \frac{\partial \mu_H^{II}}{\partial s_0}] = -\frac{(\mu_H^{II} - \mu_L^{II})^2}{(s_1 - s_0)}$ such that (17) becomes

$$\begin{aligned}
\dots &= \delta [u_H - u_L]^2 + \delta [u_H - u_L] u'_L \left[(1 - s_0) \frac{\partial \mu_L^{II}}{\partial s_0} + (1 - s_1) \frac{\partial \mu_L^{II}}{\partial s_1} \right] \\
&\quad + \delta [u_H - u_L] u'_H \left[s_0 \frac{\partial \mu_H^{II}}{\partial s_0} + s_1 \frac{\partial \mu_H^{II}}{\partial s_1} \right] - \delta u'_L u'_H (\mu_H^{II} - \mu_L^{II})^2. \tag{18}
\end{aligned}$$

It follows directly from (2a) and (4a) that $[(1 - s_0) \frac{\partial \mu_L^{II}}{\partial s_0} + (1 - s_1) \frac{\partial \mu_L^{II}}{\partial s_1}] = 0$, and from (1a) and (3a) that $[s_0 \frac{\partial \mu_H^{II}}{\partial s_0} + s_1 \frac{\partial \mu_H^{II}}{\partial s_1}] = 0$. Hence, (18) becomes

$$\dots = \delta [u_H - u_L]^2 - \delta u'_L u'_H (\mu_H^{II} - \mu_L^{II})^2. \tag{19}$$

We derive now a sufficient condition for this expression being strictly positive implying $\frac{dU_h^Y}{ds_0} > 0$.

Define $Z_\mu(\epsilon) := [u(\mu + \epsilon) - u(\mu)]^2 - u'(\mu + \epsilon)u'(\mu)\epsilon^2$ with $\epsilon > 0$. Then $Z'_\mu(\epsilon) = 2u'(\mu + \epsilon)[u(\mu + \epsilon) - u(\mu)] - [2\epsilon u'(\mu + \epsilon) + \epsilon^2 u''(\mu + \epsilon)]u'(\mu)$, $Z''_\mu(\epsilon) = 2u''(\mu + \epsilon)[u(\mu + \epsilon) - u(\mu)] + 2u'(\mu + \epsilon)[u'(\mu + \epsilon) - u'(\mu)] - [4\epsilon u''(\mu + \epsilon) + \epsilon^2 u'''(\mu + \epsilon)]u'(\mu)$, $Z'''_\mu(\epsilon) = 2u'''(\mu + \epsilon)[u(\mu + \epsilon) - u(\mu)] + 6u''(\mu + \epsilon)[u'(\mu + \epsilon) - u'(\mu)] - [6\epsilon u'''(\mu + \epsilon) + \epsilon^2 u''''(\mu + \epsilon)]u'(\mu)$, $Z_\mu(0) = 0$, $Z'_\mu(0) = 0$, $Z''_\mu(0) = 0$ and $Z'''_\mu(0) = 0$. Moreover,

$$\begin{aligned}
Z''''_\mu(\epsilon) &= 6[u''(\mu + \epsilon)^2 - u'''(\mu + \epsilon)u'(\mu)] + 2u'''(\mu + \epsilon)u'(\mu + \epsilon) \\
&\quad + 2u''''(\mu + \epsilon)[u(\mu + \epsilon) - u(\mu)] - [8\epsilon u''''(\mu + \epsilon) + \epsilon^2 u'''''(\mu + \epsilon)]u'(\mu). \tag{20}
\end{aligned}$$

The second expression in the first line is non-negative by Assumption 1. The expression in the second line is zero since $u''''(\mu) = 0$ (second part of Assumption 2). Furthermore, $u''(\mu + \epsilon) \geq u''(\mu)$ by Assumption 1 and $u'''(\mu + \epsilon) = u'''(\mu)$ by the second part of Assumption 2. This implies

$$Z_\mu''''(\epsilon) \geq 6[u''(\mu)^2 - u'''(\mu)u'(\mu)]$$

which is strictly positive by the first part of Assumption 2.

