Test design under voluntary participation

(This version: April 27, 2016)

Frank Rosar[a]

[a]*Department of Economics, University of Bonn, Lennéstr. 37, 53113 Bonn, Germany. Tel.: + 49 228 73 6192. Fax: + 49 228 73 7940. E-mail: email@frankrosar.de.*

Abstract

An agent who is imperfectly informed about his binary quality can voluntarily participate in a test that generates a public signal. I study the design of the test that allows for optimal learning of the agent's quality when the agent strives for a high perception of his quality but is averse towards perception risk. For a large class of reduced-form utility functions that reflect these properties, the optimal test is binary and not subject to false positives. I uncover the forces that drive this result and develop a method to derive it. Furthermore, for a non-reduced version of my model where the designer chooses a probabilistic estimate of the agent's quality but suffers either more from false positives or from false negatives, I show that the same type of test is optimal.

JEL classification: D02, D82, D83

Keywords: test design; Bayesian learning; false positive; false negative; asymmetric information; voluntary participation

1. Introduction

Technological advances have in many areas improved the capability to reveal information about an agent through testing procedures. In contrast to signaling and screening mechanisms that were subject of intense study in the past, such procedures are in many environments capable of revealing information that goes beyond what the agent knows himself. However, either because a regulating agency saw a need to protect the agent or for some technical reason, testing often requires the agreement of the agent. This article studies the design of the testing procedure in such environments.

To fix ideas, think of the following examples: A genetic test can determine whether a patient will develop a certain illness, but testing necessitates for legal reasons the formal agreement of the patient. Big data analysis can be used to assess whether a bank is viable in a scenario that is unknown to the bank, but the bank must grant access to the necessary data. Polygraphs and fMRI scanners can be used to judge about the eligibility of a potential employee, but using them requires the candidate's agreement. Similar problems arise also in more classical environments: The judge in an inquisitorial legal system controls the generation of information, but she requires an imperfectly informed prosecutor to bring cases to court.[1] Product testing can reveal information about the safety of a product, but the imperfectly informed producer has to apply for certification.[2]

In this kind of applications, the agent's quality is either good (i.e., he is healthy/viable/eligible/...) or bad (i.e., he is ill/nonviable/ineligible/...), but he is often only imperfectly informed about his true quality. A test is any device that, if used, issues a public signal that depends directly on the agent's true quality.[3] An accurate test perfectly reveals the agent's quality. It generates a *positive* result if the agent is good and a *negative* result if he is bad.[4] Any inaccurate test is subject to false positives, false negatives, or both. After the test is designed by the principal, the agent decides upon participation. The principal uses the generated test result or the agent's decision not to participate to draw a Bayesian inference about the agent's quality. I consider first a version of the model with stylized reduced-form utility specifications that depend directly on this inference. Then I discuss a non-reduced version and a class of utility functions with standard properties that is interesting from an applied perspective.

In general, the principal strives for uncovering information about the agent's quality. She dislikes false positives and false negatives. However, depending on the application, she may dislike one type of error more than the other. The agent cares about how his quality is perceived

[1] Kamenica and Gentzkow (2011) discuss in their leading example the complementary problem where information generation is controlled by an uninformed prosecutor.

[2] See Harbaugh and Rasmusen (2014) for a deeper discussion of the certification application.

[3] I am interested in the case where the generated information is always learned by the principal. Matthews and Postlewaite (1985) compare an agent's incentive to generate information through an accurate test for voluntary and mandatory disclosure of the generated information. See also Farhi et al. (2013).

[4] Note that I fix the wording such that "positive" refers to the result that is better for the agent. For some applications, the wording is interchanged. E.g., a positive HIV test is the result that is worse for the agent.

by the principal. He benefits when the perception is higher but he is in many problems averse towards perception risk. For instance, such an aversion may be due to the Hirshleifer effect or it may arise for psychological reasons.[5]

What makes the test design problem interesting is that there is no full unraveling under an accurate test. The agent faces a trade-off: non-participation signals unfavorable private information about his quality but participation comes along with a perception risk. The usual unraveling argument fails as the imperfection of the agent's private information limits how adverse the inference associated to non-participation can be. Because of his aversion towards perception risk, the agent with the worst possible private information strictly prefers to be perceived as having the worst possible private information to his quality being perfectly revealed by an accurate test. The relevance of this kind of effect was already observed by Stiglitz (1975):

> "If individuals are very risk averse and not perfectly certain of their abilities, then they may prefer to be treated simply as average rather than to undertake the chance of being screened and labelled below average."

Thus, even if there are no technical constraints that limit the test accuracy and accuracy is costless, the principal may have an incentive to choose an inaccurate test to foster participation.

This gives rise to a number of questions: Is optimal learning achieved through an accurate test? If not, what is it optimal to test for? Is the optimal test subject to false positives, false negatives or both? How is this affected by whether the principal suffers more from false positives or from false negatives?

My paper has three main contributions. The first two are theoretical. First, I use the version of my model with reduced-form utility specifications to uncover the forces that drive the optimal test design. For a large class of reduced-form utility functions that reflect the interest in learning on the principal's side and perception risk-aversion on the agent's side, I find that a single test structure is optimal: the optimal test is binary and not subject to false positives. Second, to derive this optimal test structure, I develop a method to solve binary belief design problems with a participation constraint. I show how such problems can be used to solve my non-binary test design problem with voluntary participation of an imperfectly informed agent.

The third contribution is more applied. I discuss the non-reduced version of my model where the principal has to give a probabilistic estimate of the agent's quality but suffers either more from a high estimate when it turns out that the agent is bad or from a low estimate when it turns out that he is good. That is, either false positives or false negative are more costly for the principal. Expectedly, tests that are not subject to false positives perform well when the principal suffers more from false positives; more surprisingly, such tests turn out to be also optimal when the principal suffers more from false negatives.

[5] According to Hirshleifer (1971), such an aversion can arise because the revelation of public information takes away hedging opportunities. Moreover, in medical contexts, learning aversion arises often for psychological reasons. An indicator for such an aversion is that participation rates in medical tests are often low even when testing is costless. See Lyter et al. (1987) and Hull et al. (1988).

2. Literature

The classic literature on education observed already that signaling and testing are two important channels of learning. Spence (1973) focuses on the signaling role of education. Arrow (1973) and Stiglitz (1975) point to the role of schooling as an information generating device and a costly signal at the same time. Weiss (1983) and Alós-Ferrer and Prat (2012) take an information generation technology as given and study the incentives to engage in costly signaling prior to information generation. My article is concerned with the design of an information generation technology that optimally exploits signaling incentives.

My article contributes mainly to the literature on information design in sender-receiver games. I study a problem where the information structure is controlled by the receiver and a constraint derives from voluntary participation of the privately informed sender.

The literature on Bayesian persuasion studies the design of the information structure by an uninformed sender. In a seminal contribution, Kamenica and Gentzkow (2011) study this problem for a general class of environments.[6] Bayes' Law restricts the expected value of the posterior belief but the design of the information structure is not subject to any further constraints. As the concavification approach is feasible for this problem, it has a simple solution that can be extended into various directions.[7] My analysis requires different techniques that can cope with an additional constraint that derives from the designer's inability to unilaterally impose an information structure.

Some articles investigate persuasion problems where the sender is already informed at the design stage (e.g., Perez-Richet, 2014; Gill and Sgroi, 2012; Li and Li, 2013). The information structure is then jointly determined by the sender's information generation technology choice and the signaling effect that is associated with it.[8] A related effect arises also in my article. The information structure is jointly determined by the information generation technology and the participation behavior that it implies.

In problems where the information structure is controlled by the receiver, participation of a privately informed sender is often an issue. Most closely related to my article is independent work by Harbaugh and Rasmusen (2014). Like me, they are interested in information generation when a constraint derives from voluntary participation. In contrast to me, they model the sender differently such that different incentive problems arise: he is perfectly informed about

[6] See also Rayo and Segal (2010) and Kolotilin (2015). Persuasion problems where the sender engages in specific kinds of market interaction after the disclosure of information are studied by Ostrovsky and Schwarz (2010) (matching markets) and Goldstein and Leitner (2015) (banking).

[7] This includes the disclosure of information by many senders (Gentzkow and Kamenica, 2016), the interpretation of information by many receivers (Wang, 2013; Alonso and Câmara, 2015; Taneva, 2016), heterogeneity in priors (Alonso and Câmara, 2016), and costly information structures (Gentzkow and Kamenica, 2014).

[8] For a (potentially) unconstrained and a constrained interim design problem where the sender is perfectly informed about his binary type, Perez-Richet (2014) and Gill and Sgroi (2012) find that only pooling equilibria can arise. By contrast, signaling can occur in Li and Li (2013) where the sender faces a constrained interim design problem and has imperfect binary information. Other constrained interim design problems are studied in Titman and Trueman (1986), Gill and Sgroi (2008) and in Section 4 of Daley and Green (2014).

his continuous quality, risk-neutral with respect to how his quality is perceived, and he has to bear an exogenously given fixed cost when he participates. Manipulating how a participating sender's quality is perceived on average is the only instrument for setting participation incentives. In my article, risk-aversion makes participation endogenously costly and there are different ways to induce participation. Perception risk considerations, which are mute in their article, are the central theme in mine. In a less related article, Perez-Richet and Prady (2012) study similar channels of learning as I do. The sender, who is perfectly informed about his binary quality, chooses the cost of test accuracy ("complexity") and the receiver chooses the test accuracy ("the level of understanding"). In contrast to my article, signaling occurs before the test design and tests are restricted to a specific one-dimensional class.

Schweizer and Szech (2014) and Caplin and Eliaz (2003) study test design problems in medical contexts with anticipatory costs of learning. In Schweizer and Szech (2014), a doctor designs a test to maximize the expected utility of his patient. Since participation is no issue, the concavification approach does apply. For a reasonable class of preferences, an inaccurate binary test that is not subject to false positives is optimal. Caplin and Eliaz (2003) study the design of a pass-fail test that is capable of stopping the spread of HIV when participation is voluntary and agents condition their sexual matching behavior on the generated information. Unfavorable test results can be hidden and agents possess no private information. Setting participation incentives may 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 same test structure that is optimal in my article is for different reasons optimal in these articles.

Benoît and Dubra (2004) discuss the preferences of a sender with the median signal about two different imperfectly informative signals. They argue that he may prefer the signal that generates less information even though he would prefer his information to be perfectly revealed. There is no signaling effect and no test design in this paper, but it shares with me the effect that a party that likes information generation prefers a tests that is less accurate as possible.

Some strands of literature are related to my article but differ more strongly. One branch of the test design literature studies the incentive effects of tests on effort provision (e.g., Dubey and Geanakoplos, 2010). The literature on certification studies the design of the information structure by a profit maximizing intermediary. This design goal implies very different effects. For instance, in the seminal contribution by Lizzeri (1999), the certifier can capture the entire informational surplus in the market despite revealing no information. 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. Milgrom and Weber (1982) and Ottaviani and Prat (2001) identify environments in that some player has an interest in the public revelation of information. I take such an interest as given and study the problem where the information structure cannot be unilaterally imposed by such a player.

	$\sigma = 1$	$\sigma = 2$	$\dots$	$\sigma = Z$
$\omega = g$	p_1^g	p_2^g	$\dots$	p_Z^g
$\omega = b$	p_1^b	p_2^b	$\dots$	p_Z^b

Table 1: Tabular description of a test $T = (p^b, p^g)$

3. The reduced model

There is a principal (she) and an agent (he). The agent's quality is either good ($\omega = g$) or bad ($\omega = b$) but the agent is only imperfectly informed about his quality. He is good with probability θ and bad with probability $1 - \theta$; the agent knows the value of θ whereas the principal knows only its distribution. 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 my 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.[9] The prior probability with that the agent is good is $\theta_0 \equiv \mathbb{E}_\theta[\theta]$.

A test $T = (p^b, p^g)$ is characterized by a number Z of possible test results and a conditional probability vector $p^\omega \equiv (p_1^\omega, \dots, p_Z^\omega)$ for each $\omega \in \{b, g\}$. I will denote the different test results by $1, 2, \dots, Z$ and a generic test result by σ. p_σ^ω describes the probability with that the test generates the result σ when it is used by an agent with quality ω. Table 1 displays a tabular description of such a test. σ and θ are independent conditional on ω. Thus, from the perspective of the agent with private signal θ, the test generates the result σ with probability $p^\theta(p_\sigma^b, p_\sigma^g) \equiv (1 - \theta)p_\sigma^b + \theta p_\sigma^g$.

Any test where $p_\sigma^\omega \in \{0, 1\}$ for all ω and all σ is accurate; any test with $p^b = p^g$ is completely inaccurate; I call any other test inaccurate. E.g., the test $((1, 0), (1 - \rho, \rho))$ is accurate if $\rho = 1$, completely inaccurate if $\rho = 0$ and inaccurate if $\rho \in (0, 1)$. Moreover, I say that the test T' is more accurate than the test T'' if it generates, conditional on that it is used for any fixed, non-degenerate distribution of private signals, information that is more informative in the sense of Blackwell. Define $\mathcal{T}_Z \equiv \{(p^b, p^g) \in \Delta^{Z-1} \times \Delta^{Z-1} | p^b \neq p^g, \forall \sigma : p_\sigma^b + p_\sigma^g > 0\}$, $\mathcal{T} \equiv \bigcup_{Z=2}^\infty \mathcal{T}_Z$, and $\mathcal{T}_a \equiv \{(p^b, p^g) \in \mathcal{T} | \forall \sigma : p_\sigma^b = 0 \text{ or } p_\sigma^g = 0\}$. $\mathcal{T}$ describes the set of all not completely inaccurate tests where all test results can be obtained; $\mathcal{T}_Z$ describes the subset that contains only tests with Z test results; $\mathcal{T}_a$ describes the set of all accurate tests.

The timing is as follows:

1. *Test design.* The principal chooses a test $(p^b, p^g) \in \mathcal{T}$.

2. *Quality and private information.* Nature draws $\omega \in \{g, b\}$ and $\theta \in [\underline{\theta}, \overline{\theta}]$.

[9] By assuming that $\underline{\theta} > 0$, I restrict attention to the interesting cases. For $\underline{\theta} = 0$ the test design problem is trivial. I will discuss this in Section 4. The assumption that $\overline{\theta} < 1$ is not crucial but it simplifies the exposition of my analysis. See Weiss (1983) for a similar notion of imperfection of private information.

3. *Participation decision.* After observing the test design (p^b, p^g) and his private signal θ, the agent decides between participation $(d = Y)$ and non-participation $(d = N)$.

4. *Information generation.* If $d = Y$, nature draws a test result σ according to p^ω and reveals it to the principal; otherwise, the principal observes $d = N$.

5. *Updating and reduced-form payoffs.* The principal draws inferences from two sources of information. First, she uses the agent's participation decision to form a Bayesian belief about his private signal. Second, if the test was used, she processes the information revealed by the test to draw additional Bayesian inferences about the agent's quality. I will refer to the probability with which the principal believes that the agent is good after making these inferences as quality perception μ. After μ is formed, the reduced game ends and payoffs that depend directly on μ realize: the agent evaluates μ at a smooth function $u : [0, 1] \to \mathbb{R}$ with $u' > 0$, $u'' < 0$, and $u''' \geq 0$; the principal evaluates μ at a smooth function $v : [0, 1] \to \mathbb{R}$ with $v'' > 0$ and $v''' \geq 0$.[10] I will discuss these reduced-form utility functions in Section 3.2.

