Test design under voluntary participation and conflicting preferences☆

(This version: July 11, 2014)

Frank Rosar[a]

[a] *University of Bonn*

Abstract

At issue is the design of a test which can generate a signal about an agent's quality when the agent is imperfectly informed and his participation in the test is voluntary. The designer strives for information generation, whereas the agent strives for a high quality perception and is averse to information generation. When quality is binary, the optimal test is binary and not subject to false positives. For an extended setting with non–binary quality types, a possibly non–binary adapted version of such a test is optimal.

JEL classification: D82, D83

Keywords: test design, Bayesian learning, persuasion, asymmetric information, voluntary participation, false positives, false negatives, signaling, pooling

1. Introduction

This article studies how to design a test in order to learn optimally about the quality of an imperfectly informed agent who is averse to the perception risk which comes along with testing and whose participation in the test is voluntary. In such a framework, participation can be fostered by employing a test which is subject to errors. This article explains *how* the test has to be made subject to errors in order to achieve optimal learning.

Possible applications include learning through the educational system, medical testing and financial stress testing (see Section 7 for a detailed discussion). At issue is an agent's "quality" like the productivity of an individual, the health status of a patient or the viability of a bank. Information generation is possible through testing, testing is however often either implicitly or explicitly voluntary. In some situations indirect learning through the agent's participation decision is explicitly desired, but the inducement of full participation is demanded in others.

☆This paper elaborates ideas partly originating from joint work with Elisabeth Schulte under the title "Optimal test design under imperfect private information and voluntary participation". Therefore, I am deeply indebted to Elisabeth. Furthermore, I would like to thank Kfir Eliaz, Hans Peter Grüner, Emir Kamenica, Johannes Koenen, Benny Moldovanu, Tymofiy Mylovanov, Marco Ottaviani, Martin Peitz, Sergei Severinov, Konrad Stahl, Nora Szech, and various seminar participants. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Email address: email@frankrosar.de (Frank Rosar)

This gives rise to the following kind of questions: Is optimal learning achieved by a test which is accurate? If not, what is it optimal to test for? Is it optimal to test for high or low productivity, healthiness or illness, viability or non–viability? Is it optimal to induce full participation?

Most closely related to this article is independent work by Harbaugh and Rasmusen (2013). Like us, they study the design of a test to maximize the generation of information when a constraint derives from voluntary participation. In contrast to us, they model the agent differently such that different incentive problems arise. The agent is perfectly informed, perception–risk neutral and suffers an exogenously given constant cost from participation. Perception risk considerations which are the central theme in our article and through which a role for the generation of inaccurate information arises are mute in theirs. We defer a detailed discussion of the related literature to Section 6.

In the first part of this article, we consider the problem where the quality variable in question is binary, say good or bad, and where the design of the information generation technology is not subject to any technical limitations.[1] The class of games in which we are interested in consists of three stages: First, a principal designs a test. The chosen test design is publicly observable. Second, an agent who is in possession of an imperfect private signal about his probability of being good decides upon participation in this test. If he participates, a test result is generated and becomes publicly observable, otherwise, his non–participation can be inferred.[2] Third, a receiver updates her belief about the agent's quality and takes an action. We pursue a reduced–form modeling approach in which the receiver's action is not modelled explicitly. The quality perception which derives from Bayesian updating directly determines the agent's and the principal's payoffs.[3] The agent strives for a high quality perception and is averse to perception risk. The principal strives for information generation as measured by the variance of the posterior distribution of quality perceptions.

Participation in any informative test involves the generation of information to which the agent is averse. As better private information makes the generation of favorable test results more likely, only equilibria which exhibit threshold participation behavior can exist. Participation signals thus in any equilibrium favorable private information. Perfect learning of the agent's quality is generally impossible as the imperfection of the agent's private information precludes full unraveling under an accurate test. The test design simultaneously determines participation incentives and learning in a non–trivial way.

We first study optimal learning subject to the inducement of a given participation threshold.

[1]There is also no cost associated to test accuracy. This allows us to focus on the usage of inaccuracies for strategic reasons.

[2]We are interested in the case where generated information is always publicly disclosed. Matthews and Postlewaite (1985) compare the incentives to generate information through an accurate test for the cases with voluntary and with mandatory disclosure of generated information. See also Farhi et al. (2013).

[3]See Matthews and Postlewaite (1985) for a similar modeling approach. The agent's utility depends in the reduced game directly on his belief about the others' belief about his quality. It can thus be interpreted as a particular psychological game as introduced in Geanakoplos et al. (1989).

For this it is useful to interpret tests in terms of "structure" and "accuracy". For example, the structure of a test might be that it is binary (which allows for an interpretation as a "pass–fail test") and not subject to false positives (that is, the agent might only pass when he is good). The accuracy is then the likelihood of correct positives (that is, the probability with which the agent passes when he is good). We find that for any primitives of the model any inducible participation threshold is optimally induced by a binary test which is not subject to false positives.

A rough intuition is the following: Binary tests which are not subject to false positives impose no perception risk on a hypothetical agent who is certain to be bad and therewith the smallest possible perception risk on an agent who is relatively likely to be bad like the threshold agent. As learning about the quality of an imperfectly informed agent corresponds basically to imposing a perception risk on him, the optimal test structure induces relatively less learning about the threshold agent (whose participation constraint must be respected) than on agents with higher private signals (whose participation constraints are slack). Although such a property sounds intuitive at first glance, it is surprising that the result relies indeed only on perception risk considerations. This is in particular the case because false positives improve how the threshold agent is perceived on average besides increasing the perception risk that he faces. It is thus a priori not clear whether participation of the threshold agent is better induced by decreasing his perception risk or by increasing his expected perception.

The knowledge about the optimal test structure reduces the test design problem to the determination of the optimal test accuracy. When the inducement of full participation is desired for exogenous reasons, the participation constraint determines the test accuracy. Otherwise, optimal learning might rely also on indirect learning through the agent's (non–)participation decision. For the uniform–quadratic case, we establish that the principal faces a trade–off between test accuracy and participation, and we discuss how the resolution of this trade–off and the importance of indirect learning depend on the accuracy of the agent's private information.

In the second part of this article, we consider the problem where the quality variable of interest is still binary, but where the test design is subject to technical limitations. More specifically, tests are only capable of generating information about non–binary "quality types" which reveal stochastic information about the agent's true quality. In the stress testing context, this means that the bank is ultimately still either viable or not, but that it might not be possible to perfectly reveal the bank's viability at the time the test is conducted. There might for instance be three possible quality types: First, the bank is definitely viable. Second, the bank is definitely not viable. Third, the bank is only viable if the economy is sufficiently stable/no other bank breaks down. The third type allows only for the imperfect inference that the bank is viable with an intermediate probability.[4] We show how our techniques from the problem

[4] Although the inference from the third type is in the described example exogenous, it might also derive endogenously from an extended model in which multiple banks are modeled explicitly. See Li and Li (2013) for a related modeling approach in the context of political campaigning.

without technical limitations can be used to study the more general problem with non–binary quality types. As the same forces are at work, the optimal test can be interpreted as a possibly non–binary generalization of a binary test which is not subject to false positives. In particular, if the agent's signaling incentive is sufficiently weak, a pass–fail test where the agent passes with a certain probability when he has the best quality type and where he fails otherwise is optimal.

This article is organized as follows: We introduce our model in Section 2 and we use an example to motivate trade–offs and research questions in Section 3. The analysis of the optimal test design is conducted in Section 4. In Section 5 we study the extended model with non–binary quality types. Section 6 relates our article to the literature and Section 7 discusses our modeling assumptions and our results in the light of different applications. Finally, we conclude in Section 8. All proofs are relegated to the appendix.

2. The model

There is a principal (she) and an agent (he). The agent's quality is either good ($\omega = g$) or bad ($\omega = b$), but he is only imperfectly informed about his quality. He privately knows that he is good with probability θ. θ is distributed according to a cumulative distribution function F with a strictly positive density f on an interval support $[\underline{\theta}, \overline{\theta}]$ with $0 < \underline{\theta} < \overline{\theta} < 1$. The assumption on the support of θ describes our notion of imperfection of private information: As $\underline{\theta} < \overline{\theta}$, the agent is endowed with some meaningful information about his quality. As $\overline{\theta} < 1$ and $\underline{\theta} > 0$, he is never certain to be good or bad.[5] The prior probability with which the agent is good is $\theta_0 := \mathbf{E}_\theta[\theta]$.

Information about the agent's quality can be generated through a test. A test $T = (p_b, p_g)$ is characterized by a number $Z \in \{2, 3, \ldots\}$ and conditional probability vectors $p_\omega := (p_\omega^1, \ldots, p_\omega^Z) \in \Delta^{Z-1}$. Z determines the set of possible test results $\Sigma := \{1, \ldots, Z\}$. p_ω^σ describes the probability with which test result σ is generated when the test is used by an agent with quality ω. σ and θ are independent conditional on ω.[6] Let $\mathcal{T}$ be the set of all informative tests (that is, $p_b \neq p_g$) where each test result occurs with a strictly positive probability (that is, $p_b^\sigma + p_g^\sigma > 0$ for all σ). Important subclasses of $\mathcal{T}$ are tests with Z test results, $\mathcal{T}^Z := \{(p_b, p_g) \in \mathcal{T} | (p_b, p_g) \in \Delta^{Z-1} \times \Delta^{Z-1}\}$, and accurate tests, $\mathcal{T}_a := \{(p_b, p_g) \in \mathcal{T} | \forall \sigma : p_b^\sigma = 0 \text{ or } p_g^\sigma = 0\}$.

The timing is as follows (see also Figure 1): First, the principal designs a test $T = (p_b, p_g) \in \mathcal{T}$ and the test design becomes publicly observable. Second, nature draws the agent's quality ω and his information about his quality θ. Third, the agent learns θ and chooses a participation probability. We denote his participation strategy by $x \in \mathcal{X} := \{x : [\underline{\theta}, \overline{\theta}] \to [0, 1]\}$. Fourth, if

[5]If $\underline{\theta} = 0$, the test design problem is trivial as there is full unraveling under an accurate test. See the discussion in Subsection 2.3. By assuming $\underline{\theta} > 0$, we restrict attention to the interesting cases. The assumption $\overline{\theta} < 1$ simplifies only the exposition.

[6]That is, the test can reveal information about the agent's quality, but it cannot reveal information about what the agent thinks about his quality which goes beyond what can already be inferred from the knowledge of his quality.

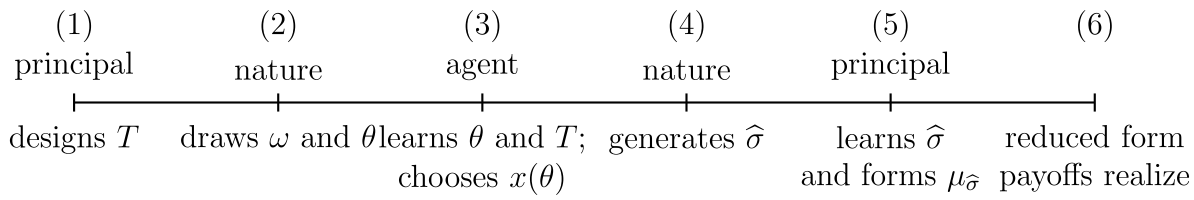


Figure 1: Timing of the reduced game

the agent participates, a test result $\sigma \in \Sigma$ is generated according to p_ω and the signal $\widehat{\sigma} = \sigma$ becomes publicly observable. Otherwise, $\widehat{\sigma} = N$ can be inferred. Let $\widehat{\sigma} \in \widehat{\Sigma} := \Sigma \cup \{N\}$. Fifth, a receiver updates her belief about the agent's quality according to a procedure which we describe below. For the reduced problem in which we are interested in, we can assume without loss of generality that the role of the receiver is assumed by the principal. We denote the probability with which she believes the agent to be good by $\mu_{\widehat{\sigma}} \in [0, 1]$ and we refer to $\mu_{\widehat{\sigma}}$ henceforth as quality perception. Let $\mu := (\mu_N, \mu_1, \ldots, \mu_Z)$. Sixth, the reduced game ends and payoffs which depend directly on $\mu_{\widehat{\sigma}}$ realize.

The agent is weakly prudent. He evaluates $\mu_{\widehat{\sigma}}$ at a thrice differentiable utility function $u : [0, 1] \to \mathbf{R}$ with $u' > 0$, $u'' < 0$ and $u''' \geq 0$ almost everywhere.[7] If he does not participate, he obtains the certain utility $u(\mu_N)$. If he participates, he obtains an uncertain utility with expected value $U_\theta(T, \mu) := \sum_{\sigma \in \Sigma} p_\theta^\sigma u(\mu_\sigma)$ where $p_\theta := (1 - \theta)p_b + \theta p_g$. The principal evaluates $\mu_{\widehat{\sigma}}$ at the convex utility function $v(\mu_{\widehat{\sigma}}) = (\theta_0 - \mu_{\widehat{\sigma}})^2$. Her ex ante expected utility is $V(T, x, \mu) := \mathbf{E}_\theta[x(\theta) \sum_{\sigma \in \Sigma} p_\theta^\sigma v(\mu_\sigma) + (1 - x(\theta))v(\mu_N)]$.

2.1. The updating procedure

There are two sources of information available to the principal: information about the agent's private signal revealed through his participation decision and, if the test is used, information about his quality revealed through the test result. Suppose first the principal uses only the agent's supposed participation strategy x and his actual participation decision to draw inferences about his quality. When quality perceptions are formed according to Bayes' Law whenever possible and when the principal believes that the agent has the lowest (resp. highest) private signal when she observes non–participation (resp. participation) although the agent is supposed to participate with probability one (resp. zero),[8] the quality perceptions associated

[7] The assumptions that $u' > 0$ a.e. and that $u'' < 0$ a.e. are essential for the reduced problem in which we are interested in. The assumption that $u''' \geq 0$ a.e. is sufficient but not necessary for deriving our results. We impose it as it simplifies our proofs and as it is satisfied for a broad class of HARA utility functions including quadratic utility, cubic utility, exponential utility/CARA utility, logarithmic utility and CRRA utility. See the appendix for a parametrization of HARA utility and a graphical illustration of the parameter space and the special cases therein.

[8] Alternatively, we could assume that when the principal observes a participation decision that is supposed to occur with probability zero, she believes that the agent has the private signal which gives him the strongest incentive to choose this decision. The assumed property follows then from Lemma 1 (a) below.

to non–participation and participation are

$$\mu_N(x) := \begin{cases} \frac{\mathbf{E}_\theta[(1-x(\theta))\theta]}{\mathbf{E}_\theta[(1-x(\theta))]} & \text{if } \mathbf{E}_\theta[x(\theta)] < 1 \\ \underline{\theta} & \text{if } \mathbf{E}_\theta[x(\theta)] = 1 \end{cases} \text{ and } \mu_Y(x) := \begin{cases} \overline{\theta} & \text{if } \mathbf{E}_\theta[x(\theta)] = 0 \\ \frac{\mathbf{E}_\theta[x(\theta)\theta]}{\mathbf{E}_\theta[x(\theta)]} & \text{if } \mathbf{E}_\theta[x(\theta)] > 0 \end{cases},$$

respectively. As imperfection of the agent's private information implies $\mu_Y(x) \in (0,1)$ for any $x \in \mathcal{X}$, Bayes' Law is always applicable to process the additional information revealed by the test result σ:

$$\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma) := \frac{\mu_Y(x)p_g^\sigma}{\mu_Y(x)p_g^\sigma + (1-\mu_Y(x))p_b^\sigma}.$$

Our updating procedure is given by $\mu(x,T) := (\mu_N(x), \mu_1(\mu_Y(x), p_b^1, p_g^1), \ldots, \mu_Z(\mu_Y(x), p_b^Z, p_g^Z))$.

2.2. Equilibrium concept and design problem

We are interested in behavior which is rational given quality perceptions, in quality perceptions which are consistent with behavior, and in equilibrium selection by the principal. For expositional reasons, we define an equilibrium concept for the subgame that starts after the principal has designed a test and let the principal then pick a test and an equilibrium. For a given test $T \in \mathcal{T}$, $(x,\mu) \in \mathcal{X} \times [0,1]^{Z+1}$ specifies an equilibrium if

$$x(\theta) \in \arg\max_{y\in[0,1]} yU_\theta(T,\mu) + (1-y)u(\mu_N) \text{ for any } \theta \in [\underline{\theta}, \overline{\theta}] \tag{1}$$

and if $\mu = \mu(x,T)$. Picking an equilibrium for test T corresponds to selecting a strategy x from $\mathcal{X}^\star(T) := \{x \in \mathcal{X} | (1) \text{ holds for } \mu = \mu(x,T)\}$. The principal's design problem corresponds thus to choosing $T \in \mathcal{T}$ and $x \in \mathcal{X}^\star(T)$ to maximize $V(T, x, \mu(x,T))$.

The ex ante expected quality perception is by Bayesian plausibility always θ_0 such that $V(T,x,\mu(x,T))$ describes the variance of the posterior quality perception distribution. The principal maximizes thus information generation/learning as measured by the variance of the posterior quality perception distribution. Such an objective is for example implied by a non–reduced version of our model in which the principal estimates the probability that the agent is good and in which she minimizes the expected quadratic error of this estimate.[9]

2.3. The role of our notion of imperfection of private information

We explain in this subsection why our notion of imperfection of private information implies that some learning is possible but that perfect learning is not. As the agent's private information allows only for imperfect inferences about his quality, perfect learning requires full participation in an accurate test. When an agent with private signal θ participates in such a test, he incurs a fair quality perception lottery. The lottery is non–degenerate as the agent is uncertain about

[9] Harbaugh and Rasmusen (2013) measure informativeness by applying a loss function to the error of the public's quality estimate. Our reduced form utility specification can be interpreted as an adaptation of their utility specification with a quadratic loss function to our setting with a binary quality. Quadratic loss functions are a standard assumption in sender–receiver games.

his quality. As he is averse to perception risk, participation leads to an expected utility strictly smaller than $u(\theta)$. When he decides not to participate, the principal cannot infer something from his behavior which the agent does not know himself. This limits how bad the principal's perception of his quality can be. The most adverse perception is that he has the worst possible private signal $\underline{\theta}$. As such a perception implies a certain utility of $u(\underline{\theta})$, the agent has for private signals close to $\underline{\theta}$ a strict incentive not to participate. There can thus not be full unraveling for an accurate test rendering perfect learning impossible.[10,11] On the other hand, learning is generally possible as there is full unraveling for an informative but sufficiently inaccurate test. Intuitively, when the test is very inaccurate, participation allows for the pooling with higher private signals at almost no "cost".

Note that it is indeed our notion of imperfection which makes the test design problem interesting. If the agent's private information is not imperfect in the sense that he is never certain to be bad (that is, if $\underline{\theta} = 0$), an unraveling equilibrium exists for an accurate test. Perfect learning is possible. If there is no meaningful private information (that is, if $\underline{\theta} = \overline{\theta}$), there is no signaling motive on the agent's side. The sole effect of participation is that additional, unbiased information is generated. The agent is never willing to participate in an informative test rendering learning about his quality impossible.

3. An example to motivate effects and research questions

To get an impression of the relevant effects, we discuss in this section a specific class of test for which effects are particularly transparent. Consider tests with the "structure" $T^{gb}(\rho) := ((\rho, 1-\rho, 0), (0, 1-\rho, \rho))$ and the "accuracy" $\rho \in (0, 1]$ (see Table 1). When the agent participates, two things can happen. With probability ρ his quality is perfectly revealed. The test result is either $\sigma = 1$ or $\sigma = 3$ and allows for the inference that the agent is bad ($\mu_1 = 0$) or good ($\mu_3 = 1$), respectively. With probability $1 - \rho$ no new information about the agent's quality is generated. Nevertheless, the test result $\sigma = 2$ reveals information about the agent's private information through his participation decision ($\mu_2 = \mathbf{E}_\theta[\theta|\text{"participation"}]$). Likewise, the observation of non–participation reveals information about his private information ($\mu_N = \mathbf{E}_\theta[\theta|\text{"non–participation"}]$).