Note that

$$\begin{aligned} Z_\mu(\epsilon) &> 0 \\ \Leftrightarrow \int_0^\epsilon Z_\mu'(e_1)de_1 + Z_\mu(0) &> 0 \\ \Leftrightarrow \int_0^\epsilon \left[\int_0^{e_1} Z_\mu''(e_2)de_2 + Z_\mu'(0) \right] de_1 &> 0 \\ \Leftrightarrow \int_0^\epsilon \int_0^{e_1} \left[\int_0^{e_2} Z_\mu'''(e_3)de_3 + Z_\mu''(0) \right] de_2de_1 &> 0 \\ \Leftrightarrow \int_0^\epsilon \int_0^{e_1} \int_0^{e_2} \left[\int_0^{e_3} Z_\mu''''(e_4)de_4 + Z_\mu'''(0) \right] de_3de_2de_1 &> 0 \\ \Leftrightarrow \int_0^\epsilon \int_0^{e_1} \int_0^{e_2} \int_0^{e_3} Z_\mu''''(e_4)de_4de_3de_2de_1 &> 0. \end{aligned}$$

Hence, Assumption 1 and Assumption 2 are sufficient for $Z_\mu(\epsilon) > 0$ such that the two assumptions imply $\frac{dU_\mu^Y}{ds_0} > 0$. q.e.d.

Proof (Lemma 9). We have to show that expression (19) in the proof to Lemma 8 is strictly positive for CARA preferences with arbitrary risk parameter, i.e. for $u(\mu) = -\exp(-\gamma\mu)$ with $\gamma > 0$. By making the change in variables $\mu_H^I = x + \epsilon$ and $\mu_L^I = x$, a sufficient condition for the result is

$$\delta[u(x + \epsilon) - u(x)]^2 - \delta u'(x + \epsilon)u'(x)\epsilon^2 > 0 \text{ for any } \epsilon, \gamma > 0.$$

By using $\frac{u(x+\epsilon)}{u(x)} = \exp(-\gamma\epsilon)$, $\frac{u'(x+\epsilon)}{u'(x)} = -\gamma \exp(-\gamma\epsilon)$ and $\frac{u'(x)}{u(x)} = -\gamma$ we can write this condition as

$$[\exp(-\gamma\epsilon) - 1]^2 - (\gamma\epsilon)^2 \exp(-\gamma\epsilon) > 0 \text{ for any } \epsilon, \gamma > 0.$$

After substituting $\tilde{\gamma} = \gamma\epsilon$ and doing some rearrangements this becomes

$$1 > \exp(-\tilde{\gamma}) + \tilde{\gamma} \exp(-\frac{1}{2}\tilde{\gamma}) \text{ for any } \tilde{\gamma} > 0.$$

Define $\xi(\tilde{\gamma}) := \exp(-\tilde{\gamma}) + \tilde{\gamma} \exp(-\frac{1}{2}\tilde{\gamma})$. Since $\xi(0) = 1$, $\xi'(\tilde{\gamma}) = \exp(-\frac{1}{2}\tilde{\gamma})(-\exp(-\frac{1}{2}\tilde{\gamma}) + 1 - \frac{1}{2}\tilde{\gamma}) < 0$ for any $\tilde{\gamma}$ is sufficient for the result. After doing some more rearrangements, we obtain that

$$1 < \exp(-\frac{1}{2}\tilde{\gamma}) + \frac{1}{2}\tilde{\gamma} \text{ for any } \tilde{\gamma} > 0$$

is sufficient for the result. Define now $\eta(\tilde{\gamma}) := \exp(-\frac{1}{2}\tilde{\gamma}) + \frac{1}{2}\tilde{\gamma}$. Since $\eta(0) = 1$ and $\eta'(\tilde{\gamma}) = -\frac{1}{2} \exp(-\frac{1}{2}\tilde{\gamma}) + \frac{1}{2}$ is strictly positive for $\tilde{\gamma} > 0$, we are done. q.e.d.