3.1. Equilibrium and design problem

I denote the quality perception associated to the test result σ by μ_σ and the one associated to non-participation by μ_N. When the agent's private signal is θ and the quality perceptions associated to the test results are $\mu_Y \equiv (\mu_1, \ldots, \mu_Z)$, the agent realizes an interim expected payoff of $U^\theta(T, \mu_Y) \equiv \sum_\sigma p^\theta(p^b_\sigma, p^g_\sigma)u(\mu_\sigma)$ from participating in the test $T = (p^b, p^g)$. By not participating, he realizes a certain payoff of $u(\mu_N)$. A participation strategy is an indicator function $x : [\underline{\theta}, \overline{\theta}] \to \{0, 1\}$ that describes for which realizations of θ the agent chooses $d = Y$. The participation strategy x is optimal for test T and quality perceptions (μ_N, μ_Y) if $x(\theta) = 1$ whenever $U^\theta(T, \mu_Y) > u(\mu_N)$ and $x(\theta) = 0$ whenever $U^\theta(T, \mu_Y) < u(\mu_N)$.

The belief that the principal holds about the agent's private signal after observing $d \in \{N, Y\}$ is described by a probability distribution over $[\underline{\theta}, \overline{\theta}]$. For my purposes, this distribution will matter only through the expected signal that it implies, say θ_d. θ_d describes the probability with that the principal believes that the agent's quality is good conditional on observing d. Thus, $\mu_N = \theta_N$. If $d = Y$, the principal can use the information revealed by the test result to draw additional inferences. Since $\theta_Y \in [\underline{\theta}, \overline{\theta}] \subset (0, 1)$ for any possible belief about the agent's private signal, Bayes' Law is always applicable to process the additional information revealed

[10] Implicit in this formulation is that participation in the test does not cause cost and that test accuracy is costless. It will be a direct consequence of my solution approach that my results will extend straightforwardly to the case where participation causes cost for principal and/or agent. See also Remark 3 below and Remark S-1 in the supplementary material. The assumption that accuracy is costless allows me to focus on the use of inaccurate tests for strategic reasons. Such an assumption is standard in the information design literature. See Gentzkow and Kamenica (2014) for a notable exception.

by any test result σ:

$$\mu_\sigma = \mu(p^b_\sigma, p^g_\sigma; \theta_Y) \equiv \frac{\theta_Y p^g_\sigma}{\theta_Y p^g_\sigma + (1-\theta_Y)p^b_\sigma}.$$

Let $\mu_Y(T;\theta_Y) \equiv (\mu(p^b_1, p^g_1; \theta_Y), \ldots, \mu(p^b_Z, p^g_Z; \theta_Y))$. I call (μ_N, μ_Y) consistent with the test T and the participation strategy x if $\mu_N = \theta_N$ and $\mu_Y = \mu_Y(T;\theta_Y)$ for some (θ_N, θ_Y) that is consistent with x in the usual game-theoretic sense.[11]

For expositional reasons, I 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.[12] For a given test $T \in \mathcal{T}$, $(x, (\mu_N, \mu_Y))$ is an equilibrium if (EQ1) x is optimal given T and (μ_N, μ_Y), and (EQ2) (μ_N, μ_Y) is consistent with T and x. I denote the set of all equilibria for test T by $\mathcal{E}(T)$. The principal's test design problem corresponds to choosing a test $T \in \mathcal{T}$ and an equilibrium $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T)$ to maximize her expected payoff

$$V(T, x, (\mu_N, \mu_Y)) \equiv \mathbb{E}_\theta[x(\theta) \sum\nolimits_\sigma p^\theta(p^b_\sigma, p^g_\sigma) v(\mu_\sigma) + (1 - x(\theta)) v(\mu_N)]. \tag{1}$$

3.2. Discussion of the reduced-form utility specification

The assumptions on the first two derivatives of the agent's utility function u are essential for the problems in that I am interested in. They capture that the agent likes his quality to be perceived as better ($u' > 0$) but dislikes risk ($u'' < 0$). The assumption on the third derivative ($u''' \geq 0$) is not necessary for deriving my results. It simplifies the analysis and it is satisfied for a large class of commonly used utility functions. For example, my assumptions are satisfied for a broad class of HARA utility functions including quadratic utility, cubic utility, exponential utility/CARA utility, logarithmic utility, and CRRA utility.[13]

The assumption $v'' > 0$ is essential for the problems in that I am interested in. Since updating is Bayesian, the release of any additional information about the agent's quality transforms the posterior quality perception distribution by a mean-preserving spread. The release of any additional information increases the principal's expected payoff if, and only if, $v'' > 0$. Thus, $v'' > 0$ captures that the principal is interested in learning. The assumption that $v''' \geq 0$ is not necessary for my results but it simplifies the analysis. I will discuss the role of this assumption in Remark 1 below and I will relax it in Section 6. An interesting special case that fits my assumptions is $v(\mu) = (\mu - \theta_0)^2$. The principal's objective function $V(T, x, (\mu_N, \mu_Y))$ corresponds then to the variance of the posterior quality perception distribution.

[11] That is, if x prescribes that the decision $d = N$ is taken with positive probability, then $\theta_N = \mathbb{E}_\theta[\theta | x(\theta) = 0]$; if x prescribes that the decision $d = N$ is taken with probability zero, θ_N can assume any value $[\underline{\theta}, \overline{\theta}]$. An analogous reasoning applies for $d = Y$.

[12] I will show below that an equilibrium exists for the subgame that starts after any test choice. The principal's choice is thus well-defined. This two-step procedure is equivalent to considering equilibria of the overall game where the principal's strategy is his test choice and using equilibrium selection by the principal as refinement.

[13] See Section 1 of the supplementary material for a parametrization of HARA utility and for a graphical illustration of the part of the parameter space to that my analysis applies.

	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
$\omega = g$	0	$1 - \rho$	ρ
$\omega = b$	ρ	$1 - \rho$	0

Table 2: A parameterized class of tests $T(\rho)$ with $\rho \in (0, 1]$

4. An illustrative example: Quality estimation with symmetric error cost

Before I derive my theoretical results, I present an illustrative example that demonstrates how a non-reduced version of my model reduces to an instance of my reduced model; it explains for a parameterized class of tests that induces particularly transparent effects what the test does and how updating works; it motivates the trade-offs; and it explains the role of my notion of imperfection of private information.

A non-reduced version of the model and the reduction step. Suppose that the game does not end after the principal forms her quality perception μ but that she has to give an estimate of the probability with that the agent is good, $y \in [0, 1]$. At some stage in the future, the agent's quality comes to light and the principal suffers a quadratic loss from the distance between her estimated probability and the true probability (which is either 0 or 1); her payoff is

$$\widehat{v}(y, \omega) \equiv \begin{cases} -(y-1)^2 & \text{if } \omega = g \\ -(y-0)^2 & \text{if } \omega = b \end{cases}.$$

The agent benefits when the principal's estimate y is higher as this improves the wage he earns, the position he gets, or the status he receives, but he is risk-averse in y; his payoff is $\widehat{u}(y)$ with $\widehat{u}' > 0$, $\widehat{u}'' < 0$, and $\widehat{u}''' \geq 0$.

When the principal's quality perception is μ, she chooses y to maximize her interim expected payoff $\widehat{V}(y; \mu) \equiv -\mu(y-1)^2 - (1-\mu)(y-0)^2$. The optimal decision is $y(\mu) \equiv \mu$; that is, it coincides with the Bayesian estimate. This gives me the reduced-form utility functions $u(\mu) \equiv \widehat{u}(y(\mu)) = \widehat{u}(\mu)$ and $v(\mu) \equiv \widehat{V}(y(\mu), \mu) = -\mu(1-\mu)$, and it leaves me with an instance of my reduced model. In fact, the principal's reduced-form utility function $-\mu(1-\mu)$ induces the same preferences over tests and equilibria as $(\mu - \theta_0)^2$. Thus, the principal strives in this non-reduced example to maximize the variance of the posterior quality perception distribution (see Section 3.2).[14] In Section 6, I will study the more involved problem where one type of error is more costly than the other for the principal. It is then optimal for her to bias her estimate relative to the Bayesian estimate.

Testing and updating. Consider the class of tests $T(\rho)$ that I introduce in Table 2. A higher ρ describes a more accurate test. When the agent participates in the test $T(\rho)$, two things

[14] Harbaugh and Rasmusen (2014) measure information generation by applying a loss function to the error of the public's quality estimate. This example shows that the special case of my reduced model where the principal strives to maximize the variance of the posterior quality perception distribution can be interpreted as an adaptation of their utility specification with a quadratic loss function to my setting with a binary quality. Quadratic loss functions are a standard assumption in sender-receiver games.

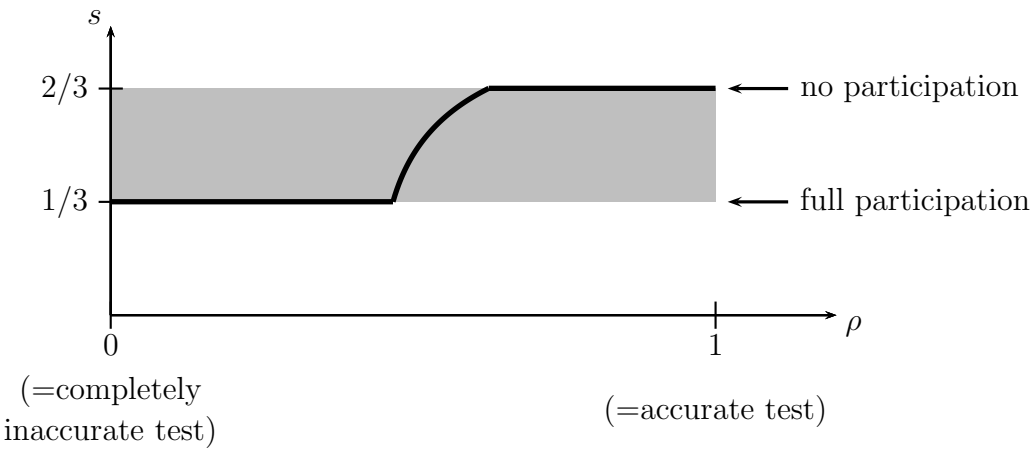


Figure 1: Relation between test accuracy and participation behavior [$u(\mu) = -(1-\mu)^2$, $\theta \sim \mathrm{U}[1/3, 2/3]$]

can happen. With probability ρ the agent's 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 ($\mu_3 = 1$). 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: $\mu_2 = \mathbb{E}_\theta[\theta|\text{“participation”}]$. Similarly, when the test is not used, the principal can still draw inferences about the agent's quality through his participation behavior: $\mu_N = \mathbb{E}_\theta[\theta|\text{“non-participation”}]$.

Trade-offs. Consider the participation incentives of the agent when his private signal is θ. If he participates, the three test results are generated with the probabilities $\rho(1-\theta)$, $1-\rho$ and $\rho\theta$. His private signal affects only the probabilities with that the perfectly informative test results $\sigma = 1$ 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. Thus, the agent has a stronger incentive to participate the higher his private signal. Only threshold participation strategies can be part of an equilibrium.

When I denote the participation threshold by s, non-participation allows the principal to draw the inference $\mathbb{E}_\theta[\theta|\text{“non-participation”}] = \mathbb{E}_\theta[\theta|\theta \leq s]$ whereas participation allows for the inference $\mathbb{E}_\theta[\theta|\text{“participation”}] = \mathbb{E}_\theta[\theta|\theta \geq s]$. From the viewpoint of the agent, participation leads to an expected quality perception of $\rho(1-\theta)\mu_1 + (1-\rho)\mu_2 + \rho\theta\mu_3 = (1-\rho)\mathbb{E}_\theta[\theta|\theta \geq s] + \rho\theta$; non-participation leads to a certain quality perception of $\mu_N = \mathbb{E}_\theta[\theta|\theta \leq s]$. This implies that the agent with the threshold signal $\theta = s$ is on average perceived as better when he participates ($(1-\rho)\mathbb{E}_\theta[\theta|\theta \geq s] + \rho s \geq s$ instead of $\mathbb{E}_\theta[\theta|\theta \leq s] \leq s$) but he faces then a perception risk.

The test accuracy determines for which private signals the agent will participate. Figure 1 illustrates for a numerical example how participation decreases as the test accuracy increases. To get an intuition for the effect of ρ on the participation behavior, consider the polar cases. Suppose first the test is accurate. Then, from the perspective of the agent with any private

signal θ, participation induces a fair quality perception lottery: the quality perception is 1 with probability θ and 0 with probability $1-\theta$. The lottery is non-degenerate as the agent is uncertain about his quality. As the agent is averse to perception risk, participation leads to an expected payoff that is strictly smaller than $u(\theta)$. When the agent does not participate, the principal cannot infer something from his behavior that the agent does not know himself. This limits how bad the principal's perception of the agent's quality can be. The most adverse perception is that the agent has the worst possible private signal $\underline{\theta}$. As such a perception implies a certain utility of $u(\underline{\theta})$, the agent has a strict incentive not to participate when his private signal is close to $\underline{\theta}$. Thus, an accurate test can induce only partial participation. On the other hand, the agent faces almost no perception risk when he participates in a very inaccurate test. When ρ is close to zero, participation allows the agent to significantly increase how he is perceived on average at almost no "cost".[15] As a consequence, there is full participation for any informative but sufficiently inaccurate test. It follows that participation decreases—at least eventually—as the test accuracy ρ increases. This gives rise to the interpretation that participation can be fostered by making the test subject to errors.[16]

The role of my notion of imperfection of private information. My notion of imperfection of private information renders the test design problem interesting. If it is not imperfect in the sense that he is never certain to be bad (i.e., if $\underline{\theta}=0$), a full unraveling equilibrium exists for any accurate test.[17] Perfect learning is possible. If there is no meaningful private information (i.e., 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 not willing to participate in any informative test. This renders learning about his quality impossible.

5. Optimal test design in the reduced model

In the preceding section I imposed an ad hoc restriction on the structure of the considered tests. This gave me a one-dimensional test design problem that was suitable for motivating effects. Most interesting is, however, the question of how the optimal test structure does actually look like. This is the main objective of this article. Once I know this structure, the principal's design problem reduces to a much simpler, one-dimensional problem that is similar to the problem that I discussed in the preceding section.

[15] When the supposed participation threshold is s, participation in a very inaccurate test increases the average perception by approximately $\mathbb{E}_\theta[\theta|\theta\geq s]-\mathbb{E}_\theta[\theta|\theta\leq s]$. The "significance" part of the statement derives from the fact that the minimal value of $\mathbb{E}_\theta[\theta|\theta\geq s]-\mathbb{E}_\theta[\theta|\theta\leq s]$ is bounded away from zero. Thus, no matter which participation threshold is induced for a given ρ, participation increases the average perception significantly. By contrast, the "cost" that is caused by the perception risk vanishes as $\rho\to 0$.

[16] In a certification context, De and Nabar (1991) show that a similar trade-off can arise with a perfectly informed, risk-neutral seller who can voluntarily decide to get tested at a fee: an inaccurate testing technology can foster participation relative to an accurate one.

[17] If $\underline{\theta}=0$, the worst possible private information is verifiable by an accurate test. The reasoning in the literature on information transmission with hard information applies to this case. See Grossman and Hart (1980), Grossman (1981), Milgrom (1981), and Okuno-Fujiwara et al. (1990).

5.1. Inducible participation behavior

Take at first any test $T \in \mathcal{T}$ as given. Test results associated with higher quality perceptions are good news in the sense of Milgrom (1981). As a higher private signal makes more favorable news more likely in the sense of first-order stochastic dominance, participation is more attractive the higher the agent's private signal. Thus, any equilibrium $(x, (\mu_N, \mu_Y))$ must exhibit threshold participation behavior. That is, $x(\theta) = 0$ if $\theta < s$ and $x(\theta) = 1$ if $\theta > s$ for some $s \in [\underline{\theta}, \overline{\theta}]$.