[10]The literature on information transmission with hard information (Grossman and Hart (1980), Grossman (1981), Milgrom (1981) and Okuno-Fujiwara et al. (1990)) finds that an agent who is in possession of verifiable private information discloses any information voluntarily. Adverse beliefs and unraveling render full disclosure, that is "full participation in the verification technology", optimal. Although participation in an accurate test discloses verifiable information about the agent's quality, an analogous reasoning does not apply to our setting. The combination of three assumptions is responsible for this: First, $\underline{\theta} > 0$ limits how adverse the quality perception associated to non–participation can be. Second, because the test can generate information which goes beyond what the agent knows, quality perceptions associated to test results can be worse than this. Third, due to the agent's perception risk–aversion, even the agent with the worst possible private signal who cannot lose in terms of expected perception when he gets tested has a strict incentive not to participate.

[11]The agent's aversion to perception risk is crucial for making our design problem interesting. If the agent is neutral to perception risk or if he loves perception risk, there exists an equilibrium with full unraveling for an accurate test rendering perfect learning possible.

Table 1: Tabular description of tests $T^{gb}(\rho)$

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
$\omega = g$	0	$1-\rho$	ρ
$\omega = b$	ρ	$1-\rho$	0
$\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma)$	$= 0$	$= \mu_Y(x)$	$= 1$

How does the agent decide on participation? If he participates, a test result is generated according to the probability vector $p_\theta = (\rho(1-\theta), 1-\rho, \rho\theta)$. His private signal affects the probabilities with which the perfectly informative signals $\sigma = 1$ and $\sigma = 3$ are generated. The higher his private signal, the more likely he is perceived as good ($\sigma = 3$) and the less likely he is perceived as bad ($\sigma = 1$). If he does not participate, it does not depend on his private signal how he is perceived. The agent has thus a stronger incentive to participate if his private signal is higher. Only threshold equilibria can exist. When we denote the participation threshold by s, the threshold agent obtains an expected perception of $\rho s + (1-\rho)\mathbf{E}_\theta[\theta|\theta \geq s]$ from participation and a certain perception of $\mathbf{E}_\theta[\theta|\theta \leq s]$ from non–participation. He faces a trade–off between being in average perceived as worse (non–participation) and bearing a perception risk (participation). How this trade–off is resolved depends on the test accuracy ρ.

Consider now the principal's choice of the test accuracy ρ. She would clearly benefit from an increase in ρ if this did not affect the participation threshold s. This is because for a given s a higher ρ transforms the induced posterior quality perception distribution by a mean–preserving spread and increases thus its variance. However, the threshold agent's participation incentive clearly deteriorates when ρ increases as participation implies then a higher perception risk and a lower expected perception.[12] By the reasoning in Subsection 2.3, an accurate test (that is, $\rho = 1$) induces partial participation, whereas a sufficiently inaccurate test (that is, ρ sufficiently close to zero) induces full participation. The principal faces thus a trade–off between accuracy and participation.[13] Inducing less than full participation might be optimal as it allows for a higher test accuracy and enables indirect learning through the agent's participation decision.

We discussed in this section tests which generate information with a specific structure. The principal can however often affect which kind of information is generated. The main goal of this article is to derive which kind of information is generated by the optimal testing procedure. The knowledge of the optimal test structure leaves the designer then with a much simpler problem which is basically as the problem which we discussed in this section.

[12]Unless the participation constraint is slack, the participation threshold must adjust in response to an increase in the test accuracy. Interestingly, it is a priori not clear in which direction it adjusts. Responsible for this is that an increase in the participation threshold improves the pool of participants as well as the pool of non–participants. Depending on the relative strength of these improvements, the threshold signal may increase or decrease.

[13]To be precise, participation is at least eventually decreasing in the test accuracy. Locally, participation might also be increasing as argued in Footnote 12.

Table 2: Test structures and implied quality perceptions

(a) $T^g(\rho)$

	$\sigma = 1$	$\sigma = 2$
$\omega = g$	$1-\rho$	ρ
$\omega = b$	1	0
$\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma)$	$\in [0, \mu_Y(x))$	$= 1$

(b) $T^b(\rho)$

	$\sigma = 1$	$\sigma = 2$
$\omega = g$	0	1
$\omega = b$	ρ	$1-\rho$
$\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma)$	$= 0$	$\in (\mu_Y(x), 1]$

4. Optimal test design

Two specific test structures are particularly important for the arguments in this section, $T^g(\rho) := ((1,0),(1-\rho,\rho))$ and $T^b(\rho) := ((\rho, 1-\rho),(0,1))$ with $\rho \in (0,1]$ (see Table 2). $T^g(\rho)$ describes a binary test which is not subject to false positives. Test result $\sigma = 2$ perfectly reveals that the agent is good, whereas there is pooling on test result $\sigma = 1$. $T^b(\rho)$ describes a binary tests which is not subject to false negatives. Test result $\sigma = 1$ perfectly reveals that the agent is bad, whereas there is pooling on test result $\sigma = 2$. ρ describes for both test structures the test accuracy.

The analysis in this section proceeds in four steps: First, we characterize which participation behavior can be induced by some test (Subsection 4.1). Then, we fix any inducible participation behavior and analyze through which test it is optimally induced (Subsection 4.2). We find that the questions about the "structure" of the optimal testing procedure and its "accuracy" can indeed be separated. Any inducible participation behavior is optimally induced by a test $T^g(\rho)$. Third, we discuss the principal's participation–accuracy trade–off under the optimal test structure for the uniform–quadratic case (Subsection 4.3). Finally, we discuss generalizations of our modeling assumptions (Subsection 4.4).

4.1. Inducible participation behavior

Important for the characterization of equilibrium behavior are threshold participation strategies. $x \in \mathcal{X}$ is a threshold strategy if there exists $s \in [\underline{\theta}, \overline{\theta}]$ such that $x(\theta) = 0$ for $\theta \in [\underline{\theta}, s)$ and $x(\theta) = 1$ for $\theta \in (s, \overline{\theta}]$. We denote the agent when he has the private signal $\theta = s$ as threshold agent. Let x^s be the specific threshold strategy with $x^s(s) = 1$ if $s \in [\underline{\theta}, \overline{\theta})$ and $x^s(s) = 0$ if $s = \overline{\theta}$.

Take any test $T \in \mathcal{T}$ as given. If the agent participates, a test result is generated which is informative about his quality. Test results associated with higher quality perceptions are good news in the sense of Milgrom (1981). As a higher private signal makes better news more likely in the sense of first–order stochastic dominance, participation is more attractive for higher private signals. Only threshold participation strategies can thus be part of an equilibrium.

Lemma 1 *Fix any $T \in \mathcal{T}$. (a) For any $x \in \mathcal{X}$, $U_\theta(T, \mu(x,T))$ is continuous and strictly increasing in θ. (b) If $x \in \mathcal{X}^\star(T)$, x is a threshold strategy. (c) $\mathcal{X}^\star(T)$ is non–empty.*[14]

[14] While $\mathcal{X}^\star(T)$ is generally non–empty, it might not be a singleton. The reason for this is that a higher

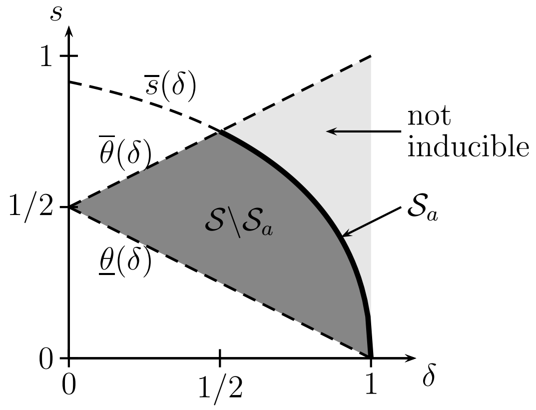


Figure 2: Inducible participation thresholds [$F = F_\delta$, $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$]

The specific threshold strategy x^s is constructed such that s is inducible by some threshold strategy if it is inducible by x^s. As the principal's expected utility is not affected by the threshold agent's behavior, it is without loss of generality to restrict attention to threshold strategies of the form x^s.

From a design perspective, only thresholds $s \in [\underline{\theta}, \overline{\theta})$ which induce participation with positive probability are interesting. We characterize now which of these thresholds can be induced by some test, $\mathcal{S} := \{s \in [\underline{\theta}, \overline{\theta}) | \exists T \in \mathcal{T} : x^s \in \mathcal{X}^\star(T)\}$, and which ones can be induced by an accurate test, $\mathcal{S}_a := \{s \in [\underline{\theta}, \overline{\theta}) | \exists T \in \mathcal{T}_a : x^s \in \mathcal{X}^\star(T)\}$. Inducing a participation threshold s consists of two parts: Motivating participation for private signals $\theta > s$ and deterring participation for private signals $\theta < s$. As the motivation part can always be satisfied by making the test sufficiently inaccurate, the deterrence part is crucial for inducibility. This part hinges on the threshold agent's incentive to participate in an accurate test.

Lemma 2 *(a)* $\mathcal{S}_a = \{s \in [\underline{\theta}, \overline{\theta}) | (1-s)u(0) + su(1) = u(\mathbf{E}_\theta[\theta | \theta \leq s])\}$. *(b)* $\mathcal{S}\backslash\mathcal{S}_a = \{s \in [\underline{\theta}, \overline{\theta}) | (1-s)u(0) + su(1) < u(\mathbf{E}_\theta[\theta | \theta \leq s])\}$. *(c) For any* $s \in \mathcal{S}\backslash\mathcal{S}_a$*, there exist* $\rho_g, \rho_b, \rho_{bg} \in (0,1)$ *such that* s *is induced by the tests* $T^g(\rho_g)$*,* $T^b(\rho_b)$ *and* $T^{gb}(\rho_{gb})$*.*

The design of the optimal test has to deal with two problems: First, not any threshold $s \in [\underline{\theta}, \overline{\theta})$ needs to be inducible. As $(1-s)u(0) + su(1)$ and $u(\mathbf{E}_\theta[\theta | \theta \leq s])$ both increase in s, the set $\mathcal{S}\backslash\mathcal{S}^a$ needs even not to be connected. Second, any threshold $s \in \mathcal{S}\backslash\mathcal{S}^a$ can be induced by multiple test, for example by tests with the structure $T^g(\rho)$, $T^b(\rho)$ and $T^{gb}(\rho)$.

Example: The uniform–quadratic case. To illustrate properties of $\mathcal{S}$ and $\mathcal{S}_a$, consider $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$ and private information which is uniformly distributed around 1/2. That is, $F = F_\delta$ with $F_\delta(\theta) := (\theta - \underline{\theta}(\delta))/(\overline{\theta}(\delta) - \underline{\theta}(\delta))$, $\underline{\theta}(\delta) := (1-\delta)/2$, $\overline{\theta}(\delta) := (1+\delta)/2$ and

supposed participation threshold makes participation as well as non–participation both more attractive. Depending on the relative speed with which participation and non–participation become more attractive when the supposed participation threshold increases, there might for a given test T coexist an equilibrium with a low stigma of failure (that is, with a low level of participation) and with a high stigma of failure (that is, with a high level of participation).

$\delta \in (0,1)$. We obtain $\mathcal{S}_a = \{s \in [\underline{\theta}(\delta), \overline{\theta}(\delta)) | s = \overline{s}(\delta)\}$ and $\mathcal{S} \backslash \mathcal{S}_a = \{s \in [\underline{\theta}(\delta), \overline{\theta}(\delta)) | s < \overline{s}(\delta)\}$ with $\overline{s}(\delta) := 2\sqrt{\underline{\theta}(\delta)} - \underline{\theta}(\delta)$.

Figure 2 illustrates how the two sets depend on δ. A higher δ transforms the distribution of the agent's private information by a mean–preserving spread. δ can thus be interpreted as a measure for the accuracy of the agent's private information. If $\delta < 1/2$, the agent's signaling motive is limited for any possible private signal. By making the test sufficiently accurate, he can for any possible private signal be deterred from participating. As a consequence, any threshold $s \in [\underline{\theta}(\delta), \overline{\theta}(\delta))$ is inducible. If $\delta > 1/2$, there exist signals for which the agent is sufficiently certain to be good such that he cannot be deterred from participating by increasing the test accuracy. The thresholds $s \in (\overline{s}(\delta), \overline{\theta}(\delta)]$ are not inducible. Due to a partial unraveling effect which becomes stronger as δ increases, the set of private signals for which the agent cannot be deterred from participating becomes larger as δ increases. If δ is close to one, there is almost full unraveling for any given test.

4.2. The optimal inducement of a given participation behavior

We investigate now how a given participation threshold $s \in \mathcal{S}$ is optimally induced. As long as we hold participation fix, we simplify notation by writing μ_Y instead of $\mu_Y(x^s)$ and μ_N instead of $\mu_N(x^s)$. The principal's problem is then

$$\max_{T \in \mathcal{T}} \; F(s)v(\mu_N) + (1 - F(s)) \left[\sum_\sigma p^\sigma_{\mu_Y} v(\mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g)) \right] \tag{2}$$

$$\text{s.t.} \quad \begin{cases} U_s(T, \mu(x^s, T)) = u(\mu_N) & \text{if } s \in (\underline{\theta}, \overline{\theta}) \\ U_s(T, \mu(x^s, T)) \geq u(\mu_N) & \text{if } s = \underline{\theta} \end{cases} . \tag{3}$$

The objective function in (2) follows from using the structure of x^s to simplify $V(T, x^s, \mu(x^s, T))$. The participation constraint in (3) follows from using Lemma 1 (a) to rephrase $x^s \in \mathcal{X}^\star(T)$. We first derive two auxiliary results (Lemma 3 and 4) which we use then to show that the optimal binary test is not subject to false positives (Proposition 1) and that the optimal test is binary (Proposition 2).

Our first auxiliary result concerns how the agent is perceived in expected terms when he has private signal θ and participates in the test T, $E_\theta(T, \mu) := \sum_\sigma p^\sigma_\theta \mu_\sigma$. By using that $p^\sigma_\theta = (p^\sigma_{\mu_Y} - (\mu_Y - \theta)(p^\sigma_g - p^\sigma_b))$ and that by Bayes' Law $\sum_\sigma p^\sigma_{\mu_Y} \mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g) = \mu_Y$, we obtain

$$E_\theta(T, \mu(x^s, T)) = \mu_Y - (\mu_Y - \theta)(E_1(T, \mu(x^s, T)) - E_0(T, \mu(x^s, T))). \tag{4}$$

$E_1(T, \mu)$ (resp. $E_0(T, \mu)$) can be interpreted as the interim expected perception from the viewpoint of a hypothetical agent who is certain to be good (resp. bad). It follows directly from the proof to Lemma 1 (a) that $E_1(T, \mu(x^s, T)) - E_0(T, \mu(x^s, T)) > 0$ and that $E_1(T, \mu(x^s, T)) - E_0(T, \mu(x^s, T)) < 1$ for any test $T \in \mathcal{T} \backslash \mathcal{T}_a$. Any non–accurate test induces only for $\theta = \mu_Y$ a fair quality perception lottery. If the agent's private signal is below (resp. above) μ_Y, he gains (resp. loses) in terms of expected perception. Intuitively, noise redistributes expected perception from the case where the agent is good to the case where he is bad and therewith from cases where it is more likely that he is good ($\theta > \mu_Y$) to cases where it is less likely that he is good ($\theta < \mu_Y$). How much he gains (resp. loses) depends on the specifics of the test design. Interestingly, for any given private signal θ, all tests among which the principal is indifferent imply the same interim expected perception of the agent:

Lemma 3 *For any $s \in \mathcal{S}$ and any θ, $V(T', x^s, \mu(x^s, T')) = V(T'', x^s, \mu(x^s, T''))$ implies that $E_\theta(T', \mu(x^s, T')) = E_\theta(T'', \mu(x^s, T''))$.*

The lemma is a consequence of Bayesian updating. Bayes' Law implies that the first moment of the quality perception distribution conditional on participation in a test T, $E_{\mu_Y}(T, \mu(x^s, T))$, is constant and that the second moment of this distribution depends only through $E_1(T, \mu(x^s, T))$ on the test design. Hence, for tests among which the principal is indifferent, $E_\theta(\cdot)$ depends neither for $\theta = \mu_Y$ nor for the (hypothetical) case with $\theta = 1$ on the test design. Linearity of $E_\theta(T, \mu(x^s, T))$ in θ implies the result.

Our second auxiliary result concerns the effects of increasing the test accuracy ρ of a test $T^g(\rho)$. There are two main effects: First, a more accurate test induces less pooling and allows thus for better signaling of favorable private information. In particular, the threshold agent's expected perception from participating clearly decreases in ρ. Second, the test accuracy affects the perception risk faced by the agent in a non–trivial way.[15] Nevertheless, we obtain that the compound effect on the threshold agent's and the principal's expected utility is clear–cut.

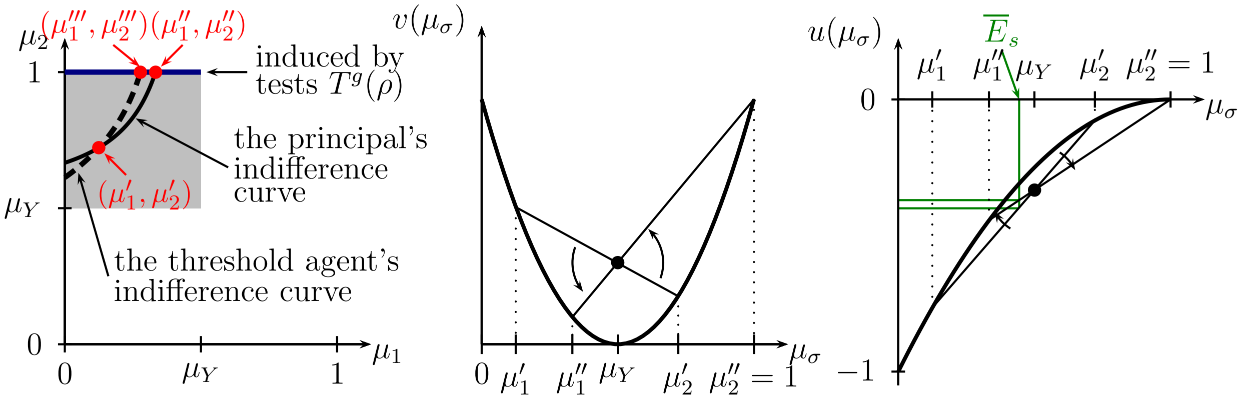
Lemma 4 *(a) For any $s \in \mathcal{S}$, $V(T^g(\rho), x^s, \mu(x^s, T^g(\rho)))$ is strictly increasing in ρ, whereas $U_s(T^g(\rho), \mu(x^s, T^g(\rho)))$ is strictly decreasing in ρ. (b) For any $s \in \mathcal{S}\backslash\mathcal{S}^a$, there exists $\rho^* \in (0, 1)$ such that $U_s(T^g(\rho^*), \mu(x^s, T^g(\rho^*))) = u(\mu_N)$. (c) For any $s \in \mathcal{S}$, $E_s(T^g(\rho), \mu(x^s, T^g(\rho)))$ is strictly decreasing in ρ.*

Crucial for understanding our main result is the following hypothetical question: When $T^g(\rho_g)$ and $T^b(\rho_b)$ specify tests for which the principal's objective function (2) attains the same value, which of these tests makes participation more attractive for the threshold agent? Intuitively, the difference between the tests is the following:

- The test $T^b(\rho_b)$ reveals a bad agent's quality with probability ρ_b and pools him otherwise with *all* good agents.
- The test $T^g(\rho_g)$ reveals a good agent's quality with probability ρ_g and pools him otherwise with all bad agents. A bad agent is thus only pooled with *some* good agents.

As the test $T^b(\rho_b)$ pools a bad agent with *more* good agents, this test seems at first glance better for an agent who is relatively likely to be bad like the threshold agent. This kind of

[15]This can be most easily seen for the case in which signals exist for which the agent is relatively certain about his quality. A hypothetical agent who is certain about his quality faces no perception risk when he participates in an accurate or a completely inaccurate test, whereas he faces a perception risk for intermediate levels of test accuracy. The non–monotonicity carries over to an agent who is only relatively certain about his quality.