Proof (Lemma 10). We have the following:

$$\frac{\partial \mu_H^I}{\partial s_0} = -\frac{(1 - [p_1 + \delta p_l])}{s_h} \mu_H^I, \quad (21)$$

$$\frac{\partial \mu_L^I}{\partial s_0} = \frac{(1 - [p_1 + \delta p_l])}{1 - s_h} \mu_L^I, \quad (22)$$

$$\frac{\partial \mu_H^I}{\partial s_1} = \frac{[p_1 + \delta p_l]}{s_h} (1 - \mu_H^I), \quad (23)$$

$$\frac{\partial \mu_L^I}{\partial s_1} = -\frac{[p_1 + \delta p_l]}{1 - s_h} (1 - \mu_L^I). \quad (24)$$

It can easily be checked that μ_L^I , μ_H^I and s_h are differentiable in (s_0, s_1) . This and differentiability of $u(\mu)$ imply that also $U_h^Y(\alpha_l, 1)$ is differentiable in (s_0, s_1) . We get:

$$\frac{\partial}{\partial s_\omega} U_h^Y(0, \alpha_h) = \frac{\partial s_h}{\partial s_\omega} [u_H - u_L] + (1 - s_h) \frac{\partial \mu_L^I}{\partial s_\omega} u'_L + s_h \frac{\partial \mu_H^I}{\partial s_\omega} u'_H. \quad (25)$$

Consider first $\omega = 0$: Inserting (21) and (22) in (25), we obtain after rearranging:

$$\frac{\partial U_h^Y}{\partial s_0} = (1 - [p_1 + \delta p_l]) \{ [u_H - \mu_H^I u'_H] - [u_L - \mu_L^I u'_L] \}.$$

Note that the derivative of $u(\mu) - \mu u'(\mu)$ with respect to μ is $-\mu u''(\mu)$ which is strictly positive for $\mu > 0$ by strict concavity of $u(\mu)$. This implies $\frac{\partial U_h^Y}{\partial s_0} > 0$.

Consider now $\omega = 1$. Inserting (23) and (24) in (25), we obtain:

$$\frac{\partial U_h^Y}{\partial s_1} = [p_1 + \delta p_l] \{ [u_H + (1 - \mu_H^I) u'_H] - [u_L - (1 - \mu_L^I) u'_L] \}.$$

Note that the derivative of $u(\mu) + (1 - \mu)u'(\mu)$ with respect to μ is $(1 - \mu)u''(\mu)$ which is strictly negative for $\mu < 1$ by strict concavity of $u(\mu)$. Hence, $\frac{\partial U_h^Y}{\partial s_1} < 0$. q.e.d.

Proof (Lemma 11). We first construct $\widehat{\delta}(p_1, p_h)$ and show then that it has the desired properties. For the perfect device $(s_0, s_1) = (0, 1)$ we have $U_h^Y(0, 1) = (1 - [p_1 + \delta p_l])u(0) + [p_1 + \delta p_l]u(1)$ and $U^N(0, 1) = u(p_1 - \delta p_h)$. While the former function is strictly increasing in δ , the latter one is strictly decreasing in δ such that there exists at most one point of intersection. If we can show that we have $U^N(0, 1) > U_h^Y(0, 1)$ for $\delta \rightarrow 0$ and $U^N(0, 1) < U_h^Y(0, 1)$ for $\delta \rightarrow \delta_{\max}$, we can apply the Intermediated Value Theorem (since $U^N(0, 1)$ and $U_h^Y(0, 1)$ are both continuous) to obtain existence of a point of intersection. Note that $\lim_{\delta \rightarrow 0} U_h^Y(0, 1) = (1 - p_1)u(0) + p_1u(1)$ which is strictly smaller than $u(p_1)$ by strict concavity of $u(\mu)$, and note that $\lim_{\delta \rightarrow 0} U^N(0, 1) = u(p_1)$. Hence, $U^N(0, 1) > U_h^Y(0, 1)$ for $\delta \rightarrow 0$. We have to distinguish two cases when we consider $\delta \rightarrow \delta_{\max}$: Case 1: $p_1 \leq p_h$ implying $\delta_{\max} = \frac{p_1}{p_h}$. In this case we have $\lim_{\delta \rightarrow \delta_{\max}} U^N(0, 1) = u(0)$ which is strictly smaller than $\lim_{\delta \rightarrow \delta_{\max}} U_h^Y(0, 1) = (1 - \frac{p_1}{p_h})u(0) + \frac{p_1}{p_h}u(1)$. Case 2: $p_1 > p_h$ implying $\delta_{\max} = \frac{1 - p_1}{1 - p_h}$. In this case we have $\lim_{\delta \rightarrow \delta_{\max}} U_h^Y(0, 1) = 0u(0) + 1u(1)$ which is strictly larger than $\lim_{\delta \rightarrow \delta_{\max}} U^N(0, 1) = u(\frac{p_1 - p_h}{1 - p_h})$. From the above we can conclude that for any p_1 and for any p_h there exists a unique $\delta \in (0, \delta_{\max})$ such that $U^N(0, 1) = U_h^Y(0, 1)$ and we will refer to it henceforth as $\widehat{\delta}(p_1, p_h)$.