For threshold participation behavior the inference that the principal can draw about the agent's private signal has a simple form: if the agent is supposed to participate with positive probability (i.e., $s \in [\underline{\theta}, \overline{\theta})$), she believes that the agent is good with probability $\theta_Y(s) \equiv \mathbb{E}_\theta[\theta|\theta \geq s]$ conditional on observing participation; if the agent is supposed to not participate with positive probability (i.e., $s \in (\underline{\theta}, \overline{\theta}]$), she believes that the agent is good with probability $\theta_N(s) \equiv \mathbb{E}_\theta[\theta|\theta \leq s]$ conditional on observing non-participation; only for behavior that is supposed to occur with probability zero the consistency requirement (EQ2) does not pin down this belief. As zero probability events do not affect the principal's expected payoff, the participation threshold s summarizes all information of an equilibrium that is relevant for the principal. This allows me to write her expected payoff as a function of T and s.

Proposition 1 (Only threshold participation strategies inducible) *Fix any test $T \in \mathcal{T}$. (a) If $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T)$, then x is a threshold strategy. If the threshold is $s \in [\underline{\theta}, \overline{\theta}]$, then*

$$V(T, x, (\mu_N, \mu_Y)) = (1 - F(s)) \sum\nolimits_\sigma p^{\theta_Y(s)}(p^b_\sigma, p^g_\sigma) v(\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s))) + F(s) v(\theta_N(s)). \quad (2)$$

(b) $\mathcal{E}(T)$ is non-empty.

Next I come to the question of which participation thresholds the principal can obtain by an appropriate design of the test. I say that the test T induces the participation threshold s if there exists a threshold participation strategy x with the threshold s and a vector of quality perceptions (μ_N, μ_Y) such that $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T)$; I say that the threshold s is inducible (inducible by an accurate test), if there exists some test $T \in \mathcal{T}$ ($T \in \mathcal{T}_a$) that induces s. I can restrict attention to the inducement of thresholds $s \in [\underline{\theta}, \overline{\theta})$ that imply participation with positive probability as only such thresholds are interesting from a design perspective. The subsequent proposition states simple conditions on the primitives of the model that determine whether a given threshold $s \in [\underline{\theta}, \overline{\theta})$ can be induced by some test and by an accurate test.

Proposition 2 (Inducible thresholds) *(a) The participation threshold $s \in [\underline{\theta}, \overline{\theta})$ is inducible by a test $T \in \mathcal{T}_a$ if, and only if, $s \in \mathcal{S}_a \equiv \{s \in [\underline{\theta}, \overline{\theta})|(1-s)u(0) + su(1) = u(\mathbb{E}_\theta[\theta|\theta \leq s])\}$. (b) The participation threshold $s \in [\underline{\theta}, \overline{\theta})$ is inducible by some test $T \in \mathcal{T}$ if, and only if, $s \in \mathcal{S} \equiv \{s \in [\underline{\theta}, \overline{\theta})|(1-s)u(0) + su(1) \leq u(\mathbb{E}_\theta[\theta|\theta \leq s])\}$.*

Intuitively, inducing a threshold s consists of two parts: the motivation of participation for private signals $\theta > s$ and the deterrence of participation for private signals $\theta < s$. By making the test sufficiently inaccurate, the motivation part can always be satisfied. Thus, crucial for inducibility is the deterrence part. This part can be satisfied if, and only if, the agent has not

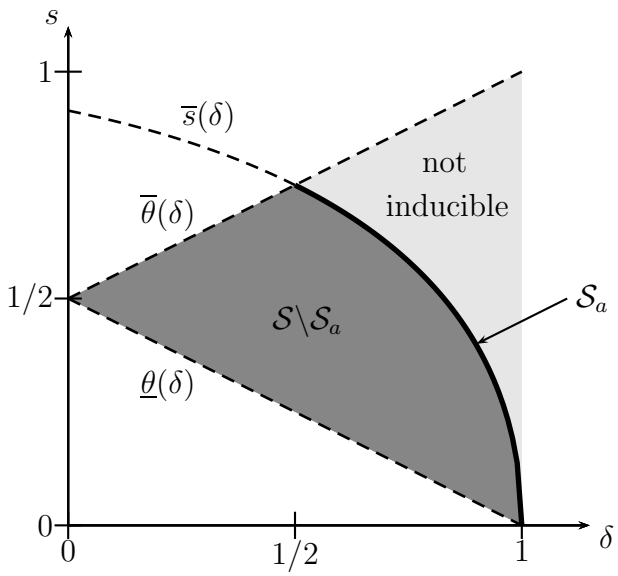


Figure 2: Inducible participation thresholds $[u(\mu) = -(1-\mu)^2,\ \theta \sim \mathrm{U}[\underline{\theta}(\delta), \overline{\theta}(\delta)]]$

a strict incentive to participate in an accurate test when his private signal is $\theta = s$. It is then possible to construct a test for that the participation constraint of the agent with the private signal $\theta = s$ is just binding. This test induces s. In fact, any class of tests that continuously transforms an accurate test into a completely inaccurate test contains a test that induces s.

Example: Inducible participation thresholds in the uniform-quadratic case. To get an intuition for what determines the inducibility of a threshold s, consider the special case where $u(\mu) = -(1-\mu)^2$ and where θ is uniformly distributed on $[\underline{\theta}(\delta), \overline{\theta}(\delta)]$ with $\underline{\theta}(\delta) \equiv (1-\delta)/2$, $\overline{\theta}(\delta) \equiv (1+\delta)/2$ and $\delta \in (0,1)$. A higher δ is associated with more informative private information. I obtain $\mathcal{S}_a = \{\overline{s}(\delta)\}$ and $\mathcal{S}\backslash\mathcal{S}_a = \{s \in [\underline{\theta}(\delta), \overline{\theta}(\delta)) | s < \overline{s}(\delta)\}$ with $\overline{s}(\delta) \equiv 2\sqrt{\underline{\theta}(\delta)} - \underline{\theta}(\delta)$. Figure 2 illustrates how the sets $\mathcal{S}$ and $\mathcal{S}_a$ depend on δ.

If $\delta < 1/2$, the agent's signaling motive is weak 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 that the agent is sufficiently certain to be good such that he cannot be deterred from participating by increasing the test accuracy. Every threshold $s \in (\overline{s}(\delta), \overline{\theta}(\delta)]$ is not inducible. Due to a partial unraveling effect that becomes stronger as δ increases, the set of private signals for that 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.

5.2. Discussion of the test design problem

Proposition 2 provides a simple characterization of the set of all inducible participation thresholds $\mathcal{S}$ in terms of the primitives of the model. This allows me to tackle the test design problem in two steps. First, I will derive the test that optimally induces any given participation threshold $s \in \mathcal{S}$. Then, I will discuss the problem of which participation threshold from $\mathcal{S}$ it is

	$\sigma=1$	$\sigma=2$
$\omega=g$	$1-p_2^g$	p_2^g
$\omega=b$	$1-p_2^b$	p_2^b

(a) General binary test

	$\sigma=1$	$\sigma=2$
$\omega=g$	$1-\rho$	ρ
$\omega=b$	1	0

(b) No false positives test: $T^{\text{NFP}}(\rho)$ with $\rho \in (0,1]$

	$\sigma=1$	$\sigma=2$
$\omega=g$	0	1
$\omega=b$	ρ	$1-\rho$

(c) No false negatives test: $T^{\text{NFN}}(\rho)$ with $\rho \in (0,1]$

Table 3: Binary tests

optimal to induce. This two-step approach will convey a better understanding of the underlying effects as it will separate the parts of the derivation that rely on the primitives of the model from those that do not: A single test structure will turn out to be optimal for all primitives of the model while the optimal participation threshold will depend on the primitives.

The subsequent lemma states that the first part of the design problem has a trivial solution when the desired threshold is inducible by an accurate test; otherwise, the problem can be simplified.

Lemma 1 (The inducement problem) *(a) Consider any $s \in \mathcal{S}_a$. Then, among all tests that induce s, each test $T \in \mathcal{T}_a$ is optimal. (b) Consider any $s \in \mathcal{S}\backslash\mathcal{S}_a$. The test T is optimal among all tests that induce s if, and only if, T solves the following problem $\mathsf{TD}(s)$:*

$$\begin{array}{ll} \max\limits_{(p^b,p^g)\in\mathcal{T}} & \sum_\sigma p^{\theta_Y(s)}(p_\sigma^b,p_\sigma^g)v(\mu(p_\sigma^b,p_\sigma^g;\theta_Y(s))) \\ s.t. & \sum_\sigma p^{s}(p_\sigma^b,p_\sigma^g)u(\mu(p_\sigma^b,p_\sigma^g;\theta_Y(s))) = u(\theta_N(s)). \end{array}$$

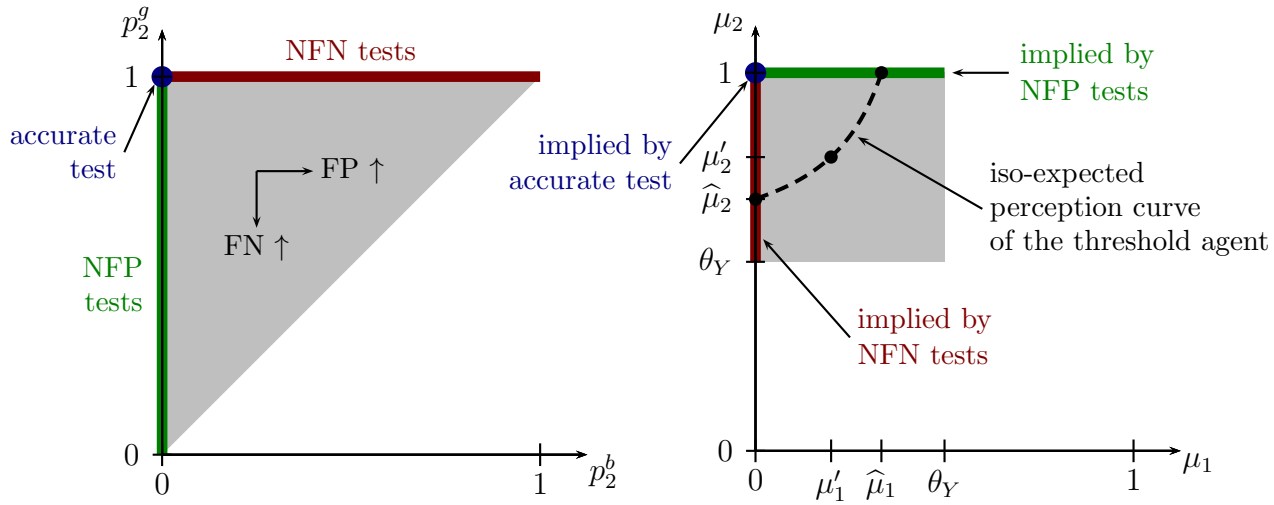
Part (b) states that three intuitive simplifications of the original design problem are without loss of generality. First, I only need to consider the principal's expected payoff conditional on participation. Second, only the participation constraint of the agent with the threshold signal is relevant. This allows me to interpret the simplified problem as a problem between the principal and the "threshold agent". Third, I can assume that $\theta_N = \theta_N(s)$ holds also for $s = \underline{\theta}$.

5.3. Quality perception design and the structure of the optimal binary test

For any $s \in \mathcal{S}\backslash\mathcal{S}_a$, I study at first the version of the problem $\mathsf{TD}(s)$ where the maximization is only over binary tests $(p^b,p^g) \in \mathcal{T}_2$. I will denote this problem by $\mathsf{BTD}(s)$. Binary tests will suffice to motivate the forces that drive the optimal test design and to derive the methods that allow me to solve also the non-binary problem.

A binary test is completely described by the two parameters p_2^b and p_2^g. See Table 3 (a). I can without loss of generality restrict attention to the case where $p_2^b < p_2^g$. Under this assumption, the test result $\sigma = 2$ ($\sigma = 1$) is relatively more often obtained by a good (bad) agent. Since this implies that $\sigma = 2$ is associated with a higher quality perception, I will refer to $\sigma = 2$ as the *positive* result and to $\sigma = 1$ as the *negative* result.

Figure 3(a) illustrates the parameter space and important special cases in it. The binary test with $p_2^b = 0$ and $p_2^g = 1$ is accurate. The agent always obtains the positive result when he is good and the negative result when he is bad. The test is subject to false negatives (FN)



(a) Binary tests

(b) Quality perception pairs

Figure 3: Relation between binary tests and the implied quality perception pairs [$\theta_Y = 1/2$]

if $p_2^g < 1$ and subject to false positives (FP) if $p_2^b > 0$. Tables 3 (b) and (c) describe two subclasses of binary tests that will be particularly important. Any test $T^{\text{NFP}}(\rho)$ is not subject to false positives (NFP): the positive result $\sigma = 2$ perfectly reveals that the agent is good whereas there is pooling on the negative result $\sigma = 1$. Likewise, any test $T^{\text{NFN}}(\rho)$ is not subject to false negatives (NFN): the negative result $\sigma = 1$ perfectly reveals that the agent is bad whereas there is pooling on the positive result $\sigma = 2$. For each of the two classes, a higher parameter ρ is associated with a higher test accuracy.

In order to be able to apply my analysis directly to the non-binary problem later on, I investigate a generalized version of the problem $\mathsf{BTD}(s)$. Consider the problem $\mathsf{BTD}(s, \theta_Y, \overline{u})$:

$$\begin{aligned} \max_{p_2^b, p_2^g \in [0,1],\ p_2^b < p_2^g} \quad & \mathcal{V}_{\mathcal{T}}(p_2^b, p_2^g; \theta_Y) \equiv p^{\theta_Y}(1-p_2^b, 1-p_2^g) \cdot v(\mu(1-p_2^b, 1-p_2^g; \theta_Y)) \\ & \qquad + p^{\theta_Y}(p_2^b, p_2^g) \cdot v(\mu(p_2^b, p_2^g; \theta_Y)) \\ \text{s.t.} \quad & \mathcal{U}^s_{\mathcal{T}}(p_2^b, p_2^g; \theta_Y) \equiv p^{s}(1-p_2^b, 1-p_2^g) \cdot u(\mu(1-p_2^b, 1-p_2^g; \theta_Y)) \\ & \qquad + p^{s}(p_2^b, p_2^g) \cdot u(\mu(p_2^b, p_2^g; \theta_Y)) = \overline{u} \end{aligned}$$

with $s \in [0, 1)$, $\theta_Y \in (s, 1)$ and $\overline{u} \in \overline{\mathcal{U}}(\theta_Y) \equiv [su(1)+(1-s)u(0), u(\theta_Y))$. θ_Y describes the average probability with that a participating agent is good; s describes the below average probability with that the threshold agent believes that he is good; $\overline{u}$ describes the outside option of the threshold agent. The problem $\mathsf{BTD}(s)$ is the special case of this generalized problem with $s \in \mathcal{S} \backslash \mathcal{S}_a$, $\theta_Y = \theta_Y(s)$ and $\overline{u} = u(\theta_N(s))$.

p_2^b and p_2^g affect the principal's objective function $\mathcal{V}_{\mathcal{T}}(p_2^b, p_2^g; \theta_Y)$ and the threshold agent's participation constraint $\mathcal{U}^s_{\mathcal{T}}(p_2^b, p_2^g; \theta_Y) = \overline{u}$ in a complex way. To get a better understanding of what drives the optimal test design, it is illuminative to consider how the objective function and the constraint are affected by changes in the implied quality perceptions. The subsequent lemma establishes the relation between binary tests and quality perception pairs.