(a) Indifference curves (b) The principal's utility (c) The agent's utility

Figure 3: Quality perceptions inducible by and utility from binary tests [$\mu_Y = 1/2$, $s = \underline{\theta} = 1/3$, $u(\mu_\sigma) = -(1-\mu_\sigma)^2$]

reasoning is however misleading. As for any given private signal all tests among which the principal is indifferent induce by Lemma 3 the same interim expected perception, the threshold agent's preferences are not driven by expected perception but by perception risk. As a bad agent—and therewith any agent who is relatively likely to be bad—incurs a lower perception risk when false positives are less likely, the threshold agent prefers the test $T^g(\rho_g)$ which is not subject to false positives. This gives the designer scope for increasing the test accuracy which in turn improves learning about the agent's quality (see Lemma 4 (a)). Hence, given any test $T^b(\rho_b)$ which satisfies (3), we can take the test $T^g(\rho_g)$ as described above and modify it by increasing the test accuracy to obtain a test $T^g(\rho_g')$ which satisfies (3) and which is better for the principal. As a similar reasoning can be applied to improve also upon any other given binary test T', we obtain the following result:

Proposition 1 *For any $s \in \mathcal{S}$, there exists $\rho^* \in (0,1]$ such that $T^g(\rho^*)$ is the optimal binary test inducing s. If $s \in \mathcal{S}_a$, $\rho^* = 1$. If $s \in \mathcal{S}\backslash\mathcal{S}_a$, ρ^* is implicitly defined by $U_s(T^g(\rho^*), \mu(x^s, T^g(\rho^*))) = u(\mu_N)$.*

We construct in the proof for any given binary test T' first a test $T^g(\rho_g)$ which makes the principal indifferent to the initial test. By Lemma 3, the threshold agent's expected perception is not affected. We then increase the accuracy of the test $T^g(\rho_g)$ implying that the threshold agent's expected perception decreases by Lemma 4 (c). This directly implies the following corollary:

Corollary 1 *Among all binary tests inducing s, the interim expected perception of the threshold agent is lowest for the optimal test.*

A graphical intuition. For explaining our subsequent results, it will be useful to give also a graphical intuition for Proposition 1. The first step of the proof establishes that the updating rules $\mu_1 = \mu_1(\mu_Y, p_b^1, p_g^1)$ and $\mu_2 = \mu_2(\mu_Y, p_b^2, p_g^2)$ imply a one–to–one correspondence between binary tests with $p_b^2 < p_g^2$ and quality perceptions with $0 \le \mu_1 < \mu_Y < \mu_2 \le 1$. It follows that the test design problem can be handled as a quality perception design problem. As the principal cannot affect μ_N under our supposition that s is fix, she chooses basically a quality perception pair $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$. The set of feasible quality perception pairs is illustrated by the grey area in Figure 3(a). The blue line segment on the top describes the quality perception pairs associated with tests which are not subject to false positives. Points which lie further to the left on this line segment are associated with tests with a higher accuracy.

Take any quality perception pair $(\mu_1', \mu_2') \in [0, \mu_Y) \times (\mu_Y, 1)$ which induces the desired participation behavior s (that is, any pair for which the associated test satisfies (3)) as given. The dashed curve in Figure 3(a) displays the indifference curve of the threshold agent through (μ_1', μ_2'). All points on this curve are associated to tests which induce s.[16] We show that a test which is not subject to false positives must be optimal by showing that it is possible to move from (μ_1', μ_2') to (μ_1''', μ_2''') in a way such that the principal's expected utility is non–decreasing. However, instead of moving directly on the threshold agent's indifference curve, we leave this curve at first and move on the principal's indifference curve (= the solid curve in Figure 3(a)). Figure 3(b) demonstrates how quality perception pairs on the principal's indifference curve relate to each other. The principal's expected utility from (μ_1, μ_2) (conditional on participation by the agent) is obtained by evaluating the secant line through $(\mu_1, v(\mu_1))$ and $(\mu_2, v(\mu_2))$ at μ_Y. The secant lines of quality perception pairs among which the principal is indifferent rotate at μ_Y. For example, the principal is indifferent between (μ_1', μ_2') and (μ_1'', μ_2'') as displayed in Figure 3(b).

Consider next how quality perception pairs which make the principal indifferent affect the threshold agent. When $u(\cdot)$ is quadratic, the principal's expected utility is a negative linear transformation of the expected utility of an agent with private signal $\theta = \mu_Y$. This implies that also the secant lines of $u(\cdot)$ rotate at μ_Y (see Figure 3(c)). By (4) and Lemma 3, the threshold agent's expected utility is obtained by evaluating the secant lines at a constant value on the left of μ_Y, say $\overline{E}_s$. As a consequence, his expected utility increases as the secant line rotates clockwise. That is, his expected utility increases as we move to the northeast on the principal's indifference curve in the (μ_1, μ_2)–space. As the threshold agent's expected utility increases by Lemma 4 from the left to the right on the blue line segment, (μ_1'', μ_1'') must indeed lie as depicted to the right of (μ_1''', μ_2'''). As the principal's expected utility increases by the same lemma from the right to the left on the blue line segment, we obtain that her expected utility increases as we move from (μ_1'', μ_1'') to (μ_1''', μ_2'''). This establishes that the principal prefers (μ_1''', μ_2''') over (μ_1', μ_2') completing the proof for the quadratic case.

Two remarks are in order: First, for non–quadratic $u(\cdot)$, the secant lines of $u(\cdot)$ for quality perception pairs which make the principal indifferent do not intersect at the same point. This

[16] Although we can compare the principal's and the threshold agent's expected utility for points (μ_1, μ_2) which do not induce s, the principal can ultimately choose only among points which induce s. If $s > \underline{\theta}$, only the points on the agent's indifference curve induce s. If $s = \underline{\theta}$, there might also exist other points which induce s as the threshold agent's participation constraint needs not to be binding.

adds an additional effect. However, prudence of $u(\cdot)$ implies that the additional effect reinforces the effect in the quadratic case. Intuitively, as concavity is decreasing under prudence, the threshold agent suffers relatively less from the perception risk imposed by quality perception pairs which lie farther to the northeast in the (μ_1, μ_2)–space. His expected utility increases thus even stronger as we move to the northeast on the principal's indifference curve. Second, it does not matter for our reasoning whether the threshold agent's participation constraint is initially binding or slack. The only difference is the following: The optimal quality perception pair is (μ_1''', μ_2''') (resp. lies on the left of (μ_1''', μ_2''')) when the agent's participation constraint is initially binding (resp. slack).

It remains to argue why a non–binary test $T' \in \mathcal{T}^Z$ which induces s cannot be optimal. Suppose without loss of generality that $\mu_1' < \mu_2' < \ldots < \mu_Z'$ for $\mu' = \mu(x^s, T')$. The result is driven by two properties:

Property 1 By considering modifications of a non–binary test which affect only two test results, we obtain basically a binary problem to which our graphical intuition applies.

Property 2 When two test results induce the same quality perception, we can merge these test results to obtain an equivalent test with a smaller number of test results.

These properties allow us to apply the following iteration procedure:

Step 1 Consider modifications of the test T' which affect only the test results $\sigma = 1$ and $\sigma = 2$. Modify the test such that the principal is indifferent and the quality perceptions associated to $\sigma = 1$ and $\sigma = 2$ increase until the quality perception associated with $\sigma = 2$ equals μ_3'.

Step 2 Construct a test $T'' \in \mathcal{T}^{Z-1}$ by merging the test results $\sigma = 2$ and $\sigma = 3$.

Step 3 Iterate Steps 1 and 2 with $T' = T''$ until the resulting test is binary.

By construction, the principal is indifferent between the constructed binary test and the initial non–binary test T', whereas the threshold agent is better off. This allows us to apply Proposition 1 to obtain a binary test which induces s and which is better for the principal.[17]

Proposition 2 *Any $s \in \mathcal{S}$ is optimally induced by a binary test.*

By Lemma 3, the transformations described in Steps 1 to 3 above do not affect the threshold agent's expected perception. It follows thus directly that Corollary 1 extends to the case in which we allow for non–binary tests.

[17] Proposition 4 in the Web Appendix of Kamenica and Gentzkow (2011) shows for a persuasion problem in which the designer is only restricted by Bayesian plausibility that the optimal information structure requires the generation of at most as many signals as there exist quality realizations. Proposition 2 here shows that an analogous result is obtained for our testing problem in which the designer is also subject to an interim participation constraint.

Corollary 2 *Among all tests inducing s, the interim expected perception of the threshold agent is lowest for the optimal test.*

Harbaugh and Rasmusen (2013) consider a related problem in which the agent is perception risk–neutral. The designer's only instrument to foster participation is to increase the expected perception of the threshold agent. Corollary 2 argues that when the agent is imperfectly informed and perception risk–averse, non–trivial effects are added. The designer can then also foster participation by reducing the threshold agent's perception risk and doing so is better for learning than increasing the threshold agent's expected perception.

To conclude our discussion of the optimal test structure, let us stress that the structure of the optimal test depends neither on the primitives of our model nor on the participation threshold that shall be induced. This means in particular that the existence of an additional constraint which requires the principal to induce full participation does not affect the structure of the optimal test. A binary test which is not subject to false positives is always optimal.

4.3. The optimal participation behavior in the uniform–quadratic case

We know from the analysis so far that only threshold participation behavior can be induced (Lemma 1), which participation thresholds can be induced (Lemma 2) and how any inducible participation threshold is optimally induced (Propositions 1 and 2). The principal's test design problem reduces thus to the choice of a participation threshold:

$$\max_{s\in\mathcal{S}} \quad F(s)v(\mu_N(x^s)) + (1-F(s))\left[(1-\mu_Y(x^s)\rho)v(\mu_1(\mu_Y(x^s),1,1-\rho)) + \mu_Y(x^s)\rho v(1)\right]$$

$$\text{s.t.} \quad \begin{cases} \rho = 1 & \text{if } s\in\mathcal{S}_a \\ (1-\rho s)u(\mu_1(\mu_Y(x^s),1,1-\rho)) + \rho s u(1) = u(\mu_N(x^s)) & \text{if } s\in\mathcal{S}\backslash\mathcal{S}_a \end{cases}$$

Although the structure of the optimal test is quite simple, the problem which participation threshold to induce is complex because s simultaneously determines who participates, the probability with which the different test results are generated conditional on participation, and what can be inferred from the different possible observations. We discuss the choice of s therefore only for the uniform–quadratic case introduced in Subsection 4.1 (that is, for $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$ and $F = F_\delta$).

First, recall that it is not clear whether a general trade–off between accuracy and participation exists (see Footnote 12). The following result establishes that this is indeed the case.[18]

Proposition 3 *Consider $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$, $F = F_\delta$ and for any inducible participation threshold s the test $T^g(\rho)$ which optimally induces it. More participation is then induced by a less accurate test. That is, a lower s is induced by a lower ρ.*

[18]De and Nabar (1991) show in a certification context that a similar trade–off can also arise with a perfectly informed, risk–neutral seller who can voluntarily decide to get tested at an exogenously given fee. An inaccurate testing technology may foster participation relative to an accurate one.

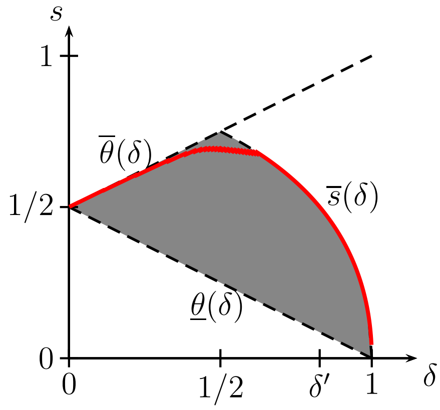


Figure 4: The optimal participation threshold $[F = F_\delta,\ u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2]$

Decreasing the accuracy of a test $T^g(\rho)$ makes the high test result $\sigma = 2$ less likely and reduces the stigma associated to the low test result $\sigma = 1$. Hence, if the optimal test structure is used, more participation is induced by "reducing the stigma of failure" and not by "inflating grades".

How does the accuracy of the agent's private information affect his participation decision? The grey area in Figure 4 indicates the inducible participation thresholds, the red solid line illustrates the optimal participation threshold.[19] If the accuracy of the agent's private signal is low, any participation threshold $s \in [\underline{\theta}(\delta), \overline{\theta}(\delta)]$ is inducible. As signals are spread out in a small interval around 1/2, the worst possible quality perception associated to non–participation is quite high and the agent's signaling motive is weak. Fostering participation is costly for the principal in terms of test accuracy. The principal prefers little participation in a relatively accurate test to higher participation in a much less accurate test. The optimal participation threshold lies thus close to $\overline{\theta}(\delta)$. If the accuracy of the agent's private signal is high, the agent's signaling motive is strong as the quality perception associated to non–participation is low. An unravelling effect kicks in causing high participation for an accurate test. As inducing participation for private signals close to $\underline{\theta}(\delta)$ requires a very inaccurate test, fostering even more participation is, again, very costly in terms of test accuracy. An accurate test and therewith the highest inducible participation threshold is optimal.

How does the accuracy of the agent's private information affect indirect learning through his participation decision? As inducing full participation is never optimal, there is always a role for indirect learning. The example is constructed such that indirect learning is for any δ the better the closer s lies to 1/2. However, if δ is low (resp. high), there is only little indirect learning as the optimal threshold lies close to $\overline{\theta}(\delta)$ (resp. $\underline{\theta}(\delta)$). Indirect learning is best for intermediate δ. By a continuity property, there exists an intermediate value of δ, say δ', for which learning through private information is optimal. If δ increases beyond δ', a higher accuracy of private information implies less learning through private information (see Figure 4).

[19]The optimal participation threshold is computed numerically using Maple.

4.4. Discussion of generalizations of our modeling assumptions

More general distributions of private information: atoms and holes. Our results concerning the optimal inducement of a given participation behavior extend to distributions of private information which include atoms and holes. A threshold strategy x is then specified by a participation threshold s and a probability with which the threshold agent participates. As it matters for our analysis only that $\mu_Y(x) > s$, but not how $\mu_Y(x)$ arises, the extension is straightforward.

More general reduced form utility functions. Extensions into two directions are possible: First, note that the analyzed case where the principal maximizes the variance of the posterior quality perception distribution corresponds basically to the general case where $v(\cdot)$ is quadratic and convex.[20] An extension to the case where $v(\cdot)$ is cubic, convex and has a positive third derivative follows as a corollary from Proposition 1 (see Corollary B 1 in the appendix). Intuitively, if the third derivative is positive, the principal has a preference for inducing quality perception lotteries with high quality perceptions. As the test which optimally induces a given participation threshold s in the quadratic case happens to be the test which induces s with the highest possible quality perceptions, this test remains optimal in the cubic case.

Second, what makes the case in which the principal's and the agent's utility functions are both quadratic which we use to provide our graphical intuition for Proposition 1 particularly tractable is that the principal's expected utility is a negative linear transformation of the agent's ex ante expected utility. The game has basically a constant–sum structure. A generalization to other utility specifications which exhibit such a structure is possible but requires a different strategy of proof.[21]

5. Non–binary quality types and limits to information generation

We extend our model in this section by introducing additional constraints. Quality is ultimately still binary, but tests can now condition only on non–binary quality types which are only imperfectly informative about the quality. This generates limits to information generation.

The aim of this section is to demonstrate how our techniques from the case without limits to information generation can be used to study the problem in which such limits exist. We demonstrate this for a special case of the general problem with limits to information generation which has an important interpretation: With a certain probability the test is not capable of generating new information, otherwise information generation is possible as before. This version of the problem is specific as the "intermediate" quality type allows for a particular inference

[20] This is a consequence of the fact that the first moment of any inducible quality perception distribution is fix by Bayes' Law.

[21] In an earlier version of this article, we derived such results for the case in which the agent has a HARA utility function and the principal's utility is a negative linear transformation of the agent's utility (see Proposition 3 (a) in Rosar and Schulte (2012)).

Table 3: Test structures and implied quality perceptions [$\gamma > 0$]

(a) $T^{(g,y)}(\rho_g)$

	$\sigma = 1$	$\sigma = 2$
$\widehat{\omega} = g$	$1 - \rho_g$	ρ_g
$\widehat{\omega} = n$	1	0
$\widehat{\omega} = b$	1	0
$\mu_\sigma(\cdot)$	$\in [0, \mu_Y(x))$	$= 1$

(b) $T^{(g,y)}(\rho_n, \rho_g)$

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
$\widehat{\omega} = g$	0	$1 - \rho_g$	ρ_g
$\widehat{\omega} = n$	$1 - \rho_n$	ρ_n	0
$\widehat{\omega} = b$	1	0	0
$\mu_\sigma(\cdot)$	$\in [0, \mu_Y(x))$	$\in [\mu_Y(x), 1)$	$= 1$

which simplifies the formulas and makes the effects most transparent. The employed ideas are non–specific and apply to the general problem.

The modified model. Suppose that the agent's quality ω is ultimately still either good or bad and that the agent is still privately informed about his probability of being good θ, but that the test might now be incapable of generating new information. Information generation is possible when $\tau = y$, but it is not possible when $\tau = n$. τ is distributed independently from ω and θ with $\text{Prob}(\tau = n) = \gamma \in [0, 1)$. Test results can then depend on the three events $(\omega, \tau) = (b, y)$, $(\omega, \tau) \in \{(g, n), (b, n)\}$ and $(\omega, \tau) = (g, y)$. There exist thus three quality types, say $\widehat{\omega} = b$, $\widehat{\omega} = n$ and $\widehat{\omega} = g$, and a test is described by a vector comprised of three conditional probability vectors, $T = (p_b, p_n, p_g)$. Using that $\text{Prob}(\omega = g|\widehat{\omega} = n) = \mu_Y(x)$, we obtain that the quality perception associated to the test result σ is

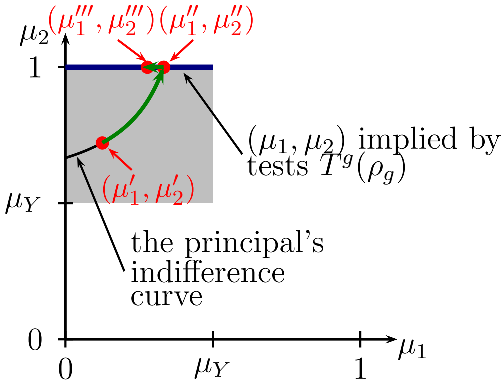
$$\mu_\sigma(\mu_Y(x), p_b^\sigma, p_n^\sigma, p_g^\sigma) = \frac{(1-\gamma)\mu_Y(x)p_g^\sigma + \gamma\mu_Y(x)p_n^\sigma}{(1-\gamma)\mu_Y(x)p_g^\sigma + \gamma p_n^\sigma + (1-\gamma)(1-\mu_Y(x))p_b^\sigma}.$$

The probability with which the agent obtains the test result σ when his private signal is θ is given by $p_\theta^\sigma = (1-\gamma)\theta p_g^\sigma + \gamma p_n^\sigma + (1-\gamma)(1-\theta)p_b^\sigma$. Everything else is as in the original model or extends straightforwardly. The original model is obtained as a special case when $\gamma = 0$.

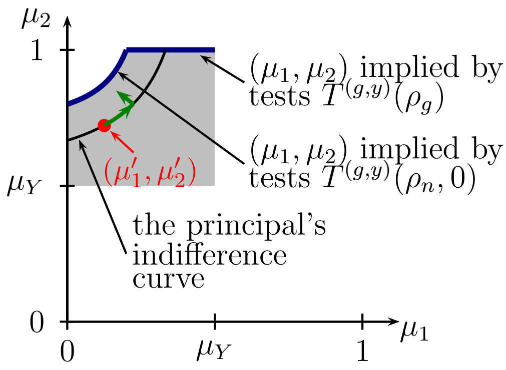
Two classes of tests are particularly important for the analysis in our modified setting, tests $T^{(g,y)}(\rho_g) := ((1,0),(1,0),(1-\rho_g,\rho_g))$ with $\rho_g \in (0,1]$ and tests $T^{(g,y)}(\rho_n,\rho_g) := ((1,0,0),(1-\rho_n,\rho_n,0),(0,1-\rho_g,\rho_g))$ with $\rho_g \in [0,1]$ and $\rho_n \in (0,1]$ (see Table 3).[22] For $\gamma = 0$, $T^{(g,y)}(\rho_g)$ corresponds to $T^g(\rho_g)$ with $\rho_g \in (0,1]$ and $T^{(g,y)}(\rho_n,\rho_g)$ corresponds to $T^g(0)$. The two classes of tests can thus be interpreted as adaptations of binary tests which are not subject to false positives to our modified setting. Tests $T^{(g,y)}(\rho_g)$ lump quality types $\widehat{\sigma} = n$ and $\widehat{\sigma} = b$ and work for the redefined types like a standard binary test which is not subject to false positives. Tests $T^{(g,y)}(\rho_n,\rho_g)$ differentiate between all three types. The highest test result identifies $\widehat{\omega} = g$ correctly, the other two test results pool adjacent types. The "accurate test" in the modified setting is the test which reveals $\widehat{\omega}$ perfectly. It is described by $T^{(g,y)}(1,1) = ((1,0,0),(0,1,0),(0,0,1))$.