(i) Assume to the contrary that $\delta < \widehat{\delta}(p_1, p_h)$ and that $\alpha_h > 0$. By construction of $\widehat{\delta}(p_1, p_h)$, we have $U^N(0, 1) > U_h^Y(0, 1)$. Since $U_h^Y(0, \alpha_h)$ does not depend on α_h and since $U^N(0, \alpha_h)$ is strictly decreasing in α_h , it follows $U^N(0, \alpha_h) > U_h^Y(0, \alpha_h)$ for any α_h . Hence, incentive compatibility is violated for $\alpha_h > 0$. Contradiction.

(ii) By construction of $\widehat{\delta}(p_1, p_h)$, $U^N(0, 1) \leq U_h^Y(0, 1)$. Hence, $\alpha_h = 1$ is consistent with incentive compatibility and it remains only to check that $\alpha_l = 0$ is also consistent. We have $U_l^Y(0, 1) = (1 - [p_1 - \delta p_h])u(0) + [p_1 - \delta p_h]u(1) < u(p_1 - \delta p_h)$ by strict concavity of $u(\mu)$. Since, $U^N(0, 1) = u(p_1 - \delta p_h)$, we obtain $U^N(0, 1) > U_l^Y(0, 1)$ and are done. q.e.d.

Proof (Lemma 12). For any $\alpha_h \in [0, 1]$ we have $U^N(0, \alpha_h) \in [u(p_1 - \delta p_h), u(p_1)]$. Moreover, $U^N(0, \alpha_h)$ does not depend on the device. $(s_0, s_1) \rightarrow (1, 1)$ implies $\mu_L^I \rightarrow p_1 + \delta p_l$ and $\mu_H^I \rightarrow p_1 + \delta p_l$ such that $U_h^Y(0, \alpha_h) \rightarrow u(p_1 + \delta p_l) > U^N(0, \alpha_h)$. Thus, an agent with a high private signal has a strict incentive to use a sufficiently uninformative device. By contrast, $\delta < \widehat{\delta}(p_1, p_h)$ and Lemma 11 imply $U_h^Y(0, \alpha_h) < U^N(0, \alpha_h)$ for $(s_0, s_1) = (0, 1)$, i.e. an agent with a high private signal has a strict incentive not to use the perfect device. Since $U_h^Y(0, \alpha_h)$ is continuous in the device by Lemma 10, we can apply the Intermediate Value Theorem to obtain that there exists a $\underline{s}_1 \in (0, 1)$ such that $U^N(0, \alpha_h) = U_h^Y(0, \alpha_h)$ for device $(0, \underline{s}_1)$. Since $U_h^Y(0, \alpha_h)$ is strictly increasing in s_1 (Lemma 10), \underline{s}_1 is unique.

It follows directly from Lemma 10 that there exists a connected and closed domain $[0, \overline{s}_0]$ and a continuous and increasing function $s_1(s_0)$ on this domain such that $U_h^Y(0, \alpha_h) = U^N(0, \alpha_h)$ if and only if $s_0 \in [0, \overline{s}_0]$ and $s_1 = s_1(s_0)$. By construction, $\underline{s}_1 = s_1(0) \in (0, 1)$. Since an agent with a high private signal has a strict incentive to use the uninformative device $(s_0, s_1) = (1, 1)$, there exists no device for $s_0 = 1$ for which the participation constraint is binding. Hence, $\underline{s}_1 < 1$.

Since $s_1(s_0)$ is monotonous, it is differentiable almost everywhere. We obtain $s_1'(s_0) = -\frac{\partial U_h^Y}{\partial s_0} / \frac{\partial U_h^Y}{\partial s_1} > 0$ for any s_0 in the domain. q.e.d.