Lemma 2 (Relation between binary tests and quality perception pairs) *Suppose that* $\theta_Y \in (0,1)$ *and consider the system of the two equations* $\mu_1 = \mu(1-p_2^b, 1-p_2^g; \theta_Y)$ *and* $\mu_2 = \mu(p_2^b, p_2^g; \theta_Y)$. *Any binary test with* $p_2^b < p_2^g$ *implies a quality perception pair* $(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y) \equiv [0, \theta_Y) \times (\theta_Y, 1]$. *Conversely, any quality perception pair* $(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y)$ *is implied by a unique binary test with* $p_2^b < p_2^g$. *This test is characterized by* $p_2^b = \beta^b(\mu_1, \mu_2; \theta_Y)$ *and* $p_2^g = \beta^g(\mu_1, \mu_2; \theta_Y)$ *with*

$$\beta^b(\mu_1, \mu_2; \theta_Y) \equiv \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1} \cdot \frac{1-\mu_2}{1-\theta_Y} \text{ and } \beta^g(\mu_1, \mu_2; \theta_Y) \equiv \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1} \cdot \frac{\mu_2}{\theta_Y}. \tag{3}$$

The grey area in Figure 3(b) displays the set of quality perception pairs that can be induced by some test, $\mathcal{Q}(\theta_Y)$. The quality perception pair $(0,1)$ is associated with the accurate test; any quality perception pair $(\mu_1, 1)$ is associated with a test $T^{\text{NFP}}(\rho)$; and any quality perception pair $(0, \mu_2)$ is associated with a test $T^{\text{NFN}}(\rho)$.

I can use Lemma 2 to describe how the agent is perceived on average as a function of the implied quality perceptions when his private signal is θ:

$$\overline{\mu}^{\theta}(\mu_1, \mu_2; \theta_Y) \equiv \mu_1 + p^{\theta}(\beta^b(\mu_1, \mu_2; \theta_Y), \beta^g(\mu_1, \mu_2; \theta_Y)) \cdot (\mu_2 - \mu_1). \tag{4}$$

Since this implies that $p^{\theta}(\beta^b(\mu_1, \mu_2; \theta_Y), \beta^g(\mu_1, \mu_2; \theta_Y)) = (\overline{\mu}^{\theta}(\mu_1, \mu_2; \theta_Y) - \mu_1)/(\mu_2 - \mu_1)$, I obtain the following quality perception design problem $\mathsf{QPD}(s, \theta_Y, \overline{u})$:

$$\begin{aligned} \max_{(\mu_1,\mu_2)\in\mathcal{Q}(\theta_Y)} \quad & \mathcal{V}_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y) \equiv v(\mu_1) + \frac{\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - \mu_1}{\mu_2 - \mu_1} \cdot (v(\mu_2) - v(\mu_1)) \\ \text{s.t.} \quad & \mathcal{U}^s_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y) \equiv u(\mu_1) + \frac{\overline{\mu}^{s}(\mu_1, \mu_2; \theta_Y) - \mu_1}{\mu_2 - \mu_1} \cdot (u(\mu_2) - u(\mu_1)) = \overline{u}. \end{aligned}$$

By construction, the quality perception pair (μ_1, μ_2) solves the problem $\mathsf{QPD}(s, \theta_Y, \overline{u})$ if, and only if, the binary test with $p_2^b = \beta^b(\mu_1, \mu_2; \theta_Y)$ and $p_2^g = \beta^g(\mu_1, \mu_2; \theta_Y)$ solves the problem $\mathsf{BTD}(s, \theta_Y, \overline{u})$.

Next I come to the question of how a change in (μ_1, μ_2) affects the threshold agent and the principal. The subsequent lemma states that movements to the northwest or the southeast of the quality perception space $\mathcal{Q}(\theta_Y)$ induce clear-cut effects.

Lemma 3 (Preferences over quality perception pairs I) *Suppose that* $s \in [0,1)$ *and that* $\theta_Y \in (s,1)$. *(a)* $\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) = \theta_Y$ *for any* $(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y)$. *(b)* $\overline{\mu}^{s}(\mu_1, \mu_2; \theta_Y)$ *is strictly increasing in* μ_1 *and strictly decreasing in* μ_2 *on* $\mathcal{Q}(\theta_Y)$. *(c)* $\mathcal{V}_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y)$ *is strictly decreasing in* μ_1 *and strictly increasing in* μ_2 *on* $\mathcal{Q}(\theta_Y)$. *(d)* $\mathcal{U}^s_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y)$ *is strictly increasing in* μ_1 *and strictly decreasing in* μ_2 *on* $\mathcal{Q}(\theta_Y)$.

Part (a) is a direct consequence of Bayesian updating. Part (b) is very intuitive. As μ_1 decreases and/or μ_2 increases, both possible inferences about the agent's quality become more accurate by moving away from θ_Y; that is, there is less pooling between the case where the agent is good and that where he is bad. This decreases the expected perception from the viewpoint of the threshold agent, who believes to be good with a below average probability. Parts (c) and (d) describe the effects on the principal's and the threshold agent's expected payoff. From

the perspective of the principal, more accurate inferences transform the quality perception distribution by a mean-preserving spread. Since v is strictly convex, this is unambiguously good for the principal. By contrast, from the perspective of the threshold agent, the quality perception distribution is transformed by a mean-decreasing spread. Because he strives for high quality perceptions ($u' > 0$) and is averse to perception risk ($u'' < 0$), he is clearly worse off.

How the principal and the threshold agent are affected when the quality perception pair moves to the northeast or the southwest depends on the details of the model. However, if I consider only modifications in (μ_1, μ_2) that leave the expected perception of the threshold agent constant, I obtain clear-cut effects again. The iso-expected perception curve of the threshold agent through any quality perception pair $(\mu_1', \mu_2') \in \mathcal{Q}(\theta_Y)$ is given by

$$\overline{\mathcal{Q}}^s(\mu_1', \mu_2'; \theta_Y) \equiv \{(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y) | \overline{\mu}^s(\mu_1, \mu_2; \theta_Y) = \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)\}.$$

The dashed curve in Figure 3(b) illustrates how this curve does typically look like. Part (b) of the subsequent lemma establishes the key property that will drive my main results. Among the quality perception pairs on any iso-expected perception curve, there is no conflict of interest between the threshold agent and the principal. Parts (a) and (c) establish technical properties that will be useful in the subsequent proofs.

Lemma 4 (Preferences over quality perception pairs II) *Suppose that $s \in [0, 1)$, that $\theta_Y \in (s, 1)$ and that $(\mu_1', \mu_2') \in \mathcal{Q}(\theta_Y)$. (a) $\overline{\mathcal{Q}}^s(\mu_1, \mu_2; \theta_Y)$ does not depend on s such that I can write $\overline{\mathcal{Q}}^s(\mu_1, \mu_2; \theta_Y) = \overline{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y)$. (b) Consider any $(\mu_1'', \mu_2'') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ with $\mu_2' < \mu_2''$. Then, $\mathcal{U}_{\mathcal{Q}}^s(\mu_1', \mu_2'; \theta_Y) < \mathcal{U}_{\mathcal{Q}}^s(\mu_1'', \mu_2''; \theta_Y)$ and $\mathcal{V}_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y) \leq \mathcal{V}_{\mathcal{Q}}(\mu_1'', \mu_2''; \theta_Y)$. (c) For any $\mu_2'' \in (\mu_2', 1]$ there exists a unique $\mu_1'' \in [0, \theta_Y)$ such that $(\mu_1'', \mu_2'') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ and a unique $\mu_1''' \in [0, \theta_Y)$ such that $\mathcal{U}_{\mathcal{Q}}^s(\mu_1''', \mu_2''; \theta_Y) = \mathcal{U}_{\mathcal{Q}}^s(\mu_1', \mu_2'; \theta_Y)$. Moreover, $\mu_1' < \mu_1''' < \mu_1''$.*

To get an intuition for Part (b), consider the preferences over the two polar quality perception pairs on a given iso-expected perception curve $\overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$, say $(0, \widehat{\mu}_2)$ and $(\widehat{\mu}_1, 1)$. The first (second) pair is associated with a NFN (NFP) test. See Figure 3(b) for an illustration. By construction, the considered quality perception pairs imply the same expected perception from the viewpoint of the threshold agent. By Part (a) of the lemma, this implies that also the expected perception from the perspective of somebody who believes that the agent is good with *any* other probability θ is the same for both quality perception pairs. Technically, this is a consequence of my supposition $\overline{\mu}^s(0, \widehat{\mu}_2; \theta_Y) = \overline{\mu}^s(\widehat{\mu}_1, 1; \theta_Y)$, the fact that Bayesian updating implies $\overline{\mu}^{\theta_Y}(0, \widehat{\mu}_2; \theta_Y) = \overline{\mu}^{\theta_Y}(\widehat{\mu}_1, 1; \theta_Y) = \theta_Y$, and the linearity of $\overline{\mu}^{\theta}(\mu_1, \mu_2; \theta_Y)$ in θ.

Suppose at first that $u''' = 0$. The agent cares then only about the interim expected perception and the interim variance of perception. In the hypothetical situation where he is certain to be bad, a NFP test does not induce a perception risk whereas a NFN test does; since he is strictly averse towards perception risk and both tests induces the same interim expected perception, he strictly prefers the NFP test. The proof shows that this intuition carries over to the case where the agent believes to be good with any below average probability $\theta \in [0, \theta_Y)$; in particular, it holds for the agent with the threshold signal. If $u''' > 0$, an additional effect comes

into play. The agent has then an additional interest in inducing high quality perceptions. As the quality perception pair that is optimal for $u''' = 0$ is already the quality perception pair on the iso-expected perception curve with the component-by-component largest quality perceptions (see, again, Figure 3(b)), the additional effect reinforces the agent's preference for $(\widehat{\mu}_1, 1)$ over $(0, \widehat{\mu}_2)$. Hence, the NFP test stays optimal for the threshold agent.

Consider next the principal and suppose at first that $v''' = 0$. As the principal is risk-loving, the effects are for her reversed compared with the agent; that is, in the hypothetical situation where she believes that the agent is good with below (above) average probability, she prefers the NFN test (NFP test). The relevant case where the principal believes that the agent is good with the average probability θ_Y describes the knife-edge case where she is just indifferent between both tests. However, if $v''' > 0$ instead of $v''' = 0$, her preference for the NFP test over the NFN test becomes strict.

I am now set to explain the structure of the optimal binary test that induces s. The solid curve in Figure 4 illustrates the quality perception pairs that satisfy a given participation constraint of the threshold agent. The problem is to find the quality perception pair on this curve that is optimal for the principal. Assume to the contrary that (μ_1', μ_2') with $\mu_2' < 1$ is optimal. The dashed curve in Figure 4 illustrates the iso-expected perception curve of the threshold agent through (μ_1', μ_2'). By Lemma 4 (b), the threshold agent and the principal face no conflict of interest among the quality perception pairs on this curve. The expected payoff of both of them increases as I move to the northeast on this curve. This means, in particular, that the agent prefers the quality perception pair that lies farthest to the northeast, $(\mu_1'', 1)$, over the original quality perception pair (μ_1', μ_2'). This gives scope for increasing the test accuracy by moving in the quality perception space to the west. By Lemma 3 (d), I so obtain a quality perception pair $(\mu_1''', 1)$ that makes the agent indifferent to (μ_1', μ_2') and that does thus also induce s. As the increase in the test accuracy makes the principal by Lemma 3 (c) even better off, I obtain a contradiction to the optimality of (μ_1', μ_2'). Hence, only a quality perception pair $(\mu_1, 1)$ that is associated to a NFP test can be optimal.

Proposition 3 (Optimal quality perception pair/binary test) *Suppose that $s \in [0, 1)$, that $\theta_Y \in (s, 1)$ and that $\overline{u} \in \overline{\mathcal{U}}(\theta_Y)$. (a) The problem $\mathsf{QPD}(s, \theta_Y, \overline{u})$ has a unique solution. The quality perception pair $(\mu_1, 1)$ where μ_1 is implicitly defined by $\mathcal{U}_Q^s(\mu_1, 1; \theta_Y) = \overline{u}$ is optimal. (b) The problem $\mathsf{BTD}(s, \theta_Y, \overline{u})$ has a unique solution. The test $T^{NFP}(\rho)$ where ρ is implicitly defined by $\mathcal{U}_T^s(0, \rho; \theta_Y) = \overline{u}$ is optimal.*

The solution to the generalized binary test design problem $\mathsf{BTD}(s, \theta_Y, \overline{u})$ directly implies the solution to the binary test design problem $\mathsf{BTD}(s)$: For any participation threshold that shall be induced and for any primitives of the model, a NFP test is strictly optimal. Before I show that the optimal test is binary, let me discuss some robustness issues.

Remark 1 (The role of the third derivative assumptions) Consider first the role of the assumption that $v''' \geq 0$. To prove Proposition 3, I transform any given quality perception pair (μ_1', μ_2') with $\mu_2' < 1$ that induces s in two steps into a quality perception pair $(\mu_1''', 1)$ that does

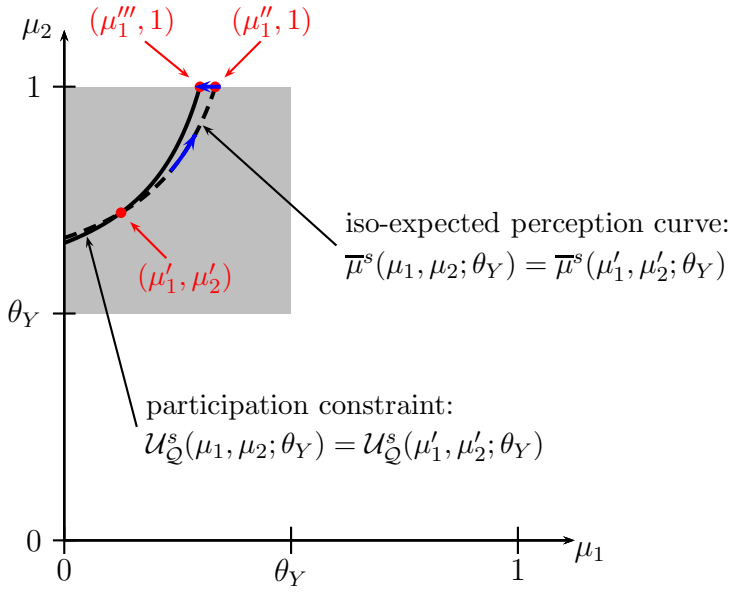


Figure 4: Illustration of the strategy of proof of Proposition 3 [$u(\mu) = -(1-\mu)^2$, $\theta_Y = 1/2$, $s = 1/3$]

also induce s. The second step is always beneficial for the principal; $v''' \geq 0$ implies that also the first step makes her weakly better off. Crucial for my result is, however, only the compound effect. When the second step makes the principal sufficiently much better off, a violation of $v''' > 0$ does not affect the optimality of a NFP test. Intuitively, if the agent's risk-aversion is strong, the second step makes the principal much better off and a violation of $v''' \geq 0$ is likely to stay without consequences. Extending the result to problems where $v''' \geq 0$ is violated requires me to impose additional structure on u and v. I demonstrate such an extension in Section 6.