Discussion of the modified design problem. It is straightforward to show that also in

[22]By allowing for tests $T^{(g,y)}(\rho_n, 0)$, we slightly abuse notation as the test result $\sigma = 3$ never occurs for such tests.



(a) original model with binary types [$\gamma = 0$]



(b) modified model with non–binary types [$\gamma = 1/4$]

Figure 5: Supportable quality perception pairs and strategy of proof [$\mu_Y = 1/2$]

the modified setting only threshold strategies can be part of an equilibrium. The problem to design a test which optimally induces a given participation threshold s is still given by (2) and (3), but the definitions of T, p_θ^σ and $\mu_\sigma(\cdot)$ have changed. To simplify notation, we write μ_Y and μ_N again without argument.

There are two major differences between the original and the modified design problem. First, any quality perception pair $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$ is in the original model supportable by a binary test, whereas this is not the case in the modified model. As the inherently noisy case where $\widehat{\omega} = n$ has to be associated with at least one of the test results, it is limited how small μ_1 and how large μ_2 can be. In particular, the quality perception pair $(0, 1)$ which corresponds to the perfect revelation of the agent's quality is not supportable. The grey areas in Figures 5(a) and 5(b) illustrate the set of supportable quality perception pairs for the original and for the modified problem, respectively. Second, any supportable quality perception pair is in the original model supportable by a unique test, whereas this is not the case for the modified model. The additional type adds a degree of freedom for the principal. Any quality perception pair in the interior of the grey area in Figure 5(b) is supportable by multiple tests. Surprisingly, the modified test design problem is nevertheless equivalent to a quality perception design problem. Moreover, it features similar properties as the original quality perception design problem.

Lemma 5 *Consider the modified model and fix any participation threshold $s \in [\underline{\theta}, \overline{\theta})$ which is inducible by some test. (a) The principal's problem to design a binary test is equivalent to a problem where she designs a quality perception pair (μ_1, μ_2). (b) The principal's preferences over quality perception pairs (μ_1, μ_2) are as in the original model. (c) The threshold agent's preferences over quality perception pairs (μ_1, μ_2) among which the principal is indifferent are as in the original model. (d) The threshold agent's preferences over quality perception pairs $(\mu_1, 1)$ are as in the original model.*

Two properties are responsible for the result: First, as the ex ante expected quality perception induced by any inducible quality perception pair is fix because updating is Bayesian, the probabilities with which the different test results occur from an ex ante perspective depend only through the induced quality perception pair (μ_1, μ_2) on the test design. This implies that the principal's preferences over tests depend only through the induced quality perception pair (μ_1, μ_2) on the test design and that they are as in the original model. Second, as an adapted version of Lemma 3 extends to the modified setting, an analogous reasoning applies for the agent's interim preferences over tests among which the principal is indifferent.

Intuition for the structure of the optimal binary test. Lemma 5 allows us to employ a similar strategy of proof for the derivation of the optimal binary test which induces a given participation threshold s as in the original model, but we have to take care of the supportability problem. The problem is that if we transform a given test as in the original model (see the movements indicated by the green arrows in Figure 5(a)), we might leave the set of supportable quality perception pairs. This problem can be avoided by considering only small movements in the quality perception space. For any given test which induces a quality perception pair which lies below the upper supportability frontier in the (μ_1, μ_2)–space (that is, below the blue curve in Figure 5(b)), we can find a test which is better for the principal and equally good for the threshold agent by making a small movement on the principal's indifference curve to the northeast followed by a small movement to the northwest (see the movements indicated by the green arrows in Figure 5(b)). It follows from this that only quality perception pairs on the upper supportability frontier can be optimal. As each point on this frontier is uniquely induced by either a test $T^{(g,y)}(\rho_g)$ or a test $T^{(g,y)}(\rho_n, 0)$, the optimal binary test must have one of these structures.

Derivation of the structure of the optimal possibly non–binary test. If we apply here the iteration procedure from the original model, a further problem might arise: Suppose a test $T \in \mathcal{T}^3$ which implies the quality perception vector μ with $\mu_1 < \mu_2 < \mu_Y < \mu_3$ is given. While it is in the original model always possible to modify the conditional probabilities that are responsible for the two lowest test results such that we obtain a quality perception vector μ' with $\mu'_1 > \mu_1$ and $\mu'_2 = \mu_3$, this might not be possible for the modified model. For instance, when μ_1 and μ_2 arise only from the pooling of $\widehat{\omega} = b$ with $\widehat{\omega} = n$, it is not possible to obtain $\mu'_2 > \mu_Y$ by shifting only probability mass between the conditional probabilities that affect the lowest two test results. The iteration procedure has thus to be adapted.

What we can do instead is the following: Step 1: We increase the lowest two quality perceptions until the second lowest quality perception hits μ_Y. Step 2: We stop modifying the lower two quality perceptions and continue by modifying the higher two quality perceptions. This leads to a test with one quality perception in $[0, \mu_Y)$, $[\mu_Y, 1)$ and $\{1\}$, respectively. Step 3: By considering now again modifications of the lower two quality perceptions, we obtain basically a binary problem to which the logic explained above applies. We so obtain either a test $T^{(g,y)}(\rho_g)$ or a test $T^{(g,y)}(\rho_n, \rho_g)$ which is equally good for the threshold agent, but which is better for the principal than the initial test. The following proposition characterizes the structure of the optimal test:

Proposition 4 *Consider the modified model and fix any participation threshold $s \in [\underline{\theta}, \overline{\theta})$ which is inducible by some test. (a) If $U_s(T^{(g,y)}(1), \mu(x^s, T^{(g,y)}(1))) \leq u(\mu_N)$, the test $T^{(g,y)}(\rho_g^*)$ where ρ_g^* is the unique solution to $U_s(T^{(g,y)}(\rho_g^*), \mu(x^s, T^{(g,y)}(\rho_g^*))) = u(\mu_N)$ is optimal. (b) If $U_s(T^{(g,y)}(1), \mu(x^s, T^{(g,y)}(1))) > u(\mu_N)$, a test $T^{(g,y)}(\rho_n, \rho_g)$ is optimal.*

The intuition for the case distinction is the following: If the threshold agent's signaling motive is weak, the test that would be optimal without limits to information generation (that is, for $\gamma = 0$) exhibits a substantial amount of pooling on the lower test result anyway. As the case where $\widehat{\omega} = n$ which is inherently noisy can be used to obtain the required pooling, the constraints which limit the informativeness of the test do not impose any binding constraints on the test design problem. A test $T^{(g,y)}(\rho_g)$ which issues only two test results and which is not subject to false positives is optimal. However, if the threshold agent's signaling motive is strong, more information can be generated by tests which issue an additional test result. A test $T^{(g,y)}(\rho_n, \rho_g)$ is then optimal. In particular, if the threshold agent's signaling motive is sufficiently strong, the test $T^{(g,y)}(1, 1)$ which perfectly reveals $\widehat{\omega}$ is optimal.

6. Discussion of the related literature

Our article contributes mainly to the literature studying the design of information structures like it is the case in the literature on test design, persuasion and certification. Three types of players are relevant in this literature: A sender who is privately informed about the variable of interest, say his quality, a receiver who is interested in the sender's quality, and a designer who determines the information structure. The role of the designer might either be assumed by the ex ante or the ex interim self of the sender, by the receiver, or by an intermediary. The literature can be classified according to what the designer maximizes and which constraints she faces.

Most closely related to our article is independent work by Harbaugh and Rasmusen (2013). The designer strives for information generation and is subject to voluntary participation as in our article, but the sender is modeled differently such that a different kind of incentive problem arises. The sender is perfectly informed about his continuous quality, risk–neutral with respect to how his quality is perceived, and he suffers an exogenously given constant cost from participation. Manipulating how a participating sender's quality is perceived on average is the only instrument for setting participation incentives. By contrast, we model the sender to be imperfectly informed and averse to perception risk. Participation in any informative test implies a non–trivial perception risk causing an endogenous cost of participation. The cost depends on the sender's private information and it is affected by the test design. Perception risk considerations which are mute in their article are the central theme in ours. In particular, we find that it is more important for setting optimal participation incentives to reduce the perception risk faced by the marginal participant than to improve how his quality is perceived on average.

Caplin and Eliaz (2003) study the design of a pass–fail test which is capable of stopping the spread of HIV when participation in the test is voluntary and when agents condition their sexual matching behavior on the generated information. In contrast to our model, unfavorable test results can be hidden and agents possess no private information. Participation is nevertheless an issue as agents suffer a psychological cost from learning. Setting participation incentives might require the test to be inaccurate. Stopping the spread of the disease requires that the test is never passed by an agent who is actually infected. The optimal test has thus for different reasons a similar structure as in our article.

Gill and Sgroi (2012) and Li and Li (2013) study problems in which a sender who is endowed with binary private information can use an information structure as a signal.[23] Gill and Sgroi (2012) investigate the design of a test and a pricing strategy by a profit–maximizing monopolist. As the low–quality monopolist has always an incentive to pool, the role of the designer is effectively assumed by the high–quality monopolist who benefits from distinguishing herself from the low–quality monopolist. The monopolist is restricted to an exogenously given class of imperfect perfect pass–fail tests. Either the toughest or the softest feasible test is optimal. By contrast, the designer is in our model restricted to a class of tests which derives endogenously from the participation constraint. The toughest feasible test is generally optimal. Li and Li (2013) consider in their base model the problem of a political candidate who is endowed with a binary private signal which is imperfectly informative about her own and her rival's binary quality. She benefits from her quality being perceived as better relative to that of her rival and she can affect perceptions by investing in a positive or a negative campaign of endogenous accuracy. The problem can be interpreted as a test design problem which is subject to technological constraints.

In the literature on persuasion, the role of the designer is assumed by the sender's ex ante self (see Kamenica and Gentzkow (2011)).[24] The information structure is chosen to manipulate a receiver's belief such that he takes decisions which are in the sender's best interest. Bayes' Law restricts the expected value of the posterior belief, but the designer is not subject to any further constraints. Besides the difference in objectives, our analysis requires different techniques as additional constraints arise from the participation decision and potentially also from technological restrictions on information generation. Schweizer and Szech (2013) study a

[23]Interim design decisions are also studied in some other articles: Gill and Sgroi (2008) consider a herding problem with testing. A principal who is perfectly informed about her binary quality can decide to get publicly tested before facing a stream of agents whose endorsement she seeks. She is restricted to a class of imperfect pass–fail tests of varying toughness. Titman and Trueman (1986) study a problem where an informed seller chooses costly test accuracy. Perez-Richet and Prady (2012) analyze a persuasion problem where the informed sender chooses complexity (= the cost of test accuracy), whereas the receiver chooses the level of understanding (= the test accuracy) after observing the complexity.

[24]See Rayo and Segal (2010) for a persuasion problem with a two–dimensional type, Wang (2012) for a problem with multiple receivers who take a collective decision by voting, and Ostrovsky and Schwarz (2010) for a problem with competition in which schools try to improve the average placement of their students in given jobs.

persuasion problem in which the design goal is related to information generation as the designer is basically the ex ante self of the receiver. A conflict of interest which necessitates "persuasion" arises because the ex ante self suffers from an anticipatory cost of learning. Under plausible conditions, a pass–fail test which is not subject to false positives is optimal as in our article.[25]

In the literature on certification, an information–generation technology is offered at a fee by a profit–maximizing intermediary. A part of this literature is concerned with the coarseness of the revealed information. Profit maximization implies design goals which differ from information generation. Lizzeri (1999) shows in his seminal article that a monopolistic intermediary who charges an informed seller for the certification has an incentive to reveal only very limited information. The intermediary can capture the entire informational surplus in the market despite revealing no information.[26]

Some important features of our test design problem play also a role in other problems. Ottaviani and Prat (2001) investigate the public revelation of information by a monopolistic seller to an imperfectly informed buyer. Affiliation of the information with the buyer's private signal allows for an interpretation as "quality information". It turns out that the seller benefits from the revelation of any such information.[27] By contrast, we take an environment in which the designer of an information revelation technology benefits from the revelation of any quality information as given and study the problem where she cannot unilaterally impose an information structure but depends on the voluntary participation of an agent.

That endogenous participation affects the value of non–participation is also important in Tirole (2012) and Philippon and Skreta (2012) who analyze mechanism design problems in the context of government interventions in financial markets. The designer faces banks who are privately informed about the quality of their assets and whose participation in the mechanism is voluntary. Although the participation behavior affects the value of non–participation like in our article, the articles differ in the instruments that are studied. Whereas the designer lacks the ability to test quality in their articles, she lacks the ability to exploit features of "mechanisms" in ours.

The revelation of information about an agent to a third party is also analyzed for different screening problems. See Calzolari and Pavan (2006a) for a monopoly problem with resale, Calzolari and Pavan (2006b) for a problem where two principals contract sequentially with the same agent, and Pancs (2014) for a problem where information about a trader's sales order can be transmitted through an electronic communication network to potential buyers. The articles consider generalized mechanisms which include a rule determining which information is

[25] Benoît and Dubra (2004) study a problem which can be interpreted as a (non–standard) persuasion problem. Technically, the sender with the median quality (instead of the sender's ex ante self) chooses between two inaccurate information structures. It is shown that even when the median sender would like his quality to be perfectly revealed, she might prefer a less accurate over a more accurate information structure.

[26] See also Peyrach and Quesada (2004) and Farhi et al. (2013). See Dranove and Jin (2010) for a survey.

[27] Milgrom and Weber (1982) obtain a similar result for the public revelation of affiliated information by an auctioneer prior to conducting different kinds of auction.

disclosed to a third party. In contrast to our article, information disclosure is not costless but requires the payment of an information rent to the agent who is in possession of the information.

In the literature studying sorting theories of education, an individual is concerned with how his productivity is perceived as competitive firms are willing to pay him his perceived expected productivity as wage. Spence (1973) focuses on the signaling role of education in an environment where schooling is less expensive for more able individuals. Arrow (1973) and Stiglitz (1975) point to the role of schooling as an information generating device and a costly signal at the same time. While Stiglitz (1975) notes already that an individual's incentive to participate in a costless, accurate test is non–trivial when he is imperfectly informed and risk–averse, we show that setting participation incentives may necessitate the generation of inaccurate information and we characterize how inaccuracies are optimally introduced when the objective is learning.

Daley and Green (2014) and Alós-Ferrer and Prat (2012) introduce an information generation technology in a signaling model à la Spence as a second source of information. The sender faces a trade–off between the two channels through which information can be transmitted. Although information can in our model also be transmitted through two channels, the sender faces no trade–off between signaling and information–generation through testing. We derive the information–generation technology which optimally exploits the sender's signaling motive.

7. Discussion of applications

Our first application is motivated by the literature studying the sorting role of **education**. At issue is whether an individual's productivity is high, say 1, or low, say 0. The individual is imperfectly informed about his productivity and can either apply directly for a job or go to an institution of higher education first. The institution generates information about the individual's productivity without affecting it. After the completion of education, competitive firms offer the individual his expected perceived productivity $\mu_{\widehat{\sigma}}$ as wage. The individual is risk–averse in his monetary income and the educational institution is designed to maximize the available information.

Our analysis suggests that an institution of higher education which is hard to pass with a good grade is optimal.[28] The individual never passes with a good grade when his productivity is low and he might even not pass with a good grade when it is high. Interestingly, although it is possible to foster more participation by inflating grades, this is not optimal. Participation is optimally incentivized by reducing the stigma of failure by making it even harder to pass with a good grade.

Our second application concerns the design of **medical tests**. At issue is whether a patient has a certain disease, say HIV. Information generation is possible through a test, testing requires

[28] As the interpretation of grades is endogenous, it does not matter whether it is "hard to pass with a good grade" and individuals who do not pass with a good grade get a "bad grade", or whether it is hard to "pass" and individuals who do not pass "fail".

in medical contexts however typically the agreement of the patient. According to stylized facts, participation rates in medical tests are often low even when testing is costless. A major reason for declining medical testing are psychological costs which come along with learning (see Lyter et al. (1987)). Moreover, the knowledge of his behavior and the observation of symptoms allow the patient to draw imperfect inferences about his health status and there exists evidence that such information influences the decision to get tested. For instance, Hull et al. (1988) find that among the patients attending a sexually transmitted disease clinic those who declined testing for HIV where sufficiently more likely to be infected than those who accepted testing. The goal of offering medical testing is typically the generation of information. Inducing full participation might serve as an additional goal when a high–risk group is offered testing. Our analysis applies to the testing of a patient who is behavioral in the sense that his testing decision is driven by its effect on the doctor's probability assessment of his health status, he prefers more favorable assessments over less favorable ones, and he is for psychological reasons averse to the perception risk which comes along with testing.

The optimal testing procedure always leaves the patient with some hope of not having the disease and it sometimes reveals him that he actually does not have it. Moreover, our analysis suggests that there might be a role for inaccurate tests, even when higher accuracy does not come at a cost. Caplin and Eliaz (2003) and Schweizer and Szech (2013) identify further reasons for why tests with these properties may be optimal in medical contexts. There might thus be a deeper role for inaccurate tests which are not subject to false positives. Using this kind of test in practical test design might thus deserve serious consideration.

Our third application concerns financial **stress testing**. Since the financial crisis in 2007, bank stress tests are routinely used to generate information about the viability of banks.[29] The testing procedures are made particularly transparent and the test results are publicly disclosed. Although participation of all major banks is typically desired, participation is often either explicitly voluntary (e.g., due to a lack of authority of the organization conducting the test) or at least implicitly (e.g, as participants have to agree to the general design). How a bank's viability is perceived affects its cost of refinancing. Stress testing typically imposes a non–trivial perception risk on the bank. As the public disclosure of information takes away hedging opportunities, the bank is likely to be averse to such risk (see Hirshleifer (1971)).[30]

We derive the test structure which is optimal when the objective is learning. When the bank's incentive to participate in the stress test for signaling reasons is weak, the optimal test is relatively coarse. It might issue only a binary test result even when it is possible to reveal much more accurate information regarding under which conditions the bank is viable

[29] See, e.g., `http://www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing`.

[30] In a recent working paper, Goldstein and Leitner (2013) study the relation between information disclosure and risk–sharing opportunities in a setting where the designer maximizes the bank's expected utility and where the bank suffers a loss when its future capital falls below a certain threshold. Besides the difference in objective functions, their analysis depends crucially on the fact that attention is restricted to cases where any private information the bank might possess can be perfectly retrieved through a stress test.

and under which it is not. The optimal testing procedure tests for viability. The desire to induce full participation limits the informativeness of the test but does not affect the structure of the generated information.[31] Although information generation is not necessarily the only goal of stress testing, our analysis suggests that it might not be conflicting with other goals like the avoidance of very bad news (which might cause the immediate breakdowns of banks) and making interventions into banks which fail the test relatively cheap (as the optimal test induces a relatively mild stigma of failure).

8. Conclusion

This article studies optimal test design under voluntary participation of an imperfectly informed, perception risk–averse agent when the principal benefits from information. A binary test which is not subject to false positives is optimal. The probability of false negatives serves as an instrument to foster participation. Learning about the agent's quality is generally imperfect either due to less than full participation, inaccuracy of the optimal test, or both. Furthermore, we have demonstrated how our techniques can be adapted to study more complicated problems with non–binary quality types.