Proof (Lemma 13). Since in any equilibrium with $\alpha_l = 0$ an agent with a low private signal does not use the device, $U_l(0, \alpha_h) = U^N(0, \alpha_h)$.

(i) By construction of the function $s_1(s_0)$ (Lemma 12), the participation constraint for an agent with a high private signal is binding for devices $(s_0, s_1(s_0))$. Hence, $U_h^Y(0, \alpha_h) = U^N(0, \alpha_h)$. Since an agent with a high private signal also obtains utility $U^N(0, \alpha_h)$ when he does not use the device, it follows $U_h = U^N(0, \alpha_h)$.

(ii) By construction of $\widehat{\delta}(p_1, p_h)$ (see the first part of the proof to Lemma 11), we have $U_h^Y(0, 1) > U^N(0, 1)$ for any $\delta > \widehat{\delta}(p_1, p_h)$. q.e.d.

Proofs of Section 7

Proof (Lemma 14). Let $u(\mu) = b_0 + b_1\mu - b_2\mu^2$. Then, $\mathbf{E}[u(\mu)] = b_0 + b_1\mathbf{E}[\mu] - b_2\mathbf{E}[\mu^2]$. Bayes' Law implies $\mathbf{E}[\mu] = p_1$ and strict concavity of $u(\mu)$ requires $b_2 > 0$ such that $\mathbf{E}[u(\mu)]$ is minimized iff $\mathbf{E}[\mu^2]$ is maximized. Let $v(\mu) = a_0 + a_1\mu + a_2\mu^2$. Then, $\mathbf{E}[v(\mu)] = a_0 + a_1\mathbf{E}[\mu] + a_2\mathbf{E}[\mu^2]$. Bayes' Law implies $\mathbf{E}[\mu] = p_1$ and strict convexity of $v(\mu)$ requires $a_2 > 0$ such that $\mathbf{E}[v(\mu)]$

is maximized iff $\mathbf{E}[\mu^2]$ is maximized. Hence, maximizing $\mathbf{E}[v(\mu)]$ is equivalent to minimizing $\mathbf{E}[u(\mu)]$. q.e.d.

Proof (Proposition 1). (i) Only for equilibria with $\alpha_l = 1$ it is possible that the participation constraint is not binding. Consider the case with $\alpha_l = 1$ and assume to the contrary that is optimal for the designer to have the constraint not binding. By Lemma 3, there exist equilibria with a higher s_1 and/or a smaller s_0 which implement the same participation behavior. For both adjustments, μ_H^{II} increases and μ_L^{II} decreases with at least one effect being strict (see the inequalities in (1), (2), (3) and (4)). I.e., the beliefs that may realize become more extreme. The designer's utility from this kind of equilibrium is $V = p_L v(\mu_L) + p_H v(\mu_H)$. By Bayes' Law $(1 - p_H)\mu_L + p_H\mu_H = p_1$ implying $p_H = \frac{p_1 - \mu_L}{\mu_H - \mu_L}$. Hence, $V = \frac{\mu_H - p_1}{\mu_H - \mu_L} v(\mu_L) + \frac{p_1 - \mu_L}{\mu_H - \mu_L} v(\mu_H)$ and we get

$$\begin{aligned}\frac{\partial V}{\partial \mu_H} &= \frac{p_1 - \mu_L}{\mu_H - \mu_L} \left\{ v'(\mu_H) - \frac{v(\mu_H) - v(\mu_L)}{\mu_H - \mu_L} \right\} \\ \frac{\partial V}{\partial \mu_L} &= \frac{\mu_H - p_1}{\mu_H - \mu_L} \left\{ v'(\mu_L) - \frac{v(\mu_H) - v(\mu_L)}{\mu_H - \mu_L} \right\}.\end{aligned}$$

By strict convexity of $v(\mu)$, $\frac{\partial V}{\partial \mu_H} > 0$ and $\frac{\partial V}{\partial \mu_L} < 0$, i.e. the designer's utility increases if beliefs become more extreme. Contradiction. We can conclude that the participation constraint must be binding.

(ii) Knowing that the participation constraint must be binding, we can apply Lemma 8 to obtain that under Assumptions 1 and 2 $\mathbf{E}[u(\mu)]$ increases strictly with s_0 . Since under Assumption 3 the designer strives for minimizing the agent's utility (see (7)), $s_0 = 0$ is optimal for the designer. q.e.d.