Consider next the role of the assumption that $u''' \geq 0$. For the threshold agent, it is crucial that the first step of the transformation is beneficial. Intuitively, this is the case even when $u''' \geq 0$ is violated when the agent is sufficiently averse towards perception risk.

Remark 2 (More general distributions of private information) The derivation of the solution to the problem $\mathsf{BTD}(s)$ depends only through the expected value of θ conditional on participation $\theta_Y(s)$ and conditional on non-participation $\theta_N(s)$ on the distribution of θ. As it does not matter for my results how these expected values arise, an extension of my analysis to the case where the distribution of θ includes atoms and holes is straightforward.

Remark 3 (Exogenous participation constraint) The analysis in this subsection did not rely on the assumption that the agent's private information is imperfect and I could easily extend it to the case where the agent is perception risk-neutral. If $u'' = 0$, the only difference in the analysis would be that some effects that are strict for $u'' > 0$ become weak. In particular, a NFP would be strictly optimal if $v''' > 0$ but only weakly optimal if $v''' = 0$. However, the two assumptions rendered the analysis in this subsection relevant. They were responsible for the emergence of an endogenous participation constraint that prevented full unraveling under an accurate test. If a participation constraint arises for other reasons, the analysis in this section applies also to problems where the agent is perfectly informed and/or cares only about expected perception. For instance, such a constraint could arise like in Harbaugh and Rasmusen (2014) where the agent suffers an exogenously given cost from participating.

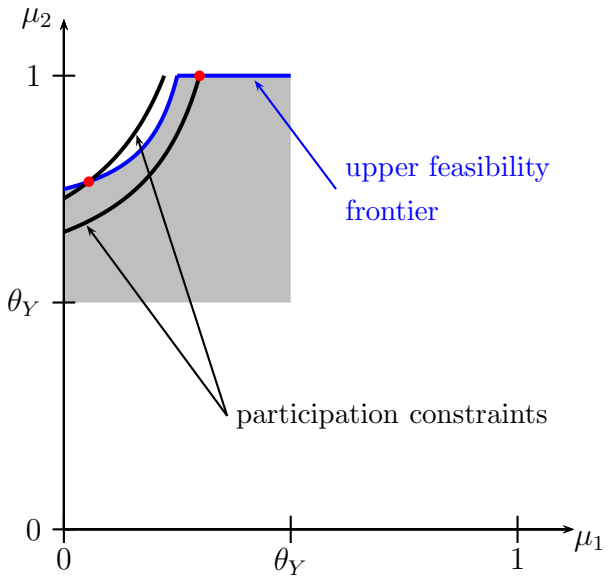


Figure 5: Illustration of Remark 4 [$u(\mu) = -(1-\mu)^2$, $\theta_Y = 1/2$, $s = 1/3$, $c = 1/3$]

Remark 4 (Exogenous technological limits on the test accuracy) The transformation of the test design problem into a quality perception design problem allows me also to easily handle additional constraints that limit the the ability to generate information. As an example, suppose that only binary tests where the joint probability of false positives and false negatives is at least $c \in (0,1)$ are feasible; that is, $p_2^b + (1-p_2^g) \geq c$ must hold. Lemma 2 allows me to transform this restriction into a restriction on the quality perception pairs (μ_1, μ_2) that can be induced: $\beta^b(\mu_1, \mu_2; \theta_Y) + (1 - \beta^g(\mu_1, \mu_2; \theta_Y)) \geq c$. The grey area in Figure 5 illustrates all quality perception pairs that can be induced by some feasible test. It follows directly from Lemma 3 and 4 that only quality perception pairs on the upper feasibility frontier can be optimal. This frontier is illustrated by the blue curve. To find the optimal quality perception pair, I need only to compare the points on this curve that satisfy the threshold agent's participation constraint. The two black curves in Figure 5 illustrate two different participation constraints; the red dot on each of the curves illustrates the unique candidate for the optimal quality perception pair. Both cases have in common that the optimal quality perception pair is associated with the test for which false positives are least likely among all tests that satisfy the participation constraint.

In the supplementary material, I explain also what can be learned from my analysis in this subsection about problems with non-binary quality. First, I discuss a version of my problem where quality is ultimately still binary but where information-generation is only possible about non-binary quality types (Section 2 of the supplementary material). As an example, think of a bank that is ultimately either viable or not. At the time of testing, it may only be possible to generate information about whether the bank is definitely viable, definitely non-viable or that it is only viable under certain circumstances. Then, I discuss a version of my problem where quality is really non-binary but where payoffs depend only on the expected value of the quality conditional on the principal's information (Section 3 of the supplementary material).

5.4. The structure of the optimal test

I explain now how the insights from the design of the optimal binary test can be used to explain why an optimal test exists that is binary. Suppose that the participation threshold $s \in \mathcal{S}\backslash\mathcal{S}_a$ shall be induced. The trick is to consider the non-binary test design problem $\mathsf{TD}(s)$ as a sequence of binary problems. To be more specific, assume to the contrary that the non-binary test $\widehat{T} \in \mathcal{T}_Z$ induces s and that it is strictly better than any binary test that induces s. Moreover, suppose without loss of generality that the implied quality perceptions are ordered and different from each other; that is, $\widehat{\mu}_1 < \widehat{\mu}_2 < \ldots < \widehat{\mu}_Z$ for $(\widehat{\mu}_1, \ldots, \widehat{\mu}_Z) \equiv \mu_Y(\widehat{T}; \theta_Y(s))$. When I consider now modifications of the test that affect only the first two test results, I obtain a binary problem. More specifically, the described problem corresponds to a generalized binary test design problem $\mathsf{BTD}(\widehat{s}, \widehat{\theta}_Y, \widehat{\overline{u}})$ with modified parameter values (see the proof of Proposition 4 for details). The analysis in Section 5.3 applies directly to this problem. In particular, I can transform this problem into the binary quality perception design problem $\mathsf{QPD}(\widehat{s}, \widehat{\theta}_Y, \widehat{\overline{u}})$ to that Lemma 3 and Lemma 4 apply. This allows me to apply modifications with the following properties:

Modification 1: (i) The first two quality perceptions increase while all other quality perceptions stay unaffected; (ii) the threshold agent (principal) becomes strictly (weakly) better off; and (iii) the expected perception of the threshold agent does not change.

Modification 2: (i) The first quality perception decreases while all other quality perceptions stay unaffected; (ii) the threshold agent (principal) becomes strictly worse (better) off; and (iii) the expected perception of the threshold agent decreases strictly.

This allows me to modify the test $\widehat{T}$ that induces s in the following way: First, apply Modification 1 until the second quality perception hits $\widehat{\mu}_3$. This gives me a test that is weakly better for the principal and strictly better for the threshold agent. Then, by applying Modification 2 to the so obtained test, I obtain a test $\widehat{T}'$ that induces again s but that is strictly better for the principal than the initial test $\widehat{T}$. By merging now the test results 2 and 3, which generate the same quality perception $\widehat{\mu}_3$, I obtain a test $\widehat{T}''$ that induces s with one test result less and that is strictly better for the principal than the initial test $\widehat{T}$. As I can iterate this procedure until I obtain a binary test, I obtain a contradiction to my supposition that $\widehat{T}$ is strictly better for the principal than any binary test that induces s. Hence, I need only to search among binary tests for an optimal test.[18] Since I know already from Lemma 1 (a) and Proposition 3 (b) that an optimal binary test exists and how it looks like, I obtain the following result:

Proposition 4 (Optimal test) *Any participation threshold $s \in \mathcal{S}$ is optimally induced by a binary test that is not subject to false positives, $T^{NFP}(\rho(s))$. If $s \in \mathcal{S}_a$, then the accurate test*

[18] Kamenica and Gentzkow (2011) show in Proposition 4 of their Web Appendix that in their persuasion problem, in that the designer is only subject to Bayesian plausibility, the optimal information structure requires at most as many different signals as there exist states of the world. My Proposition 4 shows that a similar result is obtained for my test design problem with an interim participation constraint.

$T^{NFP}(1)$ is optimal. If $s \in \mathcal{S}\backslash\mathcal{S}_a$, then the optimal test accuracy $\rho(s)$ is implicitly defined by the unique solution to $\mathcal{U}_T^s(0, \rho(s); \theta_Y(s)) = u(\theta_N(s))$.

My derivation of the optimal test structure gives rise to the following interpretation of how any participation threshold is optimally induced. I can interpret any binary test with $p_2^b < p_2^g$ as a pass-fail test where the test result $\sigma = 1$ ($\sigma = 2$) corresponds to "fail" ("pass"). If the test is accurate, the agent passes if, and only if, he is good; if the test is subject to FP, the agent does sometimes pass even when he is bad; if the test is subject to FN, the agent does sometimes fail although he is good. The quality perception associated to failing determines the stigma of failure. If it becomes relatively more likely that an agent who fails is good, the stigma of failure becomes less severe. I obtain the following result.

Corollary 1 (Properties of the optimal test) *Consider any $s \in \mathcal{S}\backslash\mathcal{S}_a$. (a) Among all tests that induce s, the expected perception of the threshold agent is lowest for the optimal test. (b) Among all binary tests that induce s, the stigma of failure is least severe for the optimal test.*

This corollary allows for the conclusion that it is better to induce participation by lowering the stigma of failure than by increasing the expected perception of the threshold agent. Harbaugh and Rasmusen (2014) consider a related problem in that the agent is perception risk-neutral. The principal's only instrument to foster participation is to increase the expected perception of the threshold agent. The corollary shows that when the agent is imperfectly informed and perception risk-averse, non-trivial effects are added. The principal can then also foster participation by reducing the threshold agent's perception risk and doing so is better for her than increasing the threshold agent's expected perception.

Remark 5 (Robustness) Because Proposition 4 arises through the repeated application of Proposition 3, Remarks 1, 2 and 3 on the robustness of this proposition apply also to Proposition 4. Remark 4 introduces limits to information generation. While the idea behind this remark stays also valid, the in this section proposed iteration procedure may get "stuck" and has thus to be adapted. The same comment applies to the modifications that I propose in Sections 2 and 3 of the supplementary material. In a previous working paper version of this article (Rosar, 2014), I demonstrate the adaptation of the iteration procedure for the extension that I propose in Section 2 of the supplementary material. I obtain from this that a possibly non-binary generalization of a NFP test is always optimal.

5.5. The optimal participation behavior

I know from my analysis so far that only threshold participation behavior can arise (Proposition 1), which participation thresholds $s \in \mathcal{S}$ can be induced (Proposition 2), and that any participation threshold $s \in \mathcal{S}$ is optimally induced by the test $T^{\text{NFP}}(\rho(s))$ (Proposition 4). If the principal wants to induce full participation for reasons that are not modeled here explicitly, the test $T^{\text{NFP}}(\rho(\underline{\theta}))$ is optimal. If not, it remains to determine which participation threshold it

is optimal for the principal to induce. My analysis so far reduces this problem to the following problem PART:

$$\max_{s\in\mathcal{S}} \quad (1-F(s))\left[p^{\theta_Y(s)}(1,1-\rho(s))v(\mu(1,1-\rho(s);\theta_Y(s))) + p^{\theta_Y(s)}(0,\rho(s))v(1)\right] \\ +F(s)v(\theta_N(s))$$

Although the structure of the optimal test is quite simple, the effect of the participation threshold s is complex. It determines simultaneously when the agent participates, with which probabilities the different test results are generated, and which inferences are drawn from the different possible observations. As PART is a one-dimensional problem, it can, however, easily be solved numerically for any primitives of the model. In Section 4 of the supplementary material, I discuss the optimal participation threshold in the uniform-quadratic case.

6. A non-reduced problem: Quality estimation with asymmetric error cost

In Section 4, I motivated my reduced-form problem with a non-reduced quality estimation problem. In this problem, FP and FN caused symmetric error cost for the principal. I relax now the symmetry assumption. Suppose that the principal's non-reduced utility function is

$$\widehat{v}(y,\omega) \equiv \begin{cases} -\alpha^{\mathrm{FN}}(y-1)^2 & \text{if } \omega = g \\ -\alpha^{\mathrm{FP}}(y-0)^2 & \text{if } \omega = b \end{cases}$$

with $\alpha^{\mathrm{FN}}, \alpha^{\mathrm{FP}} > 0$ and $\alpha^{\mathrm{FP}} \neq \alpha^{\mathrm{FN}}$. As before, y describes the principal's quality estimate. If $\alpha^{\mathrm{FP}} > \alpha^{\mathrm{FN}}$, the principal suffers more from too high estimates. This can be interpreted as FP being more costly for the principal than FN; conversely, if $\alpha^{\mathrm{FP}} < \alpha^{\mathrm{FN}}$, FN are more costly.[19] Everything else remains as in Section 4.

When the principal believes that the agent is good with probability μ, she chooses her estimate $y \in [0,1]$ to maximize her interim expected utility $\widehat{V}(y;\mu) \equiv -\mu\alpha^{\mathrm{FN}}(y-1)^2 - (1-\mu)\alpha^{\mathrm{FP}}(y-0)^2$. The optimal estimate is $y(\mu) \equiv \alpha^{\mathrm{FN}}\mu/(\alpha^{\mathrm{FN}}\mu + \alpha^{\mathrm{FP}}(1-\mu))$. While for symmetric error cost the principal's optimal estimate corresponds to the Bayesian estimate of the probability that the agent is good μ, she is biased towards lower estimates if FP are more costly and towards higher estimates if FN are more costly. The bias determines the curvature of $y(\mu)$; $y(\mu)$ is convex (concave) if FP (FN) are more costly. The optimal estimate implies the reduced-form utility function $v(\mu) \equiv \widehat{V}(y(\mu);\mu) = -\alpha^{\mathrm{FN}}\alpha^{\mathrm{FP}}\mu(1-\mu)/(\alpha^{\mathrm{FN}}\mu + \alpha^{\mathrm{FP}}(1-\mu))$ for the principal and $u(\mu) \equiv \widehat{u}(y(\mu))$ for the agent. The subsequent lemma states important properties of the optimal estimate and the reduced-form utility functions.

Lemma 5 (Properties of estimate and reduced-form utility) *Suppose that $\alpha^{FP}, \alpha^{FN} > 0$ with $\alpha^{FP} \neq \alpha^{FN}$. (a) $y' > 0$; $y'' > 0$ if $\alpha^{FP} > \alpha^{FN}$ and $y'' < 0$ if $\alpha^{FP} < \alpha^{FN}$; $y''' > 0$. (b) $v'' > 0$; $v''' > 0$ if $\alpha^{FP} > \alpha^{FN}$ and $v''' < 0$ if $\alpha^{FP} < \alpha^{FN}$. (c) $u' > 0$; if either $\alpha^{FP} < \alpha^{FN}$*

[19] In Perez-Richet (2014), the receiver faces a related decision problem. His decision is, however, binary and it is the perfectly informed sender who designs the information generation technology.

or if $\alpha^{FP} > \alpha^{FN}$ and $\widehat{u}$ is a CARA or a CRRA utility function that exhibits sufficiently strong risk-aversion, then $u'' < 0$ and $u''' > 0$; if $\alpha^{FP} > \alpha^{FN}$ and $\widehat{u}$ is a CARA utility function that exhibits sufficiently weak risk-aversion, then $u'' > 0$.

Two properties of the principal's reduced-form utility function are important. First, she always benefits from additional information ($v'' > 0$); second, as I allow for the cases where too high and where too low estimates are more costly, it is not very surprising that v''' can assume either sign. More surprisingly, a NFP test turns out to be optimal in both cases.