We believe that different extensions which are beyond the scope of this article and which bring the test design problem closer to mechanism design problems deserve further study. In the first extension, tests can not only condition on the agent's quality, but also on announcements of him. For example, the agent might self–select into tests of different difficulty. Although setting participation incentives requires then still learning to be imperfect, the extension allows it to fine–tune the structure of the generated information by exploiting the agent's signaling motive further.[32] In the second extension, the principal can use monetary transfers from participants to non–participants (or vice versa) as another instrument to affect participation incentives and—if we allow also for the first extension—also to set announcement incentives. Studying the generalized problem allows it to assess the relative importance of different instruments which can be used to foster participation.[33]

[31] That stress tests which are subject to voluntary participation exhibit only limited informativeness is also consistent with the design of other kinds of stress tests. In the aftermath of the nuclear accident that occurred at the Fukushima nuclear power plant on 11 March 2011, the European council decided that the safety of all EU nuclear power plants should be reviewed on the basis of comprehensive stress tests. The procedure should be transparent and the results should be made publicly available. The aim was generating information in the hope to restore public confidence in the European plants' safety. Although full participation was desired, the stress tests were conducted on a voluntary basis by the participating countries. Interestingly, Fukushima–like szenarios were not considered (see `http://www.greenpeace.org/eu-unit/en/Publications/2012/stress-tests-briefing/`).

[32] Bar et al. (2012) study the enrollment of students who differ in taste and quality in courses of different difficulty. They study the effect of providing information about the course difficulty to employers but not the design of course difficulties which maximizes information generation.

[33] Such an extension would make the design problem more closely related to non–standard mechanism design problems like those considered by Calzolari and Pavan (2006a), Calzolari and Pavan (2006b), Pancs (2014) and Eső and Szentes (2007). In these articles, the design of a standard mechanism is combined with the design of an information disclosure rule.

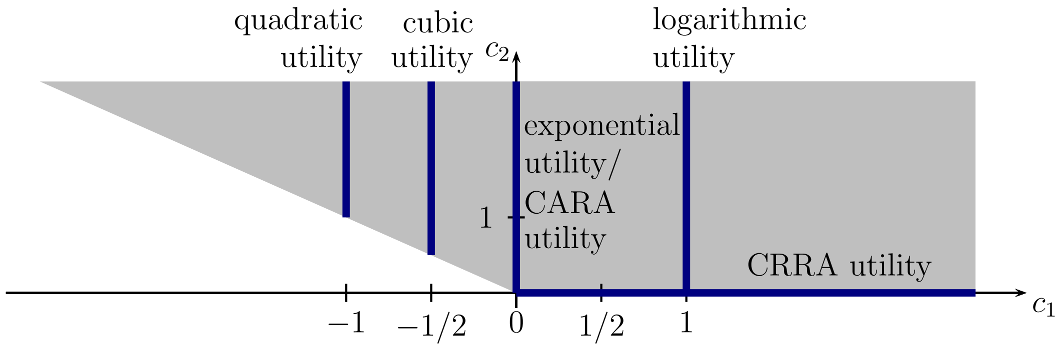


Figure 6: Parameter space and special cases of HARA utility

Appendix

Parametrization of HARA utility. The parameterized class of utility functions

$$u(\mu_{\widehat{\sigma}}) = \begin{cases} -\frac{1}{1-c_1}(c_1\mu_{\widehat{\sigma}} + c_2)^{-\frac{1}{c_1}+1} & \text{if } c_1 \in \mathbf{R}\backslash\{0,1\} \text{ and } c_2 \in [\max\{-c_1, 0\}, \infty) \\ -c_2 \exp(-\frac{1}{c_2}\mu_{\widehat{\sigma}}) & \text{if } c_1 = 0 \text{ and } c_2 \in (0,\infty) \\ \ln(\mu_{\widehat{\sigma}} + c_2) & \text{if } c_1 = 1 \text{ and } c_2 \in (0,\infty) \end{cases}$$

exhibits hyperbolic absolute risk aversion (HARA). The Arrow–Pratt measure of absolute risk–aversion is given by $-u''(\mu_{\widehat{\sigma}})/u'(\mu_{\widehat{\sigma}}) = 1/(c_1\mu_{\widehat{\sigma}} + c_2)$. The class includes as special cases quadratic utility ($c_1 = -1$), cubic utility ($c_1 = -1/2$), exponential/CARA utility ($c_1 = 0$), logarithmic utility ($c_1 = 1$) and CRRA utility ($c_2 = 0$). The parameter space is illustrated by the grey area in Figure 6. The special cases are indicated by the blue line segments. $u(\cdot)$ is increasing and concave for any parameter combination in the grey area. $u(\cdot)$ is weakly prudent for any parameter combination with $c_1 \geq -1$. The analysis in the paper applies thus to any HARA utility function with $c_1 \geq -1$.

Proof to Lemma 1. Fix any test $T \in \mathcal{T}$.

(a) Fix any participation strategy $x \in \mathcal{X}$. $U_\theta(T, \mu(x,T))$ is linear in θ and thus continuous. The coefficient of θ is given by

$$\begin{aligned} & \sum_{\sigma}(p_g^\sigma - p_b^\sigma)u(\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma)) \\ = & \sum_{\sigma'}(p_g^{\sigma'}\sum_{\sigma''}p_b^{\sigma''} - p_b^{\sigma'}\sum_{\sigma''}p_g^{\sigma''})u(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) \\ = & \sum_{\sigma'}\sum_{\sigma''\neq\sigma'}(p_g^{\sigma'}p_b^{\sigma''} - p_b^{\sigma'}p_g^{\sigma''})u(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) \\ = & \sum_{\sigma'}\sum_{\sigma''>\sigma'}(p_g^{\sigma'}p_b^{\sigma''} - p_b^{\sigma'}p_g^{\sigma''})(u(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) - u(\mu_{\sigma''}(\mu_Y(x), p_b^{\sigma''}, p_g^{\sigma''}))). \qquad (5) \end{aligned}$$

$\mu_{\sigma'}(\cdot) < \mu_{\sigma''}(\cdot)$ (resp. $\mu_{\sigma'}(\cdot) \leq \mu_{\sigma''}(\cdot)$) is equivalent to $p_g^{\sigma'}p_b^{\sigma''} - p_b^{\sigma'}p_g^{\sigma''} < 0$ (resp. $p_g^{\sigma'}p_b^{\sigma''} - p_b^{\sigma'}p_g^{\sigma''} \leq 0$). As $u(\cdot)$ is strictly increasing, each summand in (5) is weakly positive. As $\mathcal{T}$ includes only informative tests, $\mu_{\sigma'}(\cdot) < \mu_{\sigma''}(\cdot)$ for some $\sigma', \sigma'' \in \Sigma$. This implies that at least one summand in (5) is strictly positive. Hence, (5) is strictly positive rendering $U_\theta(T, \mu(x,T))$ strictly increasing in θ.

(b) This is a direct consequence of Part (a) and of $u(\mu_N(x))$ not depending on θ.

(c) We distinguish three cases:

Case 1: $U_s(T,\mu(x^s,T)) \geq u(\mu_N(x^s))$ for $s = \underline{\theta}$. $x^{\underline{\theta}} \in \mathcal{X}^\star(T)$ follows directly from Part (a).

Case 2: $U_s(T,\mu(x^s,T)) \leq u(\mu_N(x^s))$ for $s = \overline{\theta}$. $x^{\overline{\theta}} \in \mathcal{X}^\star(T)$ follows directly from Part (a).

Case 3: Neither the supposition in Case 1 nor that in Case 2 is true. Continuity of $F(\cdot)$ implies that $\mu_Y(x^s)$ and $\mu_N(x^s)$ are continuous in s. Furthermore, continuity of $u(\cdot)$ implies that $U_\theta(T,\mu(x^s,T))$ and $u(\mu_N(x^s))$ are continuous in s. From this and the supposition it follows by an Intermediate Value Theorem that there exists $s \in (\underline{\theta},\overline{\theta})$ such that $U_s(T,\mu(x^s,T)) = u(\mu_N(x^s))$. $x^s \in \mathcal{X}^\star(T)$ follows from Part (a). q.e.d.

Proof to Lemma 2.

(a) "$\subseteq$" *Suppose $s \in \mathcal{S}_a$.* That is, there exists $T \in \mathcal{T}_a$ such that $x^s \in \mathcal{X}^\star(T)$. Assume first to the contrary that $s = \underline{\theta}$. The threshold agent's expected utility from participation is then $U_{\underline{\theta}}(T,\mu(x^{\underline{\theta}},T)) = (1-\underline{\theta})u(0) + \underline{\theta}u(1)$ and his utility from non–participation is $u(\mathbf{E}_\theta[\theta|\theta \leq \underline{\theta}]) = u(\underline{\theta})$. As $(1-\underline{\theta})u(0) + \underline{\theta}u(1) < u(\underline{\theta})$ by Jensen's inequality and strict concavity of $u(\cdot)$, the threshold agent has a strict incentive not to participate. As this contradicts that the participation threshold is $s = \underline{\theta}$, $s \in (\underline{\theta},\overline{\theta})$ must be true. Lemma 1 (a) and $x^s \in \mathcal{X}^\star(T)$ with $s \in (\underline{\theta},\overline{\theta})$ imply $U_s(T,\mu(x^s,T)) = u(\mathbf{E}_\theta[\theta|\theta \leq s])$. As $U_s(T,\mu(x^s,T)) = (1-s)u(0) + su(1)$ for any $T \in \mathcal{T}_a$, we obtain $(1-s)u(0) + su(1) = u(\mathbf{E}_\theta[\theta|\theta \leq s])$.

"$\supseteq$" *Suppose $(1-s)u(0) + su(1) = u(\mathbf{E}_\theta[\theta|\theta \leq s])$ for some $s \in [\underline{\theta},\overline{\theta})$.* As for any $T \in \mathcal{T}_a$ the left–hand side corresponds to $U_s(T,\mu(x^s,T))$ and as the right–hand side corresponds to $u(\mu_N(x^s))$, the agent with private signal s is indifferent between participation and non–participation in any test $T \in \mathcal{T}_a$. Lemma 1 (a) implies $x^s \in \mathcal{X}^\star(T)$ for any $T \in \mathcal{T}_a$. Hence, $s \in \mathcal{S}_a$.

(b) "$\subseteq$" *Suppose $s \in \mathcal{S}\backslash\mathcal{S}_a$.* That is, there exists $T \in \mathcal{T}\backslash\mathcal{T}_a$ such that $x^s \in \mathcal{X}^\star(T)$. We distinguish two cases: *Case 1: $s = \underline{\theta}$.* By Jensen's inequality and strict concavity of $u(\cdot)$, $u(s) > (1-s)u(0) + su(1)$. As $u(s) = u(\mathbf{E}_\theta[\theta|\theta \leq s])$ for $s = \underline{\theta}$, we obtain the result. *Case 2: $s \in (\underline{\theta},\overline{\theta})$.* Lemma 1 (a) and $x^s \in \mathcal{X}^\star(T)$ with $s \in (\underline{\theta},\overline{\theta})$ imply that $U_s(T,\mu(x^s,T)) = u(\mathbf{E}_\theta[\theta|\theta \leq s])$. Sufficient for the result is thus $U_s(T,\mu(x^s,T)) > (1-s)u(0) + su(1)$. By definition, $U_s(T,\mu(x^s,T)) = \sum_\sigma p_s^\sigma u(\mu_\sigma(\mu_Y(x^s),p_b^\sigma,p_g^\sigma))$. By applying Jensen's inequality to the right–hand side, $U_s(T,\mu(x^s,T)) > \sum_\sigma p_s^\sigma((1-\mu_\sigma(\mu_Y(x^s),p_b^\sigma,p_g^\sigma))u(0) + \mu_\sigma(\mu_Y(x^s),p_b^\sigma,p_g^\sigma)u(1))$. Because $u(1) > u(0)$, sufficient for what we have to show is that the coefficient of $u(1)$, $p_s^\sigma\mu_\sigma(\mu_Y(x^s),p_b^\sigma,p_g^\sigma)$, is larger than s. We can write

$$\begin{aligned}
& \sum_\sigma p_s^\sigma \mu_\sigma(\mu_Y(x^s),p_b^\sigma,p_g^\sigma) - s \\
= & \sum_\sigma p_s^\sigma \frac{\mu_Y(x^s)p_g^\sigma}{\mu_Y(x^s)p_g^\sigma + (1-\mu_Y(x^s))p_b^\sigma} - \sum_\sigma s p_g^\sigma \\
= & \sum_\sigma \frac{sp_g^\sigma + (1-s)p_b^\sigma}{\mu_Y(x^s)p_g^\sigma + (1-\mu_Y(x^s))p_b^\sigma}\mu_Y(x^s)p_g^\sigma - \sum_\sigma \frac{\mu_Y(x^s)p_g^\sigma + (1-\mu_Y(x^s))p_b^\sigma}{\mu_Y(x^s)p_g^\sigma + (1-\mu_Y(x^s))p_b^\sigma} s p_g^\sigma \\
= & \sum_\sigma \frac{\mu_Y(x^s) - s}{\mu_Y(x^s)p_g^\sigma + (1-\mu_Y(x^s))p_b^\sigma} p_b^\sigma p_g^\sigma. \qquad (6)
\end{aligned}$$

As $s < \mu_Y(x^s)$ for any $s \in (\underline{\theta},\overline{\theta})$, (6) is strictly positive. This implies the result.

"$\supseteq$" *Suppose* $(1-s)u(0)+su(1) < u(\mathbf{E}_\theta[\theta|\theta \le s])$ *for some* $s \in [\underline{\theta}, \overline{\theta})$. As $s \notin \mathcal{S}^a$ follows from Part (a), we only need to show that $s \in \mathcal{S}$. Let $T : (0,1] \to \mathcal{T}$ be any continuous function such that $T(1) \in \mathcal{T}_a$ and such that $T(\rho)$ converges to an uninformative test as $\rho \to 0$. $T(\rho)$ is a function which continuously transforms a completely uninformative test into an accurate test. Examples for such functions are $T^g(\rho)$, $T^b(\rho)$ and $T^{gb}(\rho)$. By the supposition, $U_s(T(1), \mu(x^s, T(1))) < u(\mu_N(x^s))$. Moreover, by construction, $\lim_{\rho\to 0} U_s(T(\rho), \mu(x^s, T(\rho))) = u(\mu_Y(x^s)) > u(\mu_N(x^s))$. As ρ transforms $U_s(T(\rho), \mu(x^s, T(\rho)))$ continuously, there exists by an Intermediate Value Theorem a $\rho^* \in (0,1)$ such that $U_s(T(\rho^*), \mu(x^s, T(\rho^*))) = u(\mu_N(x^s))$. Lemma 1 (a) implies $x^s \in \mathcal{X}^\star(T(\rho^*))$. Hence, $s \in \mathcal{S}$.

(c) This follows directly from "$\supseteq$" in Part (b). q.e.d.

Proof to Lemma 3. Fix any participation threshold $s \in \mathcal{S}$. We argue in three steps.

Step 1: The principal's expected utility. Consider any test $T \in \mathcal{T}$. Bayes' Law implies that the first moment of the quality perception distribution conditional on participation is constant: $\sum_\sigma p^\sigma_{\mu_Y} \mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g) = \mu_Y$. Moreover, by using that quality perceptions are formed according to Bayes' Law and by simplifying, we obtain that the second moment of this distribution is given by

$$\begin{aligned}\sum_\sigma p^\sigma_{\mu_Y} \mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g)^2 &= \sum_\sigma (\mu_Y p^\sigma_g + (1-\mu_Y)p^\sigma_b) \left(\frac{\mu_Y p^\sigma_g}{\mu_Y p^\sigma_g + (1-\mu_Y)p^\sigma_b}\right)^2 \\ &= \sum_\sigma \mu_Y p^\sigma_g \mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g) = \mu_Y E_1(T, \mu(x^s, T)).\end{aligned}$$

These two properties allow us to write the principal's objective function in (2) as

$$F(s)v(\mu_N) + (1-F(s)) \left[\theta_0^2 - 2\theta_0\mu_Y + \mu_Y E_1(T, \mu(x^s, T))\right]. \tag{7}$$

The principal's objective function depends thus only through $E_1(T, \mu(x^s, T))$ on the test design.

Step 2: The agent's interim expected perception. As $E_\theta(T, \mu(x^s, T))$ is linear in θ, we can rewrite it as a linear combination of $E_{\mu_Y}(T, \mu(x^s, T))$ and of $E_1(T, \mu(x^s, T))$:

$$E_\theta(T, \mu(x^s, T)) = \sum_\sigma \left[\frac{1-\theta}{1-\mu_Y} E_{\mu_Y}(T, \mu(x^s, T)) + \frac{\theta - \mu_Y}{1-\mu_Y} E_1(T, \mu(x^s, T))\right].$$

As $E_{\mu_Y}(T, \mu(x^s, T)) = \mu_Y$ by Bayes' Law, the agent's interim expected perception depends also only through $E_1(T, \mu(x^s, T))$ on the test design.

Step 3: The agent's interim expected perception when the principal's expected utility is constant. It is a direct consequence of Steps 1 and 2 that the agent's interim expected perception is constant for tests between which the principal is indifferent. q.e.d.

Proof to Lemma 4.

(a) Fix any participation threshold $s \in \mathcal{S}$.

Step 1: The effect of test accuracy on quality perceptions. The test $T^g(\rho)$ induces the quality perceptions

$$\mu_1(\mu_Y, 1, 1-\rho) = \frac{\mu_Y(1-\rho)}{\mu_Y(1-\rho) + (1-\mu_Y)} \tag{8}$$

and $\mu_2(\mu_Y, 0, \rho) = 1$. It follows

$$\frac{d}{d\rho}\mu_1(\mu_Y, 1, 1-\rho) = \frac{-\mu_Y(1-\mu_Y)}{(\mu_Y(1-\rho)+(1-\mu_Y))^2} = -\frac{\mu_Y}{1-\mu_Y\rho}(1-\mu_1(\mu_Y, 1, 1-\rho)) \tag{9}$$

and $\frac{d}{d\rho}\mu_2(\mu_Y, 0, \rho) = 0$.

Step 2: The effect of test accuracy on the principal's expected utility. By Step 1, $\mu_1(\mu_Y, 1, 1-\rho)$ and $\mu_2(\mu_Y, 0, \rho)$ move away from each other as ρ increases. Because updating is Bayesian, this happens without affecting the expected quality perception conditional on participation. That is, an increase in ρ transforms the quality perception distribution conditional on participation by a mean–preserving spread. As the principal evaluates quality perceptions at the strictly convex function $v(\cdot)$, she strictly benefits from such mean–preserving spreads. That is, $V(T^g(\rho), x^s, \mu(x^s, T^g(\rho)))$ is strictly increasing in ρ.

Step 3: The effect of test accuracy on the threshold agent's interim expected utility from participation. As ρ does not transform the quality perception distribution which the threshold agent faces when he participates by a mean–preserving spread, we have to apply a different reasoning for the threshold agent. We have $U_\theta(T^g(\rho), \mu(x^s, T^g(\rho))) = (1-\theta\rho)u(\mu_1) + \theta\rho u(1)$ with $\mu_1 = \mu_1(\mu_Y, 1, 1-\rho)$ such that

$$\begin{aligned}
\frac{d}{d\rho}U_\theta(T^g(\rho), \mu(x^s, T^g(\rho))) &= \theta(u(1)-u(\mu_1)) - (1-\theta\rho)\frac{\mu_Y}{1-\mu_Y\rho}(1-\mu_1)u'(\mu_1) \\
&< \theta(1-\mu_1)u'(\mu_1) - (1-\theta\rho)\frac{\mu_Y}{1-\mu_Y\rho}(1-\mu_1)u'(\mu_1) \\
&= -(\mu_Y-\theta)\frac{1-\mu_1}{1-\mu_Y\rho}u'(\mu_1).
\end{aligned} \tag{10}$$

The first equality follows from differentiating and from using (9). The inequality follows from using that $u'(\mu_1) > (u(1)-u(\mu_1))/(1-\mu_1)$ by strict concavity of $u(\cdot)$. The second equality follows from simplifying. As $\mu_Y - \theta > 0$ for $\theta = s$, (10) implies that $U_s(T^g(\rho), \mu(x^s, T^g(\rho)))$ is strictly decreasing in ρ.