Proof (Proposition 2). By Proposition 1, (i) $U_l^Y(\alpha_l, 1) = U^N(\alpha_l, 1)$ and (ii) $s_0 = 0$ is true for the optimal device. Note that (ii) implies $\mu_H^{II} = 1$. Moreover, note that (i) and $U^N(\alpha_l, 1) = u(p_1 - \delta p_h)$ imply that $U_l^Y(\alpha_l, 1)$ neither depends on α_l nor on the device choice. This and Assumption 3 imply that the designer's problem simplifies to

$$\begin{aligned}\min_{\alpha_l \in (0,1]} \quad & U_h^Y(\alpha_l, 1) = (1 - s_h)u(\mu_L^{II}) + s_h u(1) \\ \text{s.t.} \quad & U_l^Y(\alpha_l, 1) = (1 - s_l)u(\mu_L^{II}) + s_l u(1) \stackrel{!}{=} u(p_1 - \delta p_h).\end{aligned}$$

Since we can apply the Implicit Function Theorem to the constraint, sufficient for the optimality of $\alpha_l = 1$ is $\frac{dU_h^Y}{d\alpha_l} = \frac{\partial U_h^Y}{\partial \alpha_l} - \left[\frac{\partial U_l^Y}{\partial \alpha_l} \right] / \left[\frac{\partial U_l^Y}{\partial s_1} \right] \frac{\partial U_h^Y}{\partial s_1} = \left[-\frac{\partial U_h^Y}{\partial \alpha_l} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial U_h^Y}{\partial s_1} \frac{\partial U_l^Y}{\partial \alpha_l} \right] / \left[-\frac{\partial U_l^Y}{\partial s_1} \right] < 0$. Since we already know that the numerator is positive by Lemma 3, it suffices to show that $-\frac{\partial U_h^Y}{\partial \alpha_l} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial U_h^Y}{\partial s_1} \frac{\partial U_l^Y}{\partial \alpha_l} < 0$.

$$\begin{aligned}& -\frac{\partial U_h^Y}{\partial \alpha_l} \frac{\partial U_l^Y}{\partial s_1} + \frac{\partial U_h^Y}{\partial s_1} \frac{\partial U_l^Y}{\partial \alpha_l} \\ &= \left[(1 - s_l) \frac{\partial s_h}{\partial s_1} - (1 - s_h) \frac{\partial s_l}{\partial s_1} \right] \frac{\partial \mu_L^{II}}{\partial \alpha_l} u'(\mu_L^{II}) (u(1) - u(\mu_L^{II}))\end{aligned}$$

$$\begin{aligned}
& + [(1-s_l)(1-s_h) - (1-s_l)(1-s_h)] \frac{\partial \mu_L^{II}}{\partial \alpha_l} u'(\mu_L^{II}) \frac{\partial \mu_L^{II}}{\partial s_1} u'(\mu_L^{II}) \\
& = \delta \frac{\partial \mu_L^{II}}{\partial \alpha_l} u'(\mu_L^{II}) [u(1) - u(\mu_L^{II})].
\end{aligned}$$

This expression is strictly negative if $\frac{\partial \mu_L^{II}}{\partial \alpha_l} = \mu'_Y(\alpha_l) \frac{1-s_1}{p_H^2 p_Y}$ is strictly negative. Since $s_1(0) < 1$ by Lemma 4, this is true if $\mu'_Y(\alpha_l)$ is strictly negative. Hence, $\mu'_Y(\alpha_l) = -\frac{\delta p_h p_l}{[p_h + p_l \alpha_l]^2}$ concludes the proof. q.e.d.

Proof (Proposition 3). Due to Assumption 3, the designer chooses device (s_0, s_1) and participation behavior (α_l, α_h) to minimize $\tilde{V} := p_l U^N(0, \alpha_h) + p_h [(1-\alpha_h)U^N(0, \alpha_h) + \alpha_h U_h^Y(0, \alpha_h)]$.