Proposition 5 (Optimal test; asymmetric error cost) *(a) Suppose that $\alpha^{FP} > \alpha^{FN}$. If $\widehat{u}$ is either a CARA or a CRRA utility function that exhibits sufficiently strong risk-aversion, then a NFP test $T^{NFP}(\rho)$ is optimal. If $\widehat{u}$ is a CARA utility function that exhibits sufficiently weak risk-aversion, then the degenerate NFP test $T^{NFP}(1)$ is optimal. (b) Suppose that $\alpha^{FP} < \alpha^{FN}$. Then, some NFP test $T^{NFP}(\rho)$ is optimal.*

Consider first the case where FP are more costly ($\alpha^{\mathrm{FP}} > \alpha^{\mathrm{FN}}$). The principal's reduced-form utility function satisfies then the assumptions of my reduced model. However, the bias of the principal towards low estimates implies that the agent gets a quite unfavorable estimate unless he is able to convince the principal that he is very likely to be good. As a consequence, if the agent is risk-neutral in the decision y, he benefits from information-generation (i.e., u is convex). For sufficiently weak risk-aversion, this effect does still dominate; perfect information generation is then possible with an accurate test. On the other hand, for sufficiently strong risk-aversion, the results of my reduced model apply directly; a NFP test is optimal.

Consider next the case where FN are more costly ($\alpha^{\mathrm{FP}} < \alpha^{\mathrm{FN}}$). Since the principal is biased towards high estimates, the agent gets a quite favorable estimate unless the test reveals that he is very likely to be bad. This does even strengthen the agent's inherent aversion towards risk (i.e., $u = \widehat{u} \circ y$ is even more concave than $\widehat{u}$). Yet because $v''' < 0$ in this case, my sufficient condition for the optimality of a NFP test is violated. However, as I motivated in Remark 1, the assumption that $v''' \geq 0$ simplifies the proof of my main result, but it is not necessary. In particular, it is far from necessary when the agent is very averse towards perception risk. I demonstrate in the proof to Proposition 5 how I can use the specific structure that I imposed on the principal's utility function to extend the result in Propositions 3 and 4. It turns out that the additional perception risk-aversion that arises through the concavity of y renders a NFP optimal no matter how strong the agent's initial risk-aversion is.

7. Conclusions

I have studied optimal test design under voluntary participation of an imperfectly informed, perception risk-averse agent when the principal benefits from information. Only threshold participation behavior where the agent participates when his information is sufficiently favorable can be induced. For a large class of reduced-form utility functions and for any participation threshold that shall be induced, a binary test that 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, I have shown that a test that is not subject to false positives is optimal for a class of non-reduced problems where the principal has to give a probabilistic estimate of the agent's quality and suffers from asymmetric error cost. Interestingly, even when the principal suffers more from false negatives, a test that avoids false positives is optimal.

My main result can also be interpreted as a foundation for the use of simple testing procedures: the optimal test can be implemented as a pass-fail-test where the agent fails when he is bad and passes only sometimes when he is good. When I employ this interpretation, the optimal test is among all pass-fail-tests that induce a given participation threshold the one that is hardest to pass and that is associated with the least severe stigma of failure.

I find two extensions that bring the test design problem closer to mechanism design problems particularly interesting. In the first extension, tests can not only condition on the agent's quality, but also on announcements of him. For example, the agent may self-select into tests of different difficulty. Although learning will then still be imperfect, the extension allows it to fine-tune the structure of the generated information by exploiting the agent's signaling motive further.[20] In the second extension, the principal can use monetary transfers between participants and non-participants as an additional instrument to set participation incentives and—if I 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 that can be used to foster participation.[21]

Acknowledgements

Previous versions of this article circulated under the titles "Test design under voluntary participation and conflicting preferences", "Optimal test design under imperfect private information and voluntary participation", and "Imperfect private information and the design of information-generating mechanisms." It elaborates ideas partly originating from joint work with Elisabeth Schulte. Therefore, I am deeply indebted to Elisabeth. Furthermore, I would like to thank Kfir Eliaz, Hans Peter Grüner, Johannes Hörner, Emir Kamenica, Johannes Koenen, Benny Moldovanu, Tymofiy Mylovanov, Volker Nocke, Marco Ottaviani, Martin Peitz, Sergei Severinov, Konrad Stahl, Nora Szech, Tymon Tatur, and various seminar participants. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

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

[21] Such an extension would make the design problem more closely related to non-standard mechanism design problems that blend information design with mechanism design (e.g., Calzolari and Pavan (2006a,b), Bergemann and Pesendorfer (2007), Eső and Szentes (2007) and Pancs (2014)).

Appendix A. Proofs

Proof of Proposition 1

I prove first an auxiliary result:

Lemma A1 *Fix any $T \in \mathcal{T}$. Consider any participation strategy x and any (μ_N, μ_Y) that is consistent with T and x. Then, $U^\theta(T, \mu_Y)$ is continuous and strictly increasing in θ.*

Proof. I can write $U^\theta(T, \mu_Y) = (\sum_\sigma (p^g_\sigma - p^b_\sigma)u(\mu_\sigma))\theta + \sum_\sigma p^b_\sigma u(\mu_\sigma)$. Since $U^\theta(T, \mu_Y)$ is linear in θ, it is continuous in θ. To prove the lemma, I need only to show that the coefficient of θ is strictly positive. I can rewrite this coefficient in the following way:

$$\begin{aligned}
\sum\nolimits_\sigma (p^g_\sigma - p^b_\sigma)u(\mu_\sigma) &= \sum\nolimits_{\sigma'} (p^g_{\sigma'} \sum\nolimits_{\sigma''} p^b_{\sigma''} - p^b_{\sigma'} \sum\nolimits_{\sigma''} p^g_{\sigma''})u(\mu_{\sigma'}) \\
&= \sum\nolimits_{\sigma'} \sum\nolimits_{\sigma'' \neq \sigma'} (p^g_{\sigma'} p^b_{\sigma''} - p^b_{\sigma'} p^g_{\sigma''})u(\mu_{\sigma'}) \\
&= \sum\nolimits_{\sigma'} \sum\nolimits_{\sigma'' > \sigma'} (p^g_{\sigma'} p^b_{\sigma''} - p^b_{\sigma'} p^g_{\sigma''})(u(\mu_{\sigma'}) - u(\mu_{\sigma''})). \qquad \text{(A.1)}
\end{aligned}$$

By consistency of (μ_N, μ_Y) with T and x, there exists $\theta_Y \in [\underline{\theta}, \overline{\theta}]$ such that $\mu_\sigma = \theta_Y p^g_\sigma / (\theta_Y p^g_\sigma + (1 - \theta_Y)p^b_\sigma)$. It follows from this that the sign of $\mu_{\sigma'} - \mu_{\sigma''}$ corresponds to the sign of $p^g_{\sigma'} p^b_{\sigma''} - p^b_{\sigma'} p^g_{\sigma''}$. Moreover, since $u' > 0$, the sign of $u(\mu_{\sigma'}) - u(\mu_{\sigma''})$ corresponds to the sign of $p^g_{\sigma'} p^b_{\sigma''} - p^b_{\sigma'} p^g_{\sigma''}$. This implies that each summand in (A.1) is weakly positive. Since $\mathcal{T}$ includes only tests that are not completely inaccurate, $\mu_{\sigma'} \neq \mu_{\sigma''}$ must be true for some σ', σ''. Thus, there exists at least one summand in (A.1) that is strictly positive. This yields the result. q.e.d.

(a) That only threshold strategies can be part of an equilibrium is a direct consequence of Lemma A1 and the fact that $u(\mu_N)$ does not depend on θ.

The formula for the principal's expected payoff follows from (1) and three observations. First, consistency implies that $\theta_N = \theta_N(s)$ for any $s \in (\underline{\theta}, \overline{\theta}]$ and that $\theta_Y = \theta_Y(s)$ for any $s \in [\underline{\theta}, \overline{\theta})$. If $s = \underline{\theta}$, θ_N may differ in equilibrium from $\theta_N(s)$. However, if $s = \underline{\theta}$, the agent does not participate with probability zero such that the value of θ_N does not matter for the principal's expected payoff. Thus, I can set also in this case $\theta_N = \theta_N(s)$ without affecting the principal's expected payoff. An analogous reasoning applies for θ_Y if $s = \overline{\theta}$. Second, $\mathbb{E}_\theta[1 - x(\theta)] = F(s)$. Third, linearity of $p^\theta(p^b_\sigma, p^g_\sigma)$ in θ implies that $\mathbb{E}_\theta[x(\theta)p^\theta(p^b_\sigma, p^g_\sigma)] = (1 - F(s))p^{\theta_Y(s)}(p^b_\sigma, p^g_\sigma)$.

(b) Define $\mu_N(s) \equiv \mathbb{E}_\theta[\theta | \theta \leq s]$ and $\mu_Y(s) \equiv (\mu_1(s), \ldots, \mu_Z(s))$ with $\mu_\sigma(s) \equiv \mathbb{E}_\theta[\theta | \theta \geq s]p^g_\sigma / (\mathbb{E}_\theta[\theta | \theta \geq s]p^g_\sigma + (1 - \mathbb{E}_\theta[\theta | \theta \geq s])p^b_\sigma)$. Note that I have constructed $(\mu_N(s), \mu_Y(s))$ such that it is consistent with the given test T and any threshold strategy with threshold s. Thus, to conclude the proof, it suffices for me to argue that there exists a threshold participation strategy with some threshold $s \in [\underline{\theta}, \overline{\theta}]$ that is optimal for the given test T and the quality perception vector $(\mu_N(s), \mu_Y(s))$. I distinguish three cases.

Case 1: $U^s(T, \mu_Y(s)) \geq u(\mu_N(s))$ for $s = \underline{\theta}$. By the supposition, the agent with signal $\theta = \underline{\theta}$ has at least a weak incentive to participate. By Lemma A1, the agent with any signal $\theta > \underline{\theta}$ has a strict incentive to participate. Hence, the participation strategy $x(\theta) = 1$ is optimal given T and $(\mu_N(s), \mu_Y(s))$.

Case 2: $U^s(T, \mu_Y(s)) \leq u(\mu_N(s))$ for $s = \overline{\theta}$. By the supposition, the agent with $\theta = \overline{\theta}$ has at least a weak incentive not to participate. By Lemma A1, the agent with any signal $\theta < \overline{\theta}$ has a strict incentive not to participate. Hence, the participation strategy $x(\theta) = 0$ is optimal given T and $(\mu_N(s), \mu_Y(s))$.

Case 3: Neither the supposition in Case 1 nor the supposition in Case 2 holds. Continuity of F implies that $\mathbb{E}_\theta[\theta|\theta \leq s]$ and $\mathbb{E}_\theta[\theta|\theta \geq s]$ are continuous in s. Furthermore, continuity of u implies that $U^\theta(T, \mu_Y(s))$ and $u(\mu_N(s))$ are continuous in s. This and the supposition allows me to apply an Intermediate Value Theorem. I obtain that there must exist $s \in (\underline{\theta}, \overline{\theta})$ such that $U^s(T, \mu_Y(s)) = u(\mu_N(s))$; that is, the agent with private signal s is indifferent between participating and not participating. By Lemma A1, the agent has a strict incentive to participate for any signal $\theta > s$ and a strict incentive not to participate for any signal $\theta < s$. Hence, the participation strategy $x(\theta) = 1$ for $\theta \geq s$ and $x(\theta) = 0$ for $\theta < s$ is optimal given T and $(\mu_N(s), \mu_Y(s))$. q.e.d.

Proof of Proposition 2

(a) "$\Rightarrow$" Suppose that the test $T \in \mathcal{T}_a$ induces the threshold $s \in [\underline{\theta}, \overline{\theta})$.

I will first argue that $s \in (\underline{\theta}, \overline{\theta})$ must be true. Assume to the contrary that $s = \underline{\theta}$. Consistency implies for this case that $\mu_N = \theta_N$ with $\theta_N \in [\underline{\theta}, \overline{\theta}]$ and that $\mu_Y = \mu_Y(T; \theta_Y(\underline{\theta}))$. The threshold agent's payoff from non-participation is $u(\theta_N)$ whereas his expected payoff from participation in the accurate test T is $U^{\underline{\theta}}(T, \mu_Y(T; \theta_Y(\underline{\theta}))) = (1-\underline{\theta})u(0)+\underline{\theta}u(1)$. By strict concavity of u and Jensen's inequality, $U^{\underline{\theta}}(T, \mu_Y(T; \theta_Y(\underline{\theta}))) < u(\underline{\theta})$. Since this implies that the agent with signal $\theta = \underline{\theta}$ has a strict incentive not to participate in the test T, it follows from Lemma A1 in the proof to Proposition 1 that the agent has also for some signals $\theta > \underline{\theta}$ a strict incentive not to participate. Contradiction.

Consistency implies for any $s \in (\underline{\theta}, \overline{\theta})$ that $(\mu_N, \mu_Y) = (\theta_N(s), \mu_Y(T; \theta_Y(s)))$. Moreover, Lemma A1 in the proof to Proposition 1 implies that for any optimal participation strategy the participation constraint must be binding for the agent with signal $\theta = s$; that is, $U^s(T, \mu_Y(T; \theta_Y(s))) = u(\theta_N(s))$. By using that the test T is accurate and that $\theta_N(s) = \mathbb{E}_\theta[\theta|\theta \leq s]$, I can write this equality as $(1-s)u(0) + su(1) = u(\mathbb{E}_\theta[\theta|\theta \leq s])$.

"$\Leftarrow$" Suppose that $(1-s)u(0) + su(1) = u(\mathbb{E}_\theta[\theta|\theta \leq s])$ for some $s \in [\underline{\theta}, \overline{\theta})$. Consider any accurate test $T \in \mathcal{T}_a$, the participation strategy defined by $x(\theta) = 0$ if $\theta < s$ and $x(\theta) = 1$ if $\theta \geq s$, and the vector of quality perceptions $(\mu_N, \mu_Y) = (\theta_N(s), \mu_Y(T; \theta_Y(s)))$. It suffices for me to show that $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T)$ under the supposition. I have $U^s(T, \mu_Y) = (1-s)u(0)+su(1)$ and $u(\mu_N) = u(\mathbb{E}_\theta[\theta|\theta \leq s])$. Thus, by the supposition, the agent with private signal $\theta = s$ is indifferent between participation and non-participation. By Lemma A1 in the proof to Proposition 1, the agent has a strict incentive to participate for any signal $\theta > s$ and a strict incentive not to participate for any signal $\theta < s$. That is, x is optimal given T and (μ_N, μ_Y). Since (μ_N, μ_Y) is constructed such that it is consistent given T and x, I obtain $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T)$.

(b) "$\Rightarrow$" Suppose that the test $T \in \mathcal{T}$ induces the threshold $s \in [\underline{\theta}, \overline{\theta})$.

Case 1: $s = \underline{\theta}$. By Jensen's inequality and strict concavity of u, $(1-s)u(0) + su(1) < u(s)$. Since $s = \mathbb{E}_\theta[\theta|\theta \leq s]$ for $s = \underline{\theta}$, I obtain $(1-s)u(0) + su(1) < u(\mathbb{E}_\theta[\theta|\theta \leq s])$.