(b) Fix any participation threshold $s \in \mathcal{S}\backslash\mathcal{S}^a$. First, Lemma 2 (b) implies that $(1-s)u(0) + su(1) < u(\mathbf{E}_\theta[\theta|\theta \leq s])$. This inequality can be written as $U_s(T^g(1), \mu(x^s, T^g(1))) < u(\mu_N)$. Second, $\lim_{\rho\to 0} U_s(T^g(\rho), \mu(x^s, T^g(\rho))) = u(\mu_Y)$ and $u(\mu_Y) > u(\mu_N)$ imply the inequality $\lim_{\rho\to 0} U_s(T^g(\rho), \mu(x^s, T^g(\rho))) > u(\mu_N)$. Third, because of continuity of $U_s(T^g(\rho), \mu(x^s, T^g(\rho)))$ in ρ and the first two points, an Intermediate Value Theorem applies. It follows that there exists $\rho^* \in (0,1)$ such that $U_s(T^g(\rho^*), \mu(x^s, T^g(\rho^*))) = u(\mu_N)$.

(c) Fix any participation threshold $s \in \mathcal{S}$. As $s < \mu_Y$, it suffices by (4) to show that $E_1(T, \mu(x^s, T)) - E_0(T, \mu(x^s, T))$ with $T = T^g(\rho)$ is strictly increasing in ρ. We have

$$\begin{aligned}
& E_1(T^g(\rho), \mu(x^s, T^g(\rho))) - E_0(T^g(\rho), \mu(x^s, T^g(\rho))) \\
= \; & [(1-\rho)\mu_1(\mu_Y, 1, 1-\rho) + \rho\mu_2(\mu_Y, 0, \rho)] - \mu_1(\mu_Y, 1, 1-\rho) \\
= \; & \rho[1-\mu_1(\mu_Y, 1, 1-\rho)]
\end{aligned}$$

As $\mu_1(\mu_Y, 1, 1-\rho)]$ is by (9) strictly decreasing in ρ, we obtain the result. q.e.d.

Proof to Proposition 1. Fix any participation threshold $s \in \mathcal{S}$.

Step 1: Equivalence of test design and quality perception design problem. The relationship between binary tests $((1-p_b^2, p_b^2), (1-p_g^2, p_g^2))$ and pairs of quality perceptions (μ_1, μ_2) is determined by the system of the two equations $\mu_1(\mu_Y, 1-p_b^2, 1-p_g^2) = \mu_1$ and $\mu_2(\mu_Y, p_b^2, p_b^2) = \mu_2$. Without loss of generality, we can restrict attention to binary tests with $p_b^2 < p_g^2$. It can be easily verified that any binary test with $p_b^2 < p_g^2$ induces a pair of quality perceptions $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$. Conversely, it can be easily verified that any pair of quality perceptions $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$ is induced by the binary test $T^\mu(\mu_1, \mu_2) := ((1-p_b^2, p_b^2), (1-p_g^2, p_g^2))$ with $p_g^2 := (\mu_Y - \mu_1)/(\mu_2 - \mu_1) \cdot \mu_2/\mu_Y$ and $p_b^2 := (\mu_Y - \mu_1)/(\mu_2 - \mu_1) \cdot (1-\mu_2)/(1-\mu_Y)$. Moreover, $T^\mu(\mu_1, \mu_2)$ is the only test which induces (μ_1, μ_2). Choosing $T \in \mathcal{T}^2$ with $p_g^2 > p_b^2$ to maximize the objective function in (2) subject to (3) is thus equivalent to choosing $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$ to maximize the objective function in (2) subject to (3) and $T = T^\mu(\mu_1, \mu_2)$.

Step 2: The principal's indifference curves in the (μ_1, μ_2)–space. Bayes' Law implies that $p_{\mu_Y}^2$ solves $(1 - p_{\mu_Y}^2)\mu_1 + p_{\mu_Y}^2 \mu_2 = \mu_Y$ for any $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$. Hence, $p_{\mu_Y}^2 = (\mu_Y - \mu_1)/(\mu_2 - \mu_1)$. Using this and the quadratic nature of the principal's utility function, we obtain that her expected utility depends only through $(1-p_{\mu_Y}^2)\mu_1^2 + p_{\mu_Y}^2 \mu_2^2 = \mu_Y(\mu_1 + \mu_2) - \mu_1\mu_2$ on (μ_1, μ_2). It follows from the Implicit Function Theorem that the slope of her indifference curves is given by $d\mu_2/d\mu_1 = (\mu_2 - \mu_Y)/(\mu_Y - \mu_1)$. Any indifference curve of the principal can thus be described by a strictly increasing and differentiable function $\mu_2(\mu_1)$ on a support $[0, \overline{\mu}_1]$ with $\mu_2(\overline{\mu}_1) = 1$ and with $\mu_2'(\mu_1) = (\mu_2 - \mu_Y)/(\mu_Y - \mu_1)$.

Step 3: The threshold agent's expected utility from participation strictly increases as we move to the northeast on an indifference curve of the principal. Fix any indifference curve $\mu_2(\mu_1)$ of the principal. Let $\mu(\mu_1) := (\mu_N, \mu_1, \mu_2(\mu_1))$. The threshold agent's expected utility from participation is then $U_s(T^\mu(\mu_1, \mu_2(\mu_1)), \mu(\mu_1)) = (1-p_s^2)u(\mu_1) + p_s^2 u(\mu_2(\mu_1))$. By Lemma 3, there exists a constant $\overline{E}_s$ such that $E_s(T^\mu(\mu_1, \mu_2(\mu_1)), \mu(\mu_1)) = \overline{E}_s$ implying that $p_s^2 = (\overline{E}_s - \mu_1)/(\mu_2(\mu_1) - \mu_1)$. Using that Bayes' Law implies $p_{\mu_Y}^2 = (\mu_Y - \mu_1)/(\mu_2(\mu_1) - \mu_1)$, we can rewrite this as $p_s^2 = p_{\mu_Y}^2 - (\mu_Y - \overline{E}_s)/(\mu_2(\mu_1) - \mu_1)$ and get

$$U_s(T^\mu(\mu_1, \mu_2(\mu_1)), \mu(\mu_1)) = U_{\mu_Y}(T^\mu(\mu_1, \mu_2(\mu_1)), \mu(\mu_1)) - (\mu_Y - \overline{E}_s)\frac{u(\mu_2(\mu_1)) - u(\mu_1)}{\mu_2(\mu_1) - \mu_1} \tag{11}$$

Consider first the second summand on the right–hand side. First, note that strict concavity of $u(\cdot)$ implies that $(u(\mu_2) - u(\mu_1))/(\mu_2 - \mu_1)$ is strictly decreasing in μ_1 and μ_2 when $\mu_1 < \mu_2$. Second, note that $\overline{E}_s < \mu_Y$ by (4). These two points together with $\mu_2'(\mu_1) > 0$ imply that $-(\mu_Y - \overline{E}_s)(u(\mu_2(\mu_1)) - u(\mu_1))/(\mu_2(\mu_1) - \mu_1)$ is strictly increasing in μ_1. It suffices thus to show that

$$U_{\mu_Y}(T^\mu(\mu_1, \mu_2(\mu_1)), \mu(\mu_1)) = \frac{\mu_2(\mu_1) - \mu_Y}{\mu_2(\mu_1) - \mu_1}u(\mu_1) + \frac{\mu_Y - \mu_1}{\mu_2(\mu_1) - \mu_1}u(\mu_2(\mu_1)) \tag{12}$$

is weakly increasing in μ_1. By differentiating (12) with respect to μ_1 and by simplifying, we obtain

$$\begin{aligned}&\left[\frac{\mu_2(\mu_1) - \mu_Y}{\mu_2(\mu_1) - \mu_1}\left(u'(\mu_1) - \frac{u(\mu_2(\mu_1)) - u(\mu_1)}{\mu_2(\mu_1) - \mu_1}\right)\right] \\ &\quad + \mu_2'(\mu_1)\left[\frac{\mu_Y - \mu_1}{\mu_2(\mu_1) - \mu_1}\left(u'(\mu_2(\mu_1)) - \frac{u(\mu_2(\mu_1)) - u(\mu_1)}{\mu_2(\mu_1) - \mu_1}\right)\right].\end{aligned}$$

By using that $\mu_2'(\mu_1) = (\mu_2(\mu_1) - \mu_Y)/(\mu_Y - \mu_1)$ by Step 2 and by simplifying again, this becomes

$$\frac{\mu_2(\mu_1)-\mu_Y}{\mu_2(\mu_1)-\mu_1}\left(u'(\mu_1)+u'(\mu_2(\mu_1))-2\frac{u(\mu_2(\mu_1))-u(\mu_1)}{\mu_2(\mu_1)-\mu_1}\right).$$

It follows that $(u'(\mu_1)+u'(\mu_2(\mu_1)))(\mu_2(\mu_1)-\mu_1)-2(u(\mu_2(\mu_1))-u(\mu_1))\geq 0$ is sufficient for (12) being weakly increasing in μ_1. Our assumption that $u'''\geq 0$ a.e. implies that $\int_0^{\mu_2(\mu_1)-\mu_1}\int_0^{\epsilon}u'''(\mu_1+\epsilon')\epsilon' d\epsilon' d\epsilon\geq 0$. As

$$\begin{aligned}
& \int_0^{\mu_2(\mu_1)-\mu_1}\int_0^{\epsilon}u'''(\mu_1+\epsilon')\epsilon' d\epsilon' d\epsilon \\
= & \int_0^{\mu_2(\mu_1)-\mu_1}\left[u''(\mu_1+\epsilon')\epsilon'-u'(\mu_1+\epsilon')\right]_{\epsilon'=0}^{\epsilon'=\epsilon} d\epsilon \\
= & \int_0^{\mu_2(\mu_1)-\mu_1}\left(u''(\mu_1+\epsilon)\epsilon-(u'(\mu_1+\epsilon)-u'(\mu_1))\right) d\epsilon \\
= & \left[(u'(\mu_1)+u'(\mu_1+\epsilon))\epsilon-2(u(\mu_1+\epsilon)-u(\mu_1))\right]_{\epsilon=0}^{\epsilon=\mu_2(\mu_1)-\mu_1} \\
= & (u'(\mu_1)+u'(\mu_2(\mu_1)))(\mu_2(\mu_1)-\mu_1)-2(u(\mu_2(\mu_1))-u(\mu_1)),
\end{aligned}$$

the assumption constitutes a sufficient condition for (12) begin weakly increasing in μ_1. Hence, (11) is strictly increasing in μ_1.

This implies in particular that the threshold agent strictly prefers the quality perception pair $(\overline{\mu}_1, 1)$ over any other quality perception pair on the principal's indifference curve. Moreover, as quality perception pairs $(\mu_1, 1)$ correspond to tests $T^g(\rho)$, we have shown with this step that for any test $T'\in\mathcal{T}^2\backslash\{T^g(\rho)|\rho\in(0,1]\}$ there exists a test $T''\in\{T^g(\rho)|\rho\in(0,1]\}$ which is equally good for the principal, but which is strictly better for the threshold agent than the test T'.

Step 4: For any test $T'\in\mathcal{T}^2\backslash\{T^g(\rho)|\rho\in(0,1]\}$ which induces s there exists a test $T'''\in\{T^g(\rho)|\rho\in(0,1]\}$ which induces s and which is strictly better for the principal. We distinguish two cases. *Case 1: $s\in\mathcal{S}\backslash\mathcal{S}_a$.* Fix any test $T'\in\mathcal{T}^2\backslash\{T^g(\rho)|\rho\in(0,1]\}$ which induces s. By Step 3, there exists a test $T''=T^g(\rho_g)$ which is equally good for the principal and which is strictly better for the threshold agent than the test T'. By Lemma 4 (b), there exists a test $T'''=T^g(\rho'_g)$ with $\rho'_g\in(0,1)$ which induces s with a binding participation constraint. As the test T''' is strictly worse for the threshold agent than the test T'', it must by Lemma 4 (a) be strictly better for the principal than the test T''. As the principal is by construction indifferent between the tests T' and T'', we have proven the existence of a test $T'''\in\{T^g(\rho)|\rho\in(0,1)\}$ which induces s and which is strictly better for the principal than the test T'. *Case 2: $s\in\mathcal{S}_a$.* As s is under this supposition inducible by the accurate test $T^g(1)$, the result is trivial.

Step 5: Existence and characterization of the optimal binary test. It follows directly from Step 4 that if the problem $\max_{T\in\{T^g(\rho)|\rho\in(0,1]\}}V(T,x^s,\mu(x^s,T))$ subject to (3) possesses a solution, then its maximizer constitutes an optimal binary test. By Lemma 2 (c), (3) holds for some test $T=T^g(\rho)$. By Lemma 4 (a) and continuity of $U_s(T^g(\rho),\mu(x^s,T^g(\rho)))$ in ρ, there exists a largest $\rho\in(0,1]$ such that (3) holds for $T=T^g(\rho)$, say ρ^*. Furthermore, tests $T^g(\rho)$ with a larger ρ are by Lemma 4 (a) better for the principal. Hence, $T^g(\rho^*)$ constitutes the optimal binary test. It remains to characterize ρ^*. If $s\in\mathcal{S}_a$, $\rho^*=1$. If $s\in\mathcal{S}\backslash\mathcal{S}_a$, it follows from Lemma 4 (a) and (b) that ρ^* is the unique solution to $U_s(T^g(\rho^*),\mu(x^s,T^g(\rho^*)))=u(\mu_N)$.q.e.d.

Proof to Proposition 2. Fix any participation threshold $s \in \mathcal{S}$.

Step 1: For any test $T' \in \mathcal{T}^Z$ we construct a test $T''(\alpha^*) \in \mathcal{T}^Z$ with nice properties. Fix any test $T' = ((p_b^1, p_b^2, p_b^3, \ldots, p_b^Z), (p_g^1, p_g^2, p_g^3, \ldots, p_g^Z)) \in \mathcal{T}^Z$ with $Z > 2$. Let $\mu' = \mu(x^s, T')$. Suppose without loss of generality that $\mu_1' < \mu_2' < \ldots < \mu_Z'$. The following table defines a family of tests $T''(\alpha)$ with $\alpha \in [0, (p_g^1 + p_g^2)/p_g^3]$ which differ from test T' only in the probabilities which generate the first two test results:

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$		$\sigma = Z$
$\omega = g$	$p_g^1 + p_g^2 - \alpha p_g^3$	αp_g^3	p_g^3	$\ldots$	p_g^Z
$\omega = b$	$p_b^1 + p_b^2 - \alpha p_b^3$	αp_b^3	p_b^3	$\ldots$	p_b^Z

Note that we slightly abuse notation here as for test $T''(0)$ the test result $\sigma = 2$ occurs with probability zero.

We make two observations: First, the constraints on α are chosen such that all conditional probabilities are non–negative. $T''(\alpha)$ is thus for any $\alpha \in (0, (p_g^1 + p_g^2)/p_g^3]$ indeed a test. The only non–trivial part of this observation is that $p_b^1 + p_b^2 - (p_g^1 + p_g^2)/p_g^3 \cdot p_b^3 > 0$. The inequality can be rewritten as $(p_g^3 p_b^2 - p_g^2 p_b^3) + (p_g^3 p_b^1 - p_g^1 p_b^3) > 0$. $\mu_3' > \mu_2'$ (resp. $\mu_3' > \mu_1'$) implies that the first (resp. second) bracketed expression is strictly positive. Second, by construction, the quality perception distribution induced by $T''((p_g^1 + p_g^2)/p_g^3)$ is a mean–preserving spread of the quality perception distribution induced by T' which is in turn a mean–preserving spread of the quality perception distribution induced by $T''(0)$. As the principal evaluates quality perceptions at a convex function, she prefers the test $T''((p_g^1 + p_g^2)/p_g^3)$ over the test T' and the test T' over the test $T''(0)$. By a continuity argument and an Intermediate Value Theorem, there exists $\alpha^* \in (0, (p_g^1 + p_g^2)/p_g^3)$ such that the principal is indifferent between the tests T' and $T''(\alpha^*)$.

Step 2: The induction step. Under the supposition that the agent prefers for any $Z > 2$ and for any test $T' \in \mathcal{T}^Z$ the test $T''(\alpha^*)$ as constructed in Step 1 over the test T', we can apply the following reasoning: Start with any test $T_1' \in \mathcal{T}^Z$ with $Z > 2$ which induces s. Apply Step 1 with $T' = T_1'$ to construct a test $T_1''(\alpha^*)$. As the second and the third test result of the test $T_1''(\alpha^*)$ induce the same quality perception, we can construct a test $T_2' \in \mathcal{T}^{Z-1}$ by merging these test results. As this neither affects the agent's nor the principal's payoffs, we have constructed a test $T_2' \in \mathcal{T}^{Z-1}$ from the test T_1' which is better for the agent and which is equally good for the principal. By iterating this procedure, we obtain a test $T_{Z-1}' \in \mathcal{T}^2$. As Step 4 in the Proof to Proposition 1 extends straightforwardly to the case in which we start with a test for which the threshold agent has a strict incentive to participate, we can apply Proposition 1 to obtain a test which induces s and which is better for the principal than the test T_1'. It follows that for any non–binary test inducing s there exists a binary test inducing s which is better for the principal.

Step 3: The supposition in Step 2 is true. Fix any test $T \in \mathcal{T}^Z$ with $Z > 2$. Suppose without loss of generality that $\mu_1 < \mu_2 < \ldots < \mu_Z$ for $\mu = \mu(x^s, T)$. $E_\theta(T, \mu) = p_\theta^1 \mu_1 + p_\theta^2 \mu_2 + \sum_{\sigma > 2} p_\theta^\sigma \mu_\sigma$ and $p_\theta^1 = 1 - p_\theta^2 - \sum_{\sigma > 2} p_\theta^\sigma$ imply

$$p_\theta^2 = (1 - \sum_{\sigma > 2} p_\theta^\sigma) \frac{\widetilde{E}_\theta(T, \mu) - \mu_1}{\mu_2 - \mu_1} \text{ and } p_\theta^1 = (1 - \sum_{\sigma > 2} p_\theta^\sigma) \frac{\mu_2 - \widetilde{E}_\theta(T, \mu)}{\mu_2 - \mu_1}$$

with $\widetilde{E}_\theta(T, \mu) := (E_\theta(T, \mu) - \sum_{\sigma > 2} p_\theta^\sigma \mu_\sigma)/(1 - \sum_{\sigma > 2} p_\theta^\sigma)$. By defining

$$\widetilde{U}_\theta(\mu_1, \mu_2, \widetilde{E}_\theta) := \frac{\mu_2 - \widetilde{E}_\theta}{\mu_2 - \mu_1} u(\mu_1) + \frac{\widetilde{E}_\theta - \mu_1}{\mu_2 - \mu_1} u(\mu_2),$$

we can write $U_\theta(T,\mu) = (1-\sum_{\sigma>2} p_\theta^\sigma)\widetilde{U}_\theta(\mu_1,\mu_2,\widetilde{E}_\theta(T,\mu)) + \sum_{\sigma>2} p_\theta^\sigma u(\mu_\sigma)$. Analogously, by defining

$$\widetilde{V}(\mu_1,\mu_2,\widetilde{E}_{\mu_Y}) := \frac{\mu_2 - \widetilde{E}_{\mu_Y}}{\mu_2-\mu_1} v(\mu_1) + \frac{\widetilde{E}_{\mu_Y} - \mu_1}{\mu_2-\mu_1} v(\mu_2),$$

we can write $V(T,x^s,\mu) = F(s)v(\mu_N) + (1-F(s))((1-\sum_{\sigma>2} p_{\mu_Y}^\sigma)\widetilde{V}(\mu_1,\mu_2,\widetilde{E}_{\mu_Y}(T,\mu)) + \sum_{\sigma>2} p_{\mu_Y}^\sigma v(\mu_\sigma))$. The result follows from two observations concerning tests T which differ only in p_b^1, p_b^2, p_g^1 and p_g^2: First, because updating is Bayesian, $\widetilde{E}_{\mu_Y}(T,\mu(T,x^s))$ is constant, say $\widetilde{E}_{\mu_Y}(T,\mu(T,x^s)) = \widetilde{\mu}_Y$. Furthermore, Lemma 3 implies that $\widetilde{E}_s(T,\mu(x^s,T))$ is constant for tests among which the principal is indifferent, say $\widetilde{E}_s(T,\mu(x^s,T)) = \widetilde{\overline{E}}_s$. Second, the considered tests affect the threshold agent's and the principal's expected utility only through $\widetilde{U}_s(\mu_1,\mu_2,\widetilde{\overline{E}}_s)$ and $\widetilde{V}(\mu_1,\mu_2,\widetilde{\mu}_Y)$, respectively. $\widetilde{V}(\mu_1,\mu_2,\widetilde{\mu}_Y)$ is a positive linear transformation of $V(T^\mu(\mu_1,\mu_2),x^s,\mu)$ with μ_Y replaced by $\widetilde{\mu}_Y$. $\widetilde{U}_s(\mu_1,\mu_2,\widetilde{\overline{E}}_s)$ corresponds to $U_s(T^\mu(\mu_1,\mu_2),\mu)$ with $\overline{E}_s$ replaced by $\widetilde{\overline{E}}_s$. Moreover, $\mu_1 < \widetilde{\overline{E}}_s < \widetilde{\mu}_Y < \mu_2$ for the considered tests. It follows that the logic of Steps 2 and 3 in the Proof to Proposition 1 applies to the here considered problem. Hence, the threshold agent prefers the test $T''(\alpha^*)$ over the test T'. This proves that the supposition in Step 2 of this proof is true.