By Lemma 11, device $(s_0, s_1) = (0, 1)$ implements participation behavior $(\alpha_l, \alpha_h) = (0, 1)$. We now show that this is indeed optimal for the designer. The strategy of proof is as follows: We ignore incentive compatibility and show that for any device \tilde{V} can be increased by increasing α_h , and for any participation behavior \tilde{V} can be increased by decreasing s_0 and by increasing s_1 . This proves that \tilde{V} evaluated at $\alpha_h = 1$ and at $(s_0, s_1) = (0, 1)$ is an upper bound on what the designer can achieve. Since this is actually implementable, we are done.

Note that $U^N(0, \alpha_h)$ does not depend (directly) on the device and that $U_h^Y(0, \alpha_h)$ does not depend (directly) on the participation behavior. It follows directly from Lemma 10 that for a given α_h we have $\frac{\partial \tilde{V}}{\partial s_0} < 0$ and $\frac{\partial \tilde{V}}{\partial s_1} > 0$. Moreover,

$$\begin{aligned}
\frac{\partial \tilde{V}}{\partial \alpha_h} & = p_h [U_h^Y(0, \alpha_h) - U^N(0, \alpha_h)] + [1 - p_h \alpha_h] \frac{\partial U^N(0, \alpha_h)}{\partial \alpha_h} \\
& = p_h \left[\frac{p_L}{\alpha_h p_h} u(\mu_L^I) + \frac{p_H}{\alpha_h p_h} u(\mu_H^I) - u(\mu_N^I) \right] + [1 - p_h \alpha_h] \frac{\partial \mu_N^I}{\partial \alpha_h} u'(\mu_N^I).
\end{aligned}$$

Using that $\frac{p_L}{\alpha_h p_h} u(\mu_L^I) + \frac{p_H}{\alpha_h p_h} u(\mu_H^I) < u(\frac{p_h 1}{p_h}) = u(p_1 + \delta p_l)$ by strict concavity of $u(\cdot)$ and that $\frac{\partial \mu_N^I}{\partial \alpha_h} = -\frac{\delta p_l p_h}{[1 - \alpha_h p_h]^2}$, we obtain

$$\frac{\partial \tilde{V}}{\partial \alpha_h} < p_h [u(p_1 + \delta p_l) - u(\mu_N^I)] - \frac{\delta p_l p_h}{1 - \alpha_h p_h} u'(\mu_N^I).$$

Using now that $[p_1 + \delta p_l] - \mu_N^I = \frac{\delta p_l}{1 - \alpha_h p_h}$, we get

$$\frac{\partial \tilde{V}}{\partial \alpha_h} < \frac{\delta p_l p_h}{1 - \alpha_h p_h} \left\{ \frac{u(p_1 + \delta p_l) - u(\mu_N^I)}{[p_1 + \delta p_l] - \mu_N^I} - u'(\mu_N^I) \right\} < 0.$$

Thereby the last inequality follows from strict concavity of $u(\cdot)$. q.e.d.

Proof (Proposition 4). For $\alpha_h = 0$ the result is obvious as the designer's utility does then not depend on the device. Note that for $\alpha_h \in (0, 1)$ it is only possible to induce participation with $U^N(0, \alpha_h) = U_h^Y(0, \alpha_h)$, while for $\alpha_h = 1$ also $U_l^Y(0, 1) \leq U^N(0, 1) < U_h^Y(0, 1)$ is possible. However, since, by Assumption 3, the designer is interested in minimizing $U_h^Y(0, 1)$, he has by Lemma 10 a strict incentive to decrease s_0 and / or to increase s_1 until $U^N(0, 1) = U_h^Y(0, 1)$. By Lemma 11 and assumption $\delta < \hat{\delta}(p_1, p_h)$, it is always possible to get this condition binding and by

Lemma 1 this is possible without violating the incentive of the agent to not use the device when his private signal is low. Hence, we can restrict attention to the case where the participation constraint of the agent is binding when his private signal is high. However, in this case the designer's utility is completely pinned down by the agent's utility from not using the device, $U^N(0, \alpha_h)$, which does not depend on the device choice. As a consequence, the designer obtains the same expected utility from any device which satisfies the participation constraint. q.e.d.