Case 2: $s \in (\underline{\theta}, \overline{\theta})$. Consistency implies then that $(\mu_N, \mu_Y) = (\theta_N(s), \mu_Y(T; \theta_Y(s)))$. Moreover, Lemma A1 in the proof to Proposition 1 implies that for any optimal participation strategy the participation constraint must be binding for the agent with signal s; that is, $U^s(T, \mu_Y(T; \theta_Y(s))) = u(\theta_N(s))$. Since the right-hand side corresponds to $u(\mathbb{E}_\theta[\theta|\theta \leq s])$, it suffices for me to show that $U^s(T, \mu_Y(T; \theta_Y(s))) \geq (1-s)u(0) + su(1)$. By using the definition of $U^s(T, \mu_Y(T; \theta_Y(s)))$ and applying then Jensen's inequality, I obtain

$$U^s(T, \mu_Y(T; \theta_Y(s)))$$

$$
\begin{aligned}
&= \sum\nolimits_{\sigma} p^s(p^b_\sigma, p^g_\sigma) u(\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s))) \\
&\geq \sum\nolimits_{\sigma} p^s(p^b_\sigma, p^g_\sigma)((1-\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s)))u(0) + \mu(p^b_\sigma, p^g_\sigma; \theta_Y(s))u(1)).
\end{aligned}
$$

Since $u(1) > u(0)$, I only need to show that the coefficient of $u(1)$, $\sum_\sigma p^s(p^b_\sigma, p^g_\sigma)\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s))$, is weakly larger than s. I can write

$$
\begin{aligned}
&\sum\nolimits_{\sigma} p^s(p^b_\sigma, p^g_\sigma)\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s)) - s \\
&= \sum\nolimits_{\sigma} p^s(p^b_\sigma, p^g_\sigma)\frac{\theta_Y(s)p^g_\sigma}{\theta_Y(s)p^g_\sigma + (1-\theta_Y(s))p^b_\sigma} - \sum\nolimits_{\sigma} s p^g_\sigma \\
&= \sum\nolimits_{\sigma} \frac{sp^g_\sigma + (1-s)p^b_\sigma}{\theta_Y(s)p^g_\sigma + (1-\theta_Y(s))p^b_\sigma}\theta_Y(s)p^g_\sigma - \sum\nolimits_{\sigma} \frac{\theta_Y(s)p^g_\sigma + (1-\theta_Y(s))p^b_\sigma}{\theta_Y(s)p^g_\sigma + (1-\theta_Y(s))p^b_\sigma} s p^g_\sigma \\
&= \sum\nolimits_{\sigma} \frac{\theta_Y(s) - s}{\theta_Y(s)p^g_\sigma + (1-\theta_Y(s))p^b_\sigma} p^b_\sigma p^g_\sigma. \qquad \text{(A.2)}
\end{aligned}
$$

Since $s < \theta_Y(s)$ for any $s \in (\underline{\theta}, \overline{\theta})$, the summand in (A.2) is weakly positive for all σ. This yields the result.

"$\Leftarrow$" Suppose that $(1-s)u(0)+su(1) \leq u(\mathbb{E}_\theta[\theta|\theta \leq s])$ for some $s \in [\underline{\theta}, \overline{\theta})$. 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 a completely inaccurate test as $\rho \to 0$. Examples for such functions are $T^{\text{NFP}}(\rho)$ and $T^{\text{NFN}}(\rho)$ as defined in Table 3. The introduced notation allows me to write the supposition as $U^s(T(1), \mu_Y(T(1); \theta_Y(s))) \leq u(\theta_N(s))$. Moreover, by construction, $\lim_{\rho\to 0} U^s(T(\rho), \mu_Y(T(\rho); \theta_Y(s))) = u(\theta_Y(s)) > u(\theta_N(s))$. These two inequalities and the fact that $U^s(T(\rho), \mu_Y(T(\rho); \theta_Y(s)))$ is continuous in ρ allow me to apply an Intermediate Value Theorem. It follows that there exists some $\rho^* \in (0,1)$ such that $U^s(T(\rho^*), \mu_Y(T(\rho^*); \theta_Y(s))) = u(\theta_N(s))$. By this and Lemma A1 in the proof to Proposition 1, any participation strategy with threshold s is optimal given the test $T(\rho^*)$ and the vector of quality perceptions $(\mu_N, \mu_Y) = (\theta_N(s), \mu_Y(T; \theta_Y(s)))$. Since $(\mu_N, \mu_Y) = (\theta_N(s), \mu_Y(T; \theta_Y(s)))$ is consistent given the test $T(\rho^*)$ and any participation strategy with threshold s, I obtain $(x, (\mu_N, \mu_Y)) \in \mathcal{E}(T(\rho^*))$. q.e.d.

Proof of Lemma 1

(a) Suppose that $s \in \mathcal{S}_a$. Any accurate test induces then s. Because updating is Bayesian, any test that induces s induces the same expected quality perception from the principal's perspective; it can easily be checked that $\sum_\sigma p^{\theta_Y(s)}(p^b_\sigma, p^g_\sigma)\mu(p^b_\sigma, p^g_\sigma; \theta_Y(s)) = \theta_Y(s)$ for any (p^b, p^g). Thus, from the principal's perspective, the quality perception distribution induced by any accurate test is a mean-preserving spread of the quality perception distribution that is induced by any other test that induces s. It follows directly from this and strict convexity of v that (2) is maximized by any accurate test.

(b) Suppose that $s \in \mathcal{S}\backslash\mathcal{S}_a$. Since $(1-F(s)) > 0$ and since $F(s)v(\theta_N(s))$ is constant for given s, I obtain that the objective function in the problem $\mathsf{TD}(s)$ is a positive linear transformation of (2). I distinguish now two cases. Suppose first that $s \in (\underline{\theta}, \overline{\theta})$. The consistency requirement (EQ1) implies then that $\mu_Y = \mu_Y(T; \theta_Y(s))$ and that $\mu_N = \theta_N(s)$. It follows directly from this and Lemma A1 in the proof to Proposition 1, that (EQ1) is satisfied for all θ if, and only if, $U^s(T, \mu_Y(T; \theta_Y(s))) = u(\theta_N(s))$. Since this is just the constraint in problem $\mathsf{TD}(s)$, I am done. Suppose next that $s = \underline{\theta}$. There are then two differences relative to the preceding case. First, any $\mu_N \in [\underline{\theta}, \overline{\theta}]$ satisfies the consistency requirement (EQ2). Second, necessary and sufficient

for (EQ1) is that $U^s(T, \mu_Y(T;\theta_Y(s))) \geq u(\mu_N)$. Since this constraint is satisfied for $\mu_N = \underline{\theta}$ if it is satisfied for some $\mu_N \in (\underline{\theta}, \overline{\theta}]$, it is without loss of generality to assume that $\mu_N = \underline{\theta}$ (which corresponds to $\theta_N(s)$ with $s = \underline{\theta}$). Hence, the participation (EQ1) is satisfied if, and only if, $U^s(T, \mu_Y(T;\theta_Y(s))) \geq u(\theta_N(s))$. It remains for me to argue why it is without loss of generality to assume that the participation constraint is binding. Assume to the contrary that the problem $\mathsf{TD}(s)$ is solved by a test T' with $U^s(T', \mu_Y(T;\theta_Y(s))) > u(\theta_N(s))$. It is then possible to construct a test $T''(\rho)$ in the spirit of the test described in Table 2 that is strictly better for the principal and for which $U^s(T''(\rho), \mu_Y(T;\theta_Y(s))) = u(\theta_N(s))$ is true: with probability ρ, the test $T''(\rho)$ does perfectly reveal the agent's quality; with probability $1-\rho$, the test T' is executed. Since the supposition that $s \in \mathcal{S}\backslash\mathcal{S}_a$ implies that $U^s(T''(1), \mu_Y(T;\theta_Y(s))) < \mu_N(\theta_N(s))$ and since $U^s(T''(0), \mu_Y(T;\theta_Y(s))) > u(\mu_Y(s))$ by assumption, I obtain a test that has the desired properties by a continuity property and an Intermediate Value Theorem. This yields the result. q.e.d.

Proof of Lemma 2

Take first any $p_2^b, p_2^g \in [0,1]$ with $p_2^b < p_2^g$ as given. It can be easily verified that $p_2^b < p_2^g$ implies that $\mu(1-p_2^b, 1-p_2^g;\theta_Y) \in [0,\theta_Y)$ and that $\mu(p_2^b, p_2^g;\theta_Y) \in (\theta_Y, 1]$. This yields the first part of the result. Take next any $(\mu_1,\mu_2) \in \mathcal{Q}(\theta_Y)$ as given. Then, $\mu_2 = \mu(p_2^b, p_2^g;\theta_Y)$ implies $p_2^g = \mu_2/(1-\mu_2)\cdot(1-\theta_Y)/\theta_Y \cdot p_2^b$ and $\mu_1 = \mu(1-p_2^b, 1-p_2^g;\theta_Y)$ implies $p_2^g = (\theta_Y - \mu_1)/(\theta_Y(1-\mu_1)) + \mu_1/(1-\mu_1)\cdot(1-\theta_Y)/\theta_Y \cdot p_2^b$. Since the so obtained equations are linear in p_2^b and since $\mu_1 < \mu_2$ implies that they have a different slope, there exists at most one binary test for that both equations hold simultaneously. It can easily be verified that both equations hold simultaneously for $p_2^b = \beta^b(\mu_1,\mu_2;\theta_Y)$ and $p_2^g = \beta^g(\mu_1,\mu_2;\theta_Y)$. Moreover, $\mu_2 > \theta_Y$ implies that $p_2^b < p_2^g$. This yields the second part of the result. q.e.d.

Proof of Lemma 3

(a) and (b) By using (3) in (4), I obtain

$$\overline{\mu}^{\theta}(\mu_1,\mu_2;\theta_Y) = \mu_1 + \frac{\theta_Y-\mu_1}{\mu_2-\mu_1}\left((1-\theta)\frac{1-\mu_2}{1-\theta_Y} + \theta\frac{\mu_2}{\theta_Y}\right)(\mu_2-\mu_1),$$

which I can rewrite as

$$\overline{\mu}^{\theta}(\mu_1,\mu_2;\theta_Y) = \frac{1-\theta}{1-\theta_Y}\theta_Y - \frac{\theta_Y-\theta}{1-\theta_Y}\left(\mu_1+\mu_2-\frac{\mu_1\mu_2}{\theta_Y}\right). \tag{A.3}$$

From this I directly obtain Part (a): $\overline{\mu}^{\theta_Y}(\mu_1,\mu_2;\theta_Y) = \theta_Y$. Then, by differentiating (A.3) partially with respect to μ_1 and μ_2, I obtain

$$\frac{\partial\overline{\mu}^{\theta}(\mu_1,\mu_2;\theta_Y)}{\partial\mu_1} = \frac{\theta_Y-\theta}{1-\theta_Y}\cdot\frac{\mu_2-\theta_Y}{\theta_Y}, \text{ and} \tag{A.4}$$

$$\frac{\partial\overline{\mu}^{\theta}(\mu_1,\mu_2;\theta_Y)}{\partial\mu_2} = -\frac{\theta_Y-\theta}{1-\theta_Y}\cdot\frac{\theta_Y-\mu_1}{\theta_Y}. \tag{A.5}$$

Part (b) follows from (A.4) and (A.5) with $\theta = s$, the fact that $(\mu_1,\mu_2) \in \mathcal{Q}(\theta_Y)$ implies that $\mu_1 < \theta_Y < \mu_2$, and the assumption that $s < \theta_Y$.

(c) By differentiating $\mathcal{V}_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)$ partially with respect to μ_1 and μ_2, and by using that

$\overline{\mu}^{\theta_Y}(\mu_1,\mu_2;\theta_Y)=\theta_Y$ by Part (a), I obtain

$$\begin{aligned}\frac{\partial \mathcal{V}_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)}{\partial \mu_1} &= -\frac{\mu_2-\theta_Y}{\mu_2-\mu_1}\left[\frac{v(\mu_2)-v(\mu_1)}{\mu_2-\mu_1}-v'(\mu_1)\right], \text{ and}\\ \frac{\partial \mathcal{V}_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)}{\partial \mu_2} &= \frac{\theta_Y-\mu_1}{\mu_2-\mu_1}\left[v'(\mu_2)-\frac{v(\mu_2)-v(\mu_1)}{\mu_2-\mu_1}\right].\end{aligned}$$

For any $(\mu_1,\mu_2)\in\mathcal{Q}(\theta_Y)$ the two quotients in front of the expressions in brackets are strictly positive. Since strict convexity of $v(\mu)$ implies that also the expressions in the brackets are strictly positive, I obtain the result in Part (c).

(d) By differentiating $\mathcal{U}^s_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)$ partially with respect to μ_1 and μ_2, and by simplifying then, I obtain

$$\begin{aligned}\frac{\partial \mathcal{U}^s_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)}{\partial \mu_1} &= \frac{\mu_2-\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)}{\mu_2-\mu_1}\left[u'(\mu_1)-\frac{u(\mu_2)-u(\mu_1)}{\mu_2-\mu_1}\right]\\ &\quad+\frac{\partial \overline{\mu}^s(\mu_1,\mu_2;\theta_Y)}{\partial \mu_1}\frac{u(\mu_2)-u(\mu_1)}{\mu_2-\mu_1}, \text{ and}\\ \frac{\partial \mathcal{U}^s_{\mathcal{Q}}(\mu_1,\mu_2;\theta_Y)}{\partial \mu_2} &= -\frac{\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)-\mu_1}{\mu_2-\mu_1}\left[\frac{u(\mu_2)-u(\mu_1)}{\mu_2-\mu_1}-u'(\mu_2)\right]\\ &\quad+\frac{\partial \overline{\mu}^s(\mu_1,\mu_2;\theta_Y)}{\partial \mu_2}\frac{u(\mu_2)-u(\mu_1)}{\mu_2-\mu_1}.\end{aligned}$$

Note that $\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)$ is a convex combination of μ_1 and μ_2. Thus, $\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)\in[\mu_1,\mu_2]$. This renders the two quotients in front of the expressions in brackets weakly positive. Since $u''<0$ implies that also the expressions in the brackets are non-negative, sufficient for the effects in the lemma are $(\frac{\partial}{\partial\mu_1}\overline{\mu}^s(\mu_1,\mu_2;\theta_Y))\cdot(u(\mu_2)-u(\mu_1))/(\mu_2-\mu_1)>0$ and $(\frac{\partial}{\partial\mu_2}\overline{\mu}^s(\mu_1,\mu_2;\theta_Y))\cdot(u(\mu_2)-u(\mu_1))/(\mu_2-\mu_1)<0$. This follows from Part (b) and $u'>0$. q.e.d.

Proof of Lemma 4
(a) Suppose that $(\mu_1',\mu_2'),(\mu_1'',\mu_2'')\in\overline{\mathcal{Q}}^s(\mu_1,\mu_2;\theta_Y)$. Then, I have $\overline{\mu}^s(\mu_1',\mu_2';\theta_Y)=\overline{\mu}^s(\mu_1'',\mu_2'';\theta_Y)$ by construction and $\overline{\mu}^{\theta_Y}(\mu_1',\mu_2';\theta_Y)=\overline{\mu}^{\theta_Y}(\mu_1'',\mu_2'';\theta_Y)=\theta_Y$ by Lemma 3 (a); that is, I have $\overline{\mu}^{\theta}(\mu_1',\mu_2';\theta_Y)=\overline{\mu}^{\theta}(\mu_1'',\mu_2'';\theta_Y)$ for two different values of θ. Since $\overline{\mu}^{\theta}(\mu_1,\mu_2;\theta_Y)$ is linear in θ by (4), it follows that $\overline{\mu}^{\theta}(\mu_1',\mu_2';\theta_Y)=\overline{\mu}^{\theta}(\mu_1'',\mu_2'';\theta_Y)$ for any value of θ. This is Part (a) of the result.

Before I come to the proof of (b) and (c), I prove two auxiliary results.