Step 4: Existence of an optimal test. As there exists by the construction in the preceding steps for any non–binary test inducing s a binary test inducing s which is better for the principal, existence of an optimal test follows from the existence of an optimal binary test. q.e.d.

Proof to Proposition 3. Consider $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$ and $F = F_\delta$. Moreover, consider for any inducible participation threshold s the test $T^g(\rho)$ which optimally induces s. If $\mathcal{S}_a$ is non–empty, it contains the highest inducible participation threshold, that is the lowest inducible level of participation (see the Example in Subsection 4.1). Moreover, $s \in \mathcal{S}_a$ is induced by the most accurate test. That any higher level of participation is induced by a less accurate test is thus trivial.

It remains to argue that starting from any participation threshold $s \in \mathcal{S}\backslash\mathcal{S}_a$, an increase in participation implies a decrease in the test accuracy. The threshold agent's participation constraint is by Proposition 1 for any $s \in \mathcal{S}\backslash\mathcal{S}_a$ binding. It is given by $g(\rho,s) := (1-\rho s)u(\mu_1(\mu_Y(x^s),1,1-\rho)) + \rho s u(1) - u(\mu_N(x^s)) = 0$. By the Implicit Function Theorem, $d\rho/ds = -(\partial g(\rho,s)/\partial s)/(\partial g(\rho,s)/\partial\rho)$. As Lemma 4 (a) implies that $\partial g(\rho,s)/\partial\rho < 0$, it remains to show that $\partial g(\rho,s)/\partial s > 0$. We have

$$\begin{aligned}
\frac{\partial}{\partial s} g(\rho,s) &= \rho(u(1) - u(\mu_1(\cdot))) + (1-\rho s)\frac{1}{2}\frac{1-\rho}{(1-\rho\mu_Y(x^s))^2} u'(\mu_1(\cdot)) - \frac{1}{2}u'(\mu_N(x^s)) \\
&> \rho(u(1) - u(\mu_1(\cdot))) + \frac{1}{2}\left(\frac{1-\rho s}{1-\rho\mu_Y(x^s)}\frac{1-\rho}{1-\rho\mu_Y(x^s)} - 1\right) u'(\mu_1(\cdot)) \\
&> \rho(u(1) - u(\mu_1(\cdot))) + \frac{1}{2}\left(\frac{1-\rho}{1-\rho\mu_Y(x^s)} - 1\right) u'(\mu_1(\cdot)) \\
&= \rho(u(1) - u(\mu_1(\cdot))) - \frac{1}{2}\rho(1-\mu_1(\cdot))u'(\mu_1(\cdot)) \\
&= 0.
\end{aligned}$$

The first equality follows from differentiating and from using that $d\mu_Y(x^s)/ds = 1/2$ for uniformly distributed signals. The first inequality follows from using that strict concavity of $u(\cdot)$ and $\mu_1(\mu_Y(x^s), 1, 1-\rho) < \mu_N(x^s)$ imply that $u'(\mu_N(x^s)) < u'(\mu_1(\mu_Y(x^s), 1, 1-\rho))$. The second inequality follows from using that $s < \mu_Y(x^s)$ implies $(1-\rho s)/(1-\rho\mu_Y(x^s)) > 1$. The second equality follows from using that $(1-\rho)/(1-\rho\mu_Y(x^s)) - 1 = -\rho \cdot (1-\mu_Y(x^s))/(1-\rho\mu_Y(x^s))$ and that $1-\mu_1(\cdot) = (1-\mu_Y(x^s))/(1-\rho\mu_Y(x^s))$. The third equality follows from using that $u(\mu_{\widehat{\sigma}}) = -(1-\mu_{\widehat{\sigma}})^2$ implies that $u(1) - u(\mu_1(\cdot)) = 1/2 \cdot (1-\mu_1(\cdot))u'(\mu_1(\cdot))$. Hence, $d\rho/ds > 0$. This proves the result. q.e.d.

Corollary B 1 *Proposition 1 extends to the case where $v(\cdot)$ includes a cubic term with a strictly positive coefficient.*

Proof to Corollary 1. Fix any participation threshold $s \in \mathcal{S}$. For any given test T' which induces s, ignore the cubic term at first and construct the tests T'' and T''' exactly like in Proposition 1. The tests T' and T''' both induce s. What we have to argue is that the principal also prefers test T''' over test T' when the cubic term is not ignored. Lemma 4 (a) which is responsible for the principal preferring test T''' over test T'' in Proposition 1 relies only on the convexity of $v(\cdot)$ and extends thus directly to the considered cubic case. It remains to argue why the principal prefers test T'' over test T'. Sufficient for this is that $p^1_{\mu_Y}\mu_1^3 + p^2_{\mu_Y}\mu_2(\mu_1)^3$ increases along the transition path from the test T' to the test T''. We have

$$\begin{aligned}
p^1_{\mu_Y}\mu_1^3 + p^2_{\mu_Y}\mu_2(\mu_1)^3 &= \frac{\mu_2(\mu_1)-\mu_Y}{\mu_2(\mu_1)-\mu_1}\mu_1^3 + \frac{\mu_Y-\mu_1}{\mu_2(\mu_1)-\mu_1}\mu_2(\mu_1)^3 \\
&= -\mu_1\mu_2(\mu_1)(\mu_1+\mu_2(\mu_1)) + \mu_Y \cdot (\mu_2(\mu_1)^2 + \mu_2(\mu_1)\mu_1 + \mu_1^2). \quad (13)
\end{aligned}$$

The first equality follows from using that Bayes' Law implies $p^2_{\mu_Y} = (\mu_Y-\mu_1)/(\mu_2(\mu_1)-\mu_1)$. The second equality follows from simplifying. By differentiating (13) with respect to μ_1 and by using that $\mu_2'(\mu_1) = (\mu_2(\mu_1)-\mu_Y)/(\mu_Y-\mu_1)$ by Step 2 in Proposition 1 (recall that we still consider the same transition path in the (μ_1,μ_2)–space as in Proposition 1 even though this path has now a different interpretation), we obtain

$$\begin{aligned}
&(-(\mu_2(\mu_1)-\mu_Y)(2\mu_1+\mu_2(\mu_1))) + \frac{\mu_2(\mu_1)-\mu_Y}{\mu_Y-\mu_1}((\mu_Y-\mu_1)(\mu_1+2\mu_2(\mu_1))) \\
= \ &(\mu_2(\mu_1)-\mu_Y)(\mu_2(\mu_1)-\mu_1).
\end{aligned}$$

As this expression is strictly positive, it follows that the principal's expected utility increases strictly along the transition path. This proves the result. q.e.d.

Proof to Lemma 5. Fix any participation threshold $s \in [\underline{\theta}, \overline{\theta})$ which is inducible by some test. Before we prove the four parts of the result, we prove an auxiliary result which holds also for non–binary tests.

Auxiliary Result 1: Lemma 3 extends to the modified setting. Note that the definitions of T, p^σ_θ and $\mu_\sigma(\cdot)$ are changed relative to the original model. However, the properties which we derived in Step 1 of the proof to Lemma 3 continue to hold as they rely only on general properties of Bayesian updating. In particular, we have $\sum_\sigma p^\sigma_{\mu_Y}\mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_n, p^\sigma_g) = \mu_Y$ and

$$\begin{aligned}
\sum_\sigma p^\sigma_{\mu_Y}\mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_n, p^\sigma_g)^2 &= \sum_\sigma p^\sigma_{\mu_Y}\left(\frac{(1-\gamma)\mu_Y p^\sigma_g + \gamma\mu_Y p^\sigma_n}{p^\sigma_{\mu_Y}}\right)^2 \\
&= \sum_\sigma \mu_Y((1-\gamma)p^\sigma_g + \gamma p^\sigma_n)\mu_\sigma(\mu_Y, p^\sigma_b, p^\sigma_g)
\end{aligned}$$

$$= \mu_Y E_1(T, \mu(x^s, T)).$$

It follows that the principal's expected utility is still described by the formula (7). As the Steps 2 and 3 of the proof to Lemma 3 are not affected by the changes in the definitions, Lemma 3 extends.

We consider in the remainder of this proof binary tests and prove Parts (b) and (c) before Part (a).

(b) Suppose that the quality perception pair (μ_1, μ_2) is supportable by some test. Bayes' Law implies that $p^2_{\mu_Y} = (\mu_Y - \mu_1)/(\mu_2 - \mu_1)$. This implies that $p^2_{\mu_Y}$ is the same for any binary test which induces the quality perception pair (μ_1, μ_2) and that it is as in the original model. It follows that the principal's expected utility depends only through (μ_1, μ_2) on the test design and that it is as in the original model.

(c) Suppose the quality perception pairs (μ_1', μ_2') and (μ_1'', μ_2'') are supportable by some tests, say T' and T'', and suppose the principal is indifferent between (μ_1', μ_2') and (μ_1'', μ_2''). By Auxiliary Result 1, indifference of the principal implies $E_s(T', \mu(x^s, T')) = E_s(T'', \mu(x^s, T'')) =: \overline{\overline{\overline{E}}}_s$. By using the definition of $E_s(T, \mu)$, we obtain that the test result $\sigma = 2$ is generated with probability $(\overline{\overline{\overline{E}}}_s - \mu_1')/(\mu_2' - \mu_1')$ (resp. $(\overline{\overline{\overline{E}}}_s - \mu_1'')/(\mu_2'' - \mu_1'')$) when the threshold agent participates in test T' (resp. T''). This implies that the threshold agent's preferences over tests among which the principal is indifferent depend only through the induced quality perception pair and the constant $\overline{\overline{\overline{E}}}_s$ on the test design. As $\overline{\overline{\overline{E}}}_s < \mu_Y$, the logic of Steps 2 and 3 in the Proof of Proposition 1 applies also to the modified setting. It follows that the threshold agent's expected utility increases as we move to the northeast on an indifference curve of the principal in the (μ_1, μ_2)–space. His preferences over the quality perception pairs on such curves are thus as in the original model.

(a) Suppose the tests T' and T'' induce the same quality perception pair (μ_1, μ_2). We need to argue that both tests imply the same expected utility for the principal and the threshold agent. For the principal, this follows directly from Part (b). For the threshold agent, this follows from Part (c) as the principal is indifferent between tests T' and T'' by Part (b).

(d) *Step 1: If* $(\mu_1, 1)$ *is supportable, it is supportable by a unique test which has the structure* $T^{(g,y)}(\rho)$. Suppose the quality perception pair $(\mu_1, 1)$ is induced by some test $T = ((1 - p_b^2, p_b^2), (1 - p_n^2, p_n^2), (1 - p_g^2, p_g^2))$. As $p_b^2 = p_n^2 = 0$ is necessary for $\mu_2(\mu_Y, p_b^2, p_n^2, p_g^2) = 1$, we have $T = T^{(g,y)}(p_g^2)$. Moreover, as $\mu_1(\mu_Y, 1, 1, 1 - p_g^2) = \mu_1$ has the unique solution $p_g^2 = (\mu_Y - \mu_1)/(1 - \mu_1) \cdot 1/((1 - \gamma)\mu_Y)$, there exists a unique test $T^{(g,y)}(p_g^2)$ which induces $(\mu_1, 1)$.

Step 2: The threshold agent's preferences over tests $T^{(g,y)}(\rho)$. We can proceed analogous to the Steps 1 and 3 in the Proof to Lemma 4 (a). The only difference lies in how (8), (9) and (10) exactly look like for the modified setting. We obtain

$$\mu_1(\mu_Y, 1, 1, 1 - \rho) = \frac{\mu_Y - (1 - \gamma)\mu_Y \rho}{1 - (1 - \gamma)\mu_Y \rho},$$

$$\frac{d}{d\rho}\mu_1(\cdot) = -\frac{(1 - \gamma)\mu_Y}{1 - (1 - \gamma)\mu_Y \rho_g}(1 - \mu_1(\cdot)) \tag{14}$$

and

$$\frac{d}{d\rho}U_s(T^{(g,y)}(\rho), \mu(x^s, T^{(g,y)}(\rho))) < -(1-\gamma)(\mu_Y - s)\frac{1-\mu_1(\cdot)}{1-(1-\gamma)\mu_Y\rho}u'(\mu_1(\cdot)) < 0.$$

Hence, the threshold agent strictly prefers test $T^{(g,y)}(\rho')$ over test $T^{(g,y)}(\rho'')$ if $\rho' < \rho''$.

Step 3: The threshold agent's preferences over quality perception pairs $(\mu_1, 1)$. Because $\mu_1(\mu_Y, 1, 1, 1-\rho)$ is strictly decreasing in ρ by (14), Step 2 implies that the threshold agent strictly prefers $(\mu_1'', 1)$ over $(\mu_1', 1)$ if $\mu_1' < \mu_1''$. That is, the threshold agent's preferences are as in the original model. q.e.d.

Proof to Proposition 4. Fix any participation threshold $s \in [\underline{\theta}, \overline{\theta})$ which is inducible by some test.

(a) Suppose $U_s(T^{(g,y)}(1), \mu(x^s, T^{(g,y)}(1))) \leq u(\mu_N)$.

Step 1: There exists $\rho_g \in (0,1]$ such that (3) holds with equality for $T = T^{(g,y)}(\rho_g)$. By the supposition, the threshold agent has at least a weak incentive not to participate in the test $T^{(g,y)}(1)$. As $\lim_{\rho\to 0} U_s(T^{(g,y)}(\rho), \mu(x^s, T^{(g,y)}(\rho))) = u(\mu_Y)$ and as $u(\mu_Y) > u(\mu_N)$, the threshold agent has a strict incentive to participate in any test $T^{(g,y)}(\rho)$ with a sufficiently low accuracy ρ. By a continuity property and an Intermediate Value Theorem, there exists $\rho \in (0,1]$ such that $U_s(T^{(g,y)}(\rho), \mu(x^s, T^{(g,y)}(\rho))) = u(\mu_N)$.

Step 2: The optimal binary test. Lemma 5 and Step 1 allow us to proceed like in the Proof of Proposition 1: By Lemma 5 (a) there exists also for the modified model a mapping $T^\mu(\mu_1, \mu_2)$ which relates quality perception pairs to tests such that choosing (μ_1, μ_2) to maximize the objective function in (2) subject to (3) with $T = T^\mu(\mu_1, \mu_2)$ is equivalent to choosing a test $T \in \mathcal{T}^2$ directly. There are however two difference. First, not any quality perception pair $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$ is supportable by a binary test. That is, the domain of $T^\mu(\cdot)$ is smaller than in the original model. However, the principal's preferences over those quality perception pairs which are supportable are by Lemma 5 (b) as in the original model. Second, the threshold agent's indifference curves in the (μ_1, μ_2)–space might differ from those in the original model. However, Lemma 5 (c) and (d) imply that the properties of the threshold agent's preferences which we actually use in Proposition 1, are as in the original model. This allows us to proceed as follows: Ignore at first the supportability problem. That is, suppose Lemma 5 (b), (c) and (d) hold for all $(\mu_1, \mu_2) \in [0, \mu_Y) \times (\mu_Y, 1]$. Use the logic of Proposition 1 to derive a candidate for the optimal quality perception pair. If the so constructed quality perception pair is supportable, it must be optimal. By Step 1, this is indeed the case under the supposition that we imposed for Part (a). Moreover, Step 1 implies that the optimal quality perception pair is uniquely induced by a test $T^{(g,y)}(\rho_g^*)$ where ρ_g^* is the unique solution to $U_s(T^{(g,y)}(\rho_g^*), \mu(x^s, T^{(g,y)}(\rho_g^*))) = u(\mu_N)$.

Step 3: The optimal test. For the original model, Proposition 2 establishes that the optimal test is binary. The proof makes only use of properties of binary test design problems. Hence, like in Step 2, Lemma 5 (b), (c) and (d) allow us to apply the same logic for the modified model. It follows that the optimal test is the optimal binary test.

Before we proceed with Part (b), we prove an auxiliary result.

Auxiliary Result 2: An adapted version of Lemma 4 (a) extends to the modified setting: $V(T, x^s, \mu(x^s, T))$ strictly increases and $U_s(T, \mu(x^s, T))$ strictly decreases (i) when ρ_g increases *for tests $T = T^{(g,y)}(\rho_n, \rho_g)$ and (ii) when ρ_n increases for tests $T^{(g,y)}(\rho_n, 1)$.* The proofs are analogous to the proof of Lemma 4 (a). The only difference lies in how the expressions (8), (9) and (10) exactly look like and that in case (ii) it is μ_2 which is affected by the change in the parameter instead of μ_1. *Case (i).* For tests $T^{(g,y)}(\rho_n, \rho_g)$ we obtain $\mu_2(\cdot) = ((1-\gamma)\mu_Y(1-\rho_g) + \gamma\mu_Y\rho_n)/((1-\gamma)\mu_Y(1-\rho_g)+\gamma\rho_n)$, $d\mu_2(\cdot)/d\rho_g = -(1-\gamma)\mu_Y(1-\mu_2(\cdot))/((1-\gamma)\mu_Y(1-\rho_g)+\gamma\rho_n)$ and $dU_\theta(T^{(g,y)}(\rho_n, \rho_g), \mu(x^s, T^{(g,y)}(\rho_n, \rho_g)))/d\rho_g < -\gamma(1-\gamma)\rho_n(\mu_Y - \theta)(1-\mu_2(\cdot))/((1-\gamma)\mu_Y(1-\rho_g) + \gamma\rho_n)u'(\mu_2(\cdot))$. *Case (ii).* For tests $T^{(g,y)}(\rho_n, 1)$ we obtain $\mu_1(\cdot) = \gamma(1-\rho_n)\mu_Y/(\gamma(1-\rho_n) + (1-\gamma)(1-\mu_Y))$, $d\mu_1(\cdot)/d\rho_n = -\gamma/(\gamma(1-\rho_n) + (1-\gamma)(1-\mu_Y))(\mu_Y - \mu_1(\cdot))$ and $dU_\theta(T^{(g,y)}(\rho_n, 1), \mu(x^s, T^{(g,y)}(\rho_n, 1)))/d\rho_n < -\gamma(1-\gamma)(\mu_Y-\theta)(\mu_Y - \mu_1(\cdot))/(\gamma(1-\rho_n) + (1-\gamma)(1-\mu_Y))u'(\mu_1(\cdot))$.

(b) Suppose $U_s(T^{(g,y)}(1), \mu(x^s, T^{(g,y)}(1))) > u(\mu_N)$. We show now that a given test $T' \in \mathcal{T}$ can only solve (2) subject to (3) if there exist ρ_n and ρ_g such that $T' = T^{(g,y)}(\rho_n, \rho_g)$.

Step 1: Construction of a test T'' which implies at most one quality perception in $[0, \mu_Y)$ and which is better for the threshold agent and equally good for the principal than the test T'. Denote the test T' by (p_b, p_n, p_g). Let $\mu' = \mu(x^s, T')$. Suppose without loss of generality that any test result induces a different quality perception. If $\nexists \sigma', \sigma''$ such that $\mu'_{\sigma'} < \mu'_{\sigma''} < \mu_Y$, set $T'' = T'$ and proceed with Step 2. Otherwise, proceed with the following Substeps 1.1 and 1.2.