Proof (Proposition 5). From Proposition 4 and Assumption 3 it follows that the designer chooses participation behavior α_h to maximize $\tilde{V} := -U^N(0, \alpha_h)$. Since $U^N(0, \alpha_h)$ strictly decreases in α_h , $\alpha_h = 1$ is optimal for the designer. q.e.d.

Proof (Proposition 6). Consider first $\delta < \hat{\delta}(p_1, p_h)$. Proposition 4 and Proposition 5 imply that if the designer wants to induce a participation behavior with $\alpha_l = 0$, then $\alpha_h = 1$ is optimal and the participation constraint is binding when the agent's private signal is high. This implies

$$V^I = -C_2 U^N(0, 1) + C_1 p_1 + C_0.$$

Proposition 1 and Proposition 2 imply that if the designer wants to induce a participation behavior with $\alpha_l > 0$, then $s_0 = 0$ and $\alpha_l = 1$ are optimal and the participation constraint is binding when the agent's private signal is low. Moreover, Lemma 1 and $U_l^Y(1, 1) = U^N(1, 1)$ imply $U_h^Y(1, 1) > U^N(1, 1)$. Hence,

$$V^{II} < -C_2 U^N(1, 1) + C_1 p_1 + C_0.$$

Since $U^N(0, 1) = U^N(1, 1)$, $V^I > V^{II}$. I.e., the designer strictly prefers the equilibrium with perfect separation of private information to the equilibrium with perfect pooling of private information.

Consider now $\delta > \hat{\delta}(p_1, p_h)$. It follows from Proposition 2 and Proposition 3 that either $(\alpha_l, \alpha_h) = (1, 1)$ or $(\alpha_l, \alpha_h) = (0, 1)$ is optimal. q.e.d.

References

- ANDERSON, A., AND L. SMITH (2010): "Dynamic Matching and Evolving Reputations," *Review of Economic Studies*, 77, 3–29.
- BURDETT, K., AND D. T. MORTENSEN (1981): "Testing for Ability in a Competitive Labor Market," *Journal of Economic Theory*, 25, 42–66.
- EECKHOUT, J., AND X. WENG (2009): "Assortative Learning," Working Paper.
- FARHI, E., J. LERNER, AND J. TIROLE (2005): "Certifying new technologies," *Journal of the European Economic Association*, 3, 734–744.
- GEANAKOPOLOS, J., AND D. PEARCE (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60–79.

- GERLING, K., H. P. GRÜNER, A. KIEL, AND E. SCHULTE (2005): “Information acquisition and decision making in committees: A survey,” *European Journal of Political Economy*, 21, 563–597.
- GROSSMAN, S. J., AND O. D. HART (1980): “Disclosure Laws and Takeover Bids,” *The Journal of Finance*, 35(2), 323–334.
- HIRSHLEIFER, J. (1971): “The Private and Social Value of Information and the Reward to Inventive Activity,” *The American Economic Review*, 61(4), 561–574.
- KAMENICA, E., AND M. GENTZKOW (2009): “Bayesian Persuasion,” NBER Working Paper Series Working Paper 15540.
- LERNER, J., AND J. TIROLE (2006): “A model of forum shopping,” *American Economic Review*, 96(4), 1091–1113.
- MACDONALD, G. M. (1980): “Person-Specific Information in the Labor Market,” *The Journal of Political Economy*, 88(3), 578–597.
- MILGROM, P. R. (1981): “Good News and Bad News: Representation Theorems and Applications,” *The Bell Journal of Economics*, 12(2), 380–391.
- OKUNO-FUJIWARA, M., A. POSTLEWAITE, AND K. SUZUMURA (1990): “Strategic Information Revelation,” *The Review of Economic Studies*, 57(1), 25–47.
- OSTROVSKY, M., AND M. SCHWARZ (2008): “Information Disclosure and Unraveling in Matching Markets,” forthcoming in *American Economic Journal: Microeconomics*.
- PRESCOTT, E. C., AND M. VISSCHER (1980): “Organization Capital,” *The Journal of Political Economy*, 88(3), 446–461.
- STIGLITZ, J. E. (1975): “The Theory of “Screening,” Education, and the Distribution of Income,” *The American Economic Review*, 65(3), 283–300.
- WALDMAN, M. (1989): “Information on Worker Ability, an Analysis of Investment Within the Firm,” *Information Economics and Policy*, 4(1), 57–80.