Substep 1.1: Construction of a test $T''_m(\alpha^*)$ with nice properties. Define a family of tests $T''(\alpha)$ with $\alpha \in [0, 1]$ which differ from test T' only in the probabilities which generate the test result $\sigma \in \{\sigma', \sigma''\}$ when $\widehat{\omega} \in \{g, n, b\}$, say $p^\sigma_{\widehat{\omega}}(\alpha)$. Let $p^\sigma_{\widehat{\omega}}(\alpha)$ as described in the following table:

	$\sigma = \sigma'$	$\sigma = \sigma''$
$\widehat{\omega} = g$	$\alpha(p_g^{\sigma'} + p_g^{\sigma''})$	$(1-\alpha)(p_g^{\sigma'} + p_g^{\sigma''})$
$\widehat{\omega} = n$	$\alpha(p_n^{\sigma'} + p_n^{\sigma''})$	$(1-\alpha)(p_n^{\sigma'} + p_n^{\sigma''})$
$\widehat{\omega} = b$	$(p_b^{\sigma'} + p_b^{\sigma''}) - (1-\alpha)(p_g^{\sigma'} + p_g^{\sigma''})$	$(1-\alpha)(p_g^{\sigma'} + p_g^{\sigma''})$

Note that for $\alpha = 1$ the initial test results σ' and σ'' are pooled on the new test result σ' and that there is no new test result σ''. Define $\mu_\sigma(\alpha) := \mu_\sigma(\mu_Y, p^\sigma_b(\alpha), p^\sigma_n(\alpha), p^\sigma_g(\alpha))$ for $\sigma \in \{\sigma', \sigma''\}$.

We make three observations: First, $T''(\alpha)$ specifies indeed a test. The only non–trivial part of this observation is that $p_b^{\sigma'}(\alpha) \geq 0$. This follows from $p_g^{\sigma'} < p_b^{\sigma'}$ and $p_g^{\sigma''} < p_b^{\sigma''}$ which are necessary conditions for $\mu'_{\sigma'} < \mu_Y$ and for $\mu'_{\sigma''} < \mu_Y$, respectively. Second, by construction, $\mu_{\sigma''}(\alpha) = \mu_Y$ for any $\alpha \in [0, 1)$. Moreover, $0 = \mu_{\sigma'}(0) \leq \mu'_{\sigma'} < \mu_{\sigma'}(1) < \mu_Y$. Third, the quality perception distribution induced by $T''(0)$ is a mean–preserving spread of the quality perception distribution induced by T' which is in turn a mean–preserving spread of the quality perception distribution induced by $T''(1)$. As the principal evaluates quality perceptions at a strictly convex function, she strictly prefers test $T''(0)$ over test T' and test T' over test $T''(1)$. By a continuity property and an Intermediate Value Theorem, there exists $\alpha^* \in (0, 1)$ such that the principal is indifferent between tests T' and $T''(\alpha^*)$. Note that $T''(\alpha^*)$ implies a quality perception distribution where the quality perceptions implied by test results σ' and σ'' may coincide with the quality perceptions implied by other test results. Construct thus a test $T''_m(\alpha^*)$ from test $T''(\alpha^*)$ by merging test results which imply the same quality perceptions.

Substep 1.2: The induction step. By Lemma 5 (c) and a reasoning like in Step 3 in the Proof of Proposition 2, the threshold agent prefers the test $T''_m(\alpha^*)$ constructed in Substep 1.1 over the test T'. By construction, the principal is indifferent between T' and $T''_m(\alpha^*)$. Moreover, $T''_m(\alpha^*)$ implies a quality perception distribution which has at least one quality perception less which lies in the interval $[0, \mu_Y)$. By applying Substep 1.1 repeatedly with the resulting test as the starting test in the next round, we obtain a test T'' which induces at most one quality perception in $[0, \mu_Y)$.

Step 2: Construction of a test T''' which implies at most one quality perception in $[0, \mu_Y)$ and in $[\mu_Y, 1)$, respectively, and which is better for the threshold agent and equally good for the principal than the test T''. Denote the test T'' by (p_b, p_n, p_g). Let $\mu'' = \mu(x^s, T'')$. If $\nexists \sigma', \sigma''$ such that $\mu_Y \leq \mu''_{\sigma'} < \mu''_{\sigma''} < 1$, set $T''' = T''$ and proceed with Step 3. Otherwise, proceed with the following Substeps 2.1 and 2.2.

Substep 2.1: Construction of a test $T'''_m(\alpha^*)$ with nice properties. Define a family of tests $T'''(\alpha)$ with $\alpha \in [0, 1]$ which differ from test T'' only in the probabilities which generate the test result $\sigma \in \{\sigma', \sigma''\}$ when $\widehat{\omega} \in \{g, n, b\}$, say $p^\sigma_{\widehat{\omega}}(\alpha)$. Let $p^\sigma_{\widehat{\omega}}(\alpha)$ as described in the following table:

	$\sigma = \sigma'$	$\sigma = \sigma''$
$\widehat{\omega} = g$	$(1-\alpha)(p^{\sigma'}_b + p^{\sigma''}_b) + \alpha(p^{\sigma'}_g + p^{\sigma''}_g)$	$(1-\alpha)(p^{\sigma'}_g + p^{\sigma''}_g) - (1-\alpha)(p^{\sigma'}_b + p^{\sigma''}_b)$
$\widehat{\omega} = n$	$p^{\sigma'}_n + p^{\sigma''}_n$	0
$\widehat{\omega} = b$	$p^{\sigma'}_b + p^{\sigma''}_b$	0

Note that for $\alpha = 1$ the initial test results σ' and σ'' are pooled on the new test result σ' and that there is no new test result σ''. Define $\mu_\sigma(\alpha) := \mu_\sigma(\mu_Y, p^\sigma_b(\alpha), p^\sigma_n(\alpha), p^\sigma_g(\alpha))$ for $\sigma \in \{\sigma', \sigma''\}$.

We make three observations: First, $T'''(\alpha)$ specifies indeed a test. The only non–trivial part of this observation is that $p^{\sigma''}_g(\alpha) \geq 0$. This follows from $p^{\sigma'}_g \geq p^{\sigma'}_b$ and $p^{\sigma''}_g > p^{\sigma''}_b$ which are necessary conditions for $\mu_Y \leq \mu''_{\sigma'}$ and for $\mu_Y < \mu''_{\sigma''}$, respectively. Second, by construction, $\mu_{\sigma''}(\alpha) = 1$ for any $\alpha \in [0, 1)$. Moreover, $\mu_Y = \mu_{\sigma'}(0) \leq \mu''_{\sigma'} < \mu_{\sigma'}(1) < 1$. Third, the quality perception distribution induced by $T'''(0)$ is a mean–preserving spread of the quality perception distribution induced by T'' which is in turn a mean–preserving spread of the quality perception distribution induced by test $T'''(1)$. As the principal evaluates quality perceptions at a strictly convex function, she strictly prefers test $T'''(0)$ over test T'' and test T'' over test $T'''(1)$. By a continuity property and an Intermediate Value Theorem, there exists $\alpha^* \in (0, 1)$ such that the principal is indifferent between tests T'' and $T'''(\alpha^*)$. Note that $T'''(\alpha^*)$ implies a quality perception distribution where the quality perceptions implied by test results σ' and σ'' may coincide with the quality perceptions implied by other test results. Construct thus a test $T'''_m(\alpha^*)$ from test $T''(\alpha^*)$ by merging test results which imply the same quality perceptions.

Substep 2.2: The induction step. By Lemma 5 (c) and a reasoning like in Step 3 in the Proof of Proposition 2, the threshold agent prefers the test $T'''_m(\alpha^*)$ constructed in Substep 2.1 over the initial test T''. By construction, the principal is indifferent between T'' and $T'''_m(\alpha^*)$. Moreover, $T'''_m(\alpha^*)$ implies a quality perception distribution which has at least one quality perception less which lies in the interval $[\mu_Y, 1)$ and which has equally many quality perceptions which lie in the interval $[0, \mu_Y)$. By applying Substep 2.2 repeatedly with the resulting test as the starting test in the next round, we obtain a test T''' which implies at most one quality perception in $[0, \mu_Y)$ and in $[\mu_Y, 1)$, respectively.

Step 3: Construction of a test $T'''' = T^{(g,y)}(\rho_n, \rho_g)$ or a test $T'''' = T^{(g,y)}(\rho_g)$ which is better for the threshold agent and equally good for the principal than the test T'''. Denote T''' by (p_b, p_n, p_g). Let $\mu''' = \mu(x^s, T''')$. Steps 1 and 2 imply that μ''' contains at most one quality perception in $[0, \mu_Y)$, in $[\mu_Y, 1)$ and in $\{0\}$, respectively. Moreover, because updating is Bayesian, there exists at least one quality perception in $[0, \mu_Y)$ and in $(\mu_Y, 1]$, respectively. It follows that the test T''' has the structure described in the following table

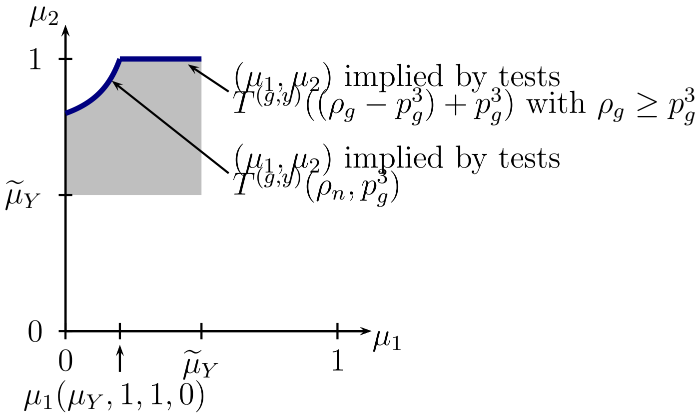


Figure 7: Quality perception pairs supportable by tests T''' with p_g^3 fix [$\widetilde{\mu}_Y = 1/2$, $\gamma = 1/4$]

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
$\widehat{\omega} = g$	p_g^1	p_g^2	p_g^3
$\widehat{\omega} = n$	p_n^1	p_n^2	0
$\widehat{\omega} = b$	p_b^1	p_b^2	0
μ_σ'''	$\in [0, \mu_Y)$	$\in [\mu_Y, 1)$	$= 1$

with $p_g^1 + p_n^1 + p_b^1 > 0$ and with $p_g^2 + p_n^2 + p_b^2 + p_g^3 > 0$. When either $p_g^2 + p_n^2 + p_b^2 = 0$ or $p_g^3 = 0$, the test generates only two test results. Note that in these cases we slightly abuse notation as we required in Section 2 that any test result occurs with positive probability. If $p_g^2 + p_n^2 + p_b^2 = 0$, $T''' = T^{(g,y)}(p_g^3)$. Set then $T'''' = T'''$ and proceed with Step 4. It remains to consider $p_g^2 + p_n^2 + p_b^2 > 0$. Consider tests which differ from test T''' only in the probabilities which generate the first two test results. As a consequence, only the quality perceptions μ_1 and μ_2 which are associated to the test results $\sigma = 1$ and $\sigma = 2$ are affected by the test design. The set of supportable quality perception pairs (μ_1, μ_2) is a subset of $[0, \widetilde{\mu}_Y) \times (\widetilde{\mu}_Y, 1]$ with $\widetilde{\mu}_Y = \mu_\sigma(\mu_Y, 1, 1, 1 - p_g^3)$.

We make four observations: First, for any $\mu_1 \in [0, \mu_1(\mu_Y, 1, 1, 0))$ the highest μ_2 such that (μ_1, μ_2) is supportable is implied by a test $T^{(g,y)}(\rho_n, p_g^3)$ with $\rho_n \in (0, 1]$. Second, for any $\mu_1 \in [\mu_1(\mu_Y, 1, 1, 0), \widetilde{\mu}_Y]$ the highest μ_2 such that (μ_1, μ_2) is supportable is implied by a test which is as described in the following table

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
$\widehat{\omega} = g$	$1 - \rho_g$	$\rho_g - p_g^3$	p_g^3
$\widehat{\omega} = n$	1	0	0
$\widehat{\omega} = b$	1	0	0
μ_σ'''	$\in [0, \mu_Y)$	$= 1$	$= 1$

with $\rho_g \geq p_g^3$. This test is equivalent to the test $T^{(g,y)}((\rho_g - p_g^3) + p_g^3)$ with $\rho_g \geq p_g^3$. Third, the upper supportability frontier (which we described in the first two observations and which is illustrated by the blue curve in Figure 7) is connected. Fourth, consider the principal's preferences over quality perception pairs $(\mu_1, \mu_2) \in [0, \widetilde{\mu}_Y) \times (\widetilde{\mu}_Y, 1]$ when the supportability problem is ignored. Then the principal's indifference curve through (μ_1''', μ_2''') is continuous, strictly increasing and contains a point $(\mu_1, 1)$ with $\mu_1 \geq \mu_1'''$.

The argument is now the following: (μ_1''', μ_2''') lies either on or below the upper supportability frontier. By the fourth observation, the principal's indifference curve through (μ_1''', μ_2''') contains a point $(\mu_1, 1)$ with $\mu_1 \geq \mu_1'''$ which lies either on or above the upper supportability frontier. By the third observation, there exists at least one point of intersection of the principal's indifference curve and the upper supportability frontier which lies in the northeast of (μ_1''', μ_2'''). By the first two observations, each point of intersection is implied by either a test $T^{(g,y)}(\rho_n, \rho_g)$ or a test $T^{(g,y)}((\rho_g - p_g^3) + p_g^3)$. Fix one such point of intersection and denote the associated test by T''''. By construction, the principal is indifferent between tests T''' and T'''', whereas the threshold agent prefers by a reasoning analogous to that in Step 3 of the proof to Proposition 2 the test T''''.

Step 4: Construction of a test $T^{(g,y)}(\rho_n^*, \rho_g^*)$ which induces s and which is better for the principal than the test T''''. We distinguish two cases.

Case 1. Suppose there exist ρ_n and ρ_g such that $T'''' = T^{(g,y)}(\rho_n, \rho_g)$. Note that we can continuously transform the test $T'''' = T^{(g,y)}(\rho_n, \rho_g)$ into the test $T^{(g,y)}(1,1)$ by first increasing ρ_g until the test $T^{(g,y)}(\rho_n, 1)$ is reached and by then increasing ρ_n. By Auxiliary Result 2, the threshold agent's expected utility decreases strictly along this transformation. By construction, the threshold agent prefers the test $T'''' = T^{(g,y)}(\rho_n, \rho_g)$ over the test T'. Moreover, it can easily be shown that the threshold agent prefers any test which induces some pooling, e.g. the test T', over the test $T^{(g,y)}(1,1)$ which perfectly reveals $\widehat{\omega}$. As $U_s(T^{(g,y)}(\rho_n, \rho_g), \mu(x^s, T^{(g,y)}(\rho_n, \rho_g)))$ is continuous in ρ_n and in ρ_g, we obtain by an Intermediate Value Theorem that there exists a test $T^{(g,y)}(\rho_n^*, \rho_g^*)$ which induces s. Because the principal is by construction indifferent between the tests T' and T'''' and because the principal's expected utility is by Auxiliary Result 2 increasing along the transformation, the principal prefers the test $T^{(g,y)}(\rho_n^*, \rho_g^*)$ over the test T'.

Case 2. Suppose there exists ρ_g such that $T'''' = T^{(g,y)}(\rho_g)$. By Lemma 5 (d), the principal's expected utility increases strictly and the threshold agent's expected utility decreases strictly in ρ_g. By the supposition of Part (b), the threshold agent has for $\rho_g = 1$ still a strict incentive to participate in the test. As $T^{(g,y)}(1) = T^{(g,y)}(0,1)$, we can apply Case 1 to obtain a test $T^{(g,y)}(\rho_n^*, \rho_g^*)$ with the desired properties. q.e.d.

References

Alós-Ferrer, C. and Prat, J. (2012). Job market signaling and employer learning. *Journal of Economic Theory*, 147:1787–1817.

Arrow, K. J. (1973). Higer education as a filter. *Journal of Public Economics*, 2:193–216.

Bar, T., Kadiyali, V., and Zussman, A. (2012). Putting grades in context. *Journal of Labor Economics*, 30:445–478.

Benoît, J.-P. and Dubra, J. (2004). Why do good cops defend bad cops? *International Economic Review*, 45:787–809.

Calzolari, G. and Pavan, A. (2006a). Monopoly with resale. *The RAND Journal of Economics*, 37:362–375.

Calzolari, G. and Pavan, A. (2006b). On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, 130:168–204.

Caplin, A. and Eliaz, K. (2003). Aids policy and psychology: A mechanism–design approach. *The RAND Journal of Economics*, 34:631–646.

Daley, B. and Green, B. (2014). Market signaling with grades. *Journal of Economic Theory*, 151:114–145.

De, S. and Nabar, P. (1991). Economic implications of imperfect quality certification. *Economics Letters*, 37:333–337.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification. *Journal of Economic Literature*, 48:935–963.

Eső, P. and Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies*, 74(3):705–731.

Farhi, E., Lerner, J., and Tirole, J. (2013). Fear of rejection? tiered certification and transparency. *The RAND Journal of Economics*, 44(4):610–631.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1:60–79.

Gill, D. and Sgroi, D. (2008). Sequential decisions with tests. *Games and Economic Behavior*, 63:663–678.

Gill, D. and Sgroi, D. (2012). The optimal choice of pre–launch reviewer. *Journal of Economic Theory*, 147:1247–1260.

Goldstein, I. and Leitner, Y. (2013). Stress tests and information disclosure. Working Paper No. 13-26, Federal Reserve Bank of Philadelphia.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24:461–483.

Grossman, S. J. and Hart, O. D. (1980). Disclosure laws and takeover bids. *The Journal of Finance*, 35(2):323–334.

Harbaugh, R. and Rasmusen, E. (2013). Coarse grades: Informing the public by withholding information. Working Paper, April 3, 2013.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61:561–574.

Hull, H. F., Bettinger, C. J., Gallaher, M. M., Keller, N. M., Wilson, J., and Mertz, G. J. (1988). Comparison of hiv-antibody prevalence in patients consenting to and declining hiv-antibody testing in an std clinic. *JAMA: the Journal of the American Medical Association*, 260:935–938.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101:2590–2615.

Li, H. and Li, W. (2013). Misinformation. *International Economic Review*, 54:253–277.

Lizzeri, A. (1999). Information revelation and certification intermediaries. *RAND Journal of Economics*, 30:214–231.

Lyter, D. W., Valdiserri, R. O., Kingsley, L. A., Amoroso, W. P., and Rinaldo, C. R. (1987). The hiv antibody test: Why gay and bisexual men want or do not want to know their results. *Public Health Reports*, 102:468–474.

Matthews, S. and Postlewaite, A. (1985). Quality testing and disclosure. *The RAND Journal of Economics*, 16:328–340.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391.

Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122.

Okuno-Fujiwara, M., Postlewaite, A., and Suzumura, K. (1990). Strategic information revelation. *The Review of Economic Studies*, 57(1):25–47.

Ostrovsky, M. and Schwarz, M. (2010). Information disclosure and unraveling in matching markets. *American Economic Journal: Microeconomics*, 2:34–63.

Ottaviani, M. and Prat, A. (2001). The value of public information in monopoly. *Econometrica*, 69:1673–1683.

Pancs, R. (2014). Designing order–book transparency in electronic networks. *Journal of the European Economic Association*, 12(3):702–723.

Perez-Richet, E. and Prady, D. (2012). Complicating to persuade? Working Paper, February 2012.

Peyrach, E. and Quesada, L. (2004). Strategic certification. Working Paper, May 2004.

Philippon, T. and Skreta, V. (2012). Optimal interventions in markets with adverse selection. *Amer-*

ican Economic Review, 102:1–30.

Rayo, L. and Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy*, 118:949–987.

Rosar, F. and Schulte, E. (2012). Optimal test design under imperfect private information and voluntary participation. Working Paper, April 10, 2012.

Schweizer, N. and Szech, N. (2013). Optimal revelation of life–changing information. Working Paper, January 2013.

Spence, M. (1973). Job market signaling. *The Quaterly Journal of Economics*, 87(3):355–374.

Stiglitz, J. E. (1975). The theory of "screening," education and the distribution of income. *American Economic Review*, 65(3):283–300.

Tirole, J. (2012). Overcoming adverse selection: How public intervention can restore market functioning. *American Economic Review*, 102:29–59.

Titman, S. and Trueman, B. (1986). Information quality and the valuation of new issues. *Journal of Accounting and Economics*, 8:159–172.

Wang, Y. (2012). Bayesian persuasion with multiple receivers. Working Paper, Novermber 15, 2012.