Lemma A2 *Suppose that $h:[0,1]\to\mathbb{R}$ is a smooth function with $h'''\geq 0$. Consider $m_1,m_2\in[0,1]$ with $m_2>m_1$. Then, $(h'(m_1)+h'(m_2))-2(h(m_2)-h(m_1))/(m_2-m_1)\geq 0$. If $h'''>0$, the inequality is strict.*

Proof. I can make the following transformations:

$$\begin{aligned}&(h'(m_1)+h'(m_2))(m_2-m_1)-2(h(m_2)-h(m_1))\\ =\;&[(h'(m_1)+h'(m_1+e_1))e_1-2(h(m_1+e_1)-h(m_1))]_{e_1=0}^{e_1=m_2-m_1}\\ =\;&\int_0^{m_2-m_1}\frac{\mathrm{d}}{\mathrm{d}e_1}\left[(h'(m_1)+h'(m_1+e_1))e_1-2(h(m_1+e_1)-h(m_1))\right]\mathrm{d}e_1\\ =\;&\int_0^{m_2-m_1}\left[h''(m_1+e_1)e_1-(h'(m_1+e_1)-h'(m_1))\right]\mathrm{d}e_1\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{m_2-m_1} \left[h''(m_1+e_2)e_2 - h'(m_1+e_2)\right]_{e_2=0}^{e_2=e_1} \mathrm{d}e_1 \\
&= \int_0^{m_2-m_1} \int_0^{e_1} \frac{\mathrm{d}}{\mathrm{d}e_2}\left[h''(m_1+e_2)e_2 - h'(m_1+e_2)\right] \mathrm{d}e_2 \mathrm{d}e_1 \\
&= \int_0^{m_2-m_1} \int_0^{e_1} \left[h'''(m_1+e_2)e_2\right] \mathrm{d}e_2 \mathrm{d}e_1.
\end{aligned}
$$

Since $h''' \geq 0$ ($h''' > 0$) implies that this expression is weakly (strictly) positive, I obtain the result. q.e.d.

Lemma A3 *Suppose that $s \in [0,1)$, that $\theta_Y \in (s,1)$, and that $(\mu_1', \mu_2') \in \mathcal{Q}(\theta_Y)$. Then,*

$$\overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y) = \{(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y) | \mu_1 \in [0, \overline{\mu}_1], \mu_2 = \mu_2(\mu_1)\}$$

for some onto function $\mu_2 : [0, \overline{\mu}_1] \to [\underline{\mu}_2, 1]$ with $\overline{\mu}_1 \in [0, \theta_Y)$ and $\underline{\mu}_2 \in (\theta_Y, 1]$. Moreover, $\mu_2(\mu_1)$ is strictly increasing and continuously differentiable with $\mu_2'(\mu_1) = (\mu_2(\mu_1) - \theta_Y)/(\theta_Y - \mu_1)$.

Proof. I first prove three properties:

Property 1: (a) There exists $\underline{\mu}_2 \in (\theta_Y, 1]$ such that $(0, \underline{\mu}_2) \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. (b) There exists $\overline{\mu}_1 \in [0, \theta_Y)$ such that $(\overline{\mu}_1, 1) \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. By (A.3), I have

$$\overline{\mu}^s(0, \mu_2; \theta_Y) = \frac{1-s}{1-\theta_Y}\theta_Y - \frac{\theta_Y - s}{1-\theta_Y}\mu_2 = \frac{1-\mu_2}{1-\theta_Y}\theta_Y + \frac{\mu_2 - \theta_Y}{1-\theta_Y}s.$$

This expression is linear in μ_2 with $\overline{\mu}^s(0,1;\theta_Y) = s$ and $\lim_{\mu_2 \downarrow \theta_Y} \overline{\mu}^s(0, \mu_2; \theta_Y) = \theta_Y$. Since $\overline{\mu}^s(\mu_1, \mu_2; \theta_Y) \in [s, \theta_Y)$ must be true, there exists by an Intermediate Value Theorem a unique $\mu_2 \in (\theta_Y, 1]$ that solves $\overline{\mu}^s(0, \mu_2; \theta_Y) = \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$. This implies Part (a) of Property 1. Part (b) arises analogously.

Property 2: If $(\mu_1'', \mu_2''), (\mu_1''', \mu_2''') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ with $\mu_1'' < \mu_1'''$, then $\mu_2'' < \mu_2'''$. This is a direct consequence of Lemma 3 (b).

Property 3: (a) For any $\mu_1'' \in [0, \overline{\mu}_1]$ there exists a unique $\mu_2'' \in [\underline{\mu}_2, 1]$ such that $(\mu_1'', \mu_2'') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. (b) For any $\mu_2'' \in [\underline{\mu}_2, 1]$ there exists a unique $\mu_1'' \in [0, \overline{\mu}_1]$ such that $(\mu_1'', \mu_2'') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. Fix any $\mu_1'' \in [0, \overline{\mu}_1]$. Since $\overline{\mu}^s(\mu_1'', \mu_2; \theta_Y)$ is continuous in μ_2 by (A.3) and strictly decreasing in μ_2 by Lemma 3 (b), there exists a unique $\mu_2'' \in [\underline{\mu}_2, 1]$ such that $\overline{\mu}^s(\mu_1'', \mu_2''; \theta_Y) = \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$ if, and only if, $\overline{\mu}^s(\mu_1'', 1; \theta_Y) \leq \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$ and $\overline{\mu}^s(\mu_1'', \underline{\mu}_2; \theta_Y) \geq \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$. First note that by Property 1 (b), $\overline{\mu}^s(\overline{\mu}_1, 1; \theta_Y) = \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$. Since Lemma 3 (b) implies that $\overline{\mu}^s(\mu_1'', 1; \theta_Y) \leq \overline{\mu}^s(\overline{\mu}_1, 1; \theta_Y)$, I obtain that the first condition holds. Note next that by Property 1 (a), $\overline{\mu}^s(0, \underline{\mu}_2; \theta_Y) = \overline{\mu}^s(\mu_1', \mu_2'; \theta_Y)$. Since Lemma 3 (b) implies that $\overline{\mu}^s(\mu_1'', \underline{\mu}_2; \theta_Y) > \overline{\mu}^s(0, \underline{\mu}_2; \theta_Y)$, I obtain that also the second condition holds. Thus, I obtain Part (a) of Property 3. Part (b) arises analogously.

It follows directly from Properties 2 and 3 that there exists a continuous and increasing function $\mu_2 : [0, \overline{\mu}_1] \to [\underline{\mu}_2, 1]$ such that $(\mu_1'', \mu_2'') \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ if, and only if, $\mu_2'' = \mu_2(\mu_1'')$. Since $\overline{\mu}^s(\mu_1, \mu_2; \theta_Y)$ is continuously differentiable with non-zero partials, I obtain by the Implicit Function Theorem that $\mu_2(\mu_1)$ is continuously differentiable with

$$\mu_2'(\mu_1) = -\frac{\partial \overline{\mu}^s(\mu_1, \mu_2(\mu_1); \theta_Y)}{\partial \mu_1} \Big/ \frac{\partial \overline{\mu}^s(\mu_1, \mu_2(\mu_1); \theta_Y)}{\partial \mu_2}.$$

By (A.4) and (A.5), $\mu_2'(\mu_1) = (\mu_2(\mu_1) - \theta_Y)/(\theta_Y - \mu_1)$. q.e.d.

The proof of the part of Lemma 4 (b) that concerns the threshold agent. By using that

$$\frac{\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)-\mu_1}{\mu_2-\mu_1} = \frac{\theta_Y-\mu_1}{\mu_2-\mu_1}-\frac{\theta_Y-\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)}{\mu_2-\mu_1},$$

I can write

$$\begin{aligned}\mathcal{U}^s_{\mathcal{Q}}(\mu_1,\mu_2(\mu_1);\theta_Y) &= u(\mu_1)+\frac{\theta_Y-\mu_1}{\mu_2(\mu_1)-\mu_1}(u(\mu_2(\mu_1))-u(\mu_1)) \\ &\quad -(\theta_Y-\overline{\mu}^s(\mu_1,\mu_2(\mu_1);\theta_Y))\frac{u(\mu_2(\mu_1))-u(\mu_1)}{\mu_2(\mu_1)-\mu_1}. \qquad \text{(A.6)}\end{aligned}$$

Since $\mu_2(\mu_1)$ is strictly increasing, I need to show that this expression is strictly increasing in μ_1. I will prove this by showing that the expression in the second (first) line is strictly (weakly) increasing in μ_1.

Effect of μ_1 on the second line of (A.6). By strict concavity of u, $(u(\mu_2)-u(\mu_1))/(\mu_2-\mu_1)$ is strictly decreasing in μ_1 and in μ_2. Since $\mu_2(\mu_1)$ is strictly increasing by Lemma A3, also $(u(\mu_2(\mu_1))-u(\mu_1))/(\mu_2(\mu_1)-\mu_1)$ is strictly decreasing in μ_1. Furthermore, since $\overline{\mu}^\theta(\mu_1,\mu_2;\theta_Y)$ is strictly increasing in θ and since $\overline{\mu}^{\theta_Y}(\mu_1,\mu_2;\theta_Y)=\theta_Y$ by Lemma 3 (a), $\theta_Y-\overline{\mu}^s(\mu_1,\mu_2;\theta_Y)>0$. Finally, since $\theta_Y-\overline{\mu}^s(\mu_1,\mu_2(\mu_1);\theta_Y)$ is constant by the supposition, I obtain the result for the second line.

Effect of μ_1 on the first line of (A.6). I have

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\mu_1}\left(u(\mu_1)+\frac{\theta_Y-\mu_1}{\mu_2(\mu_1)-\mu_1}(u(\mu_2(\mu_1))-u(\mu_1))\right) \\ =\ &\frac{\mu_2(\mu_1)-\theta_Y}{\mu_2(\mu_1)-\mu_1}\left[u'(\mu_1)-\frac{u(\mu_2(\mu_1))-u(\mu_1)}{\mu_2(\mu_1)-\mu_1}\right] \\ &+\mu_2'(\mu_1)\frac{\theta_Y-\mu_1}{\mu_2(\mu_1)-\mu_1}\left[u'(\mu_2(\mu_1))-\frac{u(\mu_2(\mu_1))-u(\mu_1)}{\mu_2(\mu_1)-\mu_1}\right] \\ =\ &\frac{\mu_2(\mu_1)-\theta_Y}{\mu_2(\mu_1)-\mu_1}\left[u'(\mu_1)+u'(\mu_2(\mu_1))-2\frac{u(\mu_2(\mu_1))-u(\mu_1)}{\mu_2-\mu_1}\right].\end{aligned}$$

The transformations arise as follows. The first equality follows from differentiating. The second equality follows from using that $\mu_2'(\mu_1)=(\mu_2(\mu_1)-\theta_Y)/(\theta_Y-\mu_1)$ by Lemma A3, and from simplifying. Since $u'''\geq 0$, the obtained expression is non-negative by Lemma A2 with $h=u$.

The proof of the part of Lemma 4 (b) that concerns the principal. Since $\overline{\mu}^{\theta_Y}(\mu_1,\mu_2;\theta_Y)=\theta_Y$ by Lemma 3 (a),

$$\mathcal{V}_{\mathcal{Q}}(\mu_1,\mu_2(\mu_1);\theta_Y) = v(\mu_1)+\frac{\theta_Y-\mu_1}{\mu_2(\mu_1)-\mu_1}(v(\mu_2(\mu_1))-v(\mu_1)).$$

By a similar reasoning as for the effect of μ_1 on the first line of (A.6), I obtain

$$\frac{\mathrm{d}}{\mathrm{d}\mu_1}\mathcal{V}_{\mathcal{Q}}(\mu_1,\mu_2(\mu_1);\theta_Y)=\frac{\mu_2(\mu_1)-\theta_Y}{\mu_2(\mu_1)-\mu_1}\left[v'(\mu_1)+v'(\mu_2(\mu_1))-2\frac{v(\mu_2(\mu_1))-v(\mu_1)}{\mu_2(\mu_1)-\mu_1}\right].$$

By Lemma A2 with $h=v$, this expression is weakly (strictly) positive if $v'''\geq 0$ ($v'''>0$).

First statement in (c): This is a direct consequence of Lemma A3.

Second statement in (c): By Lemma 3 (d), $\mathcal{U}^s_{\mathcal{Q}}(\mu_1',\mu_2'';\theta_Y)<\mathcal{U}^s_{\mathcal{Q}}(\mu_1',\mu_2';\theta_Y)$. By Part (b) of

this lemma, $\mathcal{U}^s_{\mathcal{Q}}(\mu_1'', \mu_2''; \theta_Y) > \mathcal{U}^s_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. Thus, by a continuity property and an Intermediate Value Theorem, there exists a value $\mu_1''' \in (\mu_1', \mu_1'')$ such that $\mathcal{U}^s_{\mathcal{Q}}(\mu_1''', \mu_2''; \theta_Y) = \mathcal{U}^s_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. Since $\mathcal{U}^s_{\mathcal{Q}}(\mu_1, \mu_2''; \theta_Y)$ is by Lemma 3 (d) strictly increasing in μ_1, this value is the only value μ_1 from $[0, \theta_Y)$ for that this is the case. This yields the result. q.e.d.

Proof of Proposition 3.

(a) Consider any participation constraint of the threshold agent $\mathcal{U}^s_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y) = \overline{u}$. It follows from the assumption that $\overline{u} \in \overline{\mathcal{U}}(\theta_Y)$ that there exists some quality perception pair that satisfies this constraint and it follows from Lemma 4 (c) that there exists a unique quality perception pair (μ_1, μ_2) with $\mu_2 = 1$ that satisfies this constraint, say $(\mu_1''', 1)$. I prove Part (a) of this proposition by showing that any other quality perception pair that satisfies this constraint, say (μ_1', μ_2'), is strictly worse for the principal.

Let (μ_1', μ_2') be any quality perception pair with $\mu_2' < 1$ such that $\mathcal{U}^s_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y) = \overline{u}$. By Lemma 4 (c) there exists a unique quality perception pair $(\mu_1'', 1) \in \overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$. Since also $\mu_1'' > \mu_1'$ by this lemma, it follows from Lemma 4 (b) that

$$\mathcal{V}_{\mathcal{Q}}(\mu_1'', 1; \theta_Y) \geq \mathcal{V}_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y). \tag{A.7}$$

By Lemma 4 (c), there exists further a unique quality perception pair $(\mu_1''', 1)$ such that $\mathcal{U}^s_{\mathcal{Q}}(\mu_1''', 1; \theta_Y) = \overline{u}$. Since also $\mu_1''' < \mu_1''$ by this lemma, I obtain by Lemma 3 (c) that

$$\mathcal{V}_{\mathcal{Q}}(\mu_1''', 1; \theta_Y) > \mathcal{V}_{\mathcal{Q}}(\mu_1'', 1; \theta_Y). \tag{A.8}$$

Since $\mathcal{V}_{\mathcal{Q}}(\mu_1''', 1; \theta_Y) > \mathcal{V}_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ by (A.7) and (A.8), I obtain the result.

(b) This is a direct consequence of Part (a) and the construction of the quality perception design problem. q.e.d.

Proof of Proposition 4.

The result for $s \in \mathcal{S}_a$ follows directly from Lemma 1 (a). Thus, consider $s \in \mathcal{S} \backslash \mathcal{S}_a$.

Binary modifications of a non-binary test. Take any test $\widehat{T} = (\widehat{p}^b, \widehat{p}^g) \in \mathcal{T}_Z$ with $Z > 2$ as given. Define $\overline{p}^b \equiv \widehat{p}^b_1 + \widehat{p}^b_2$ and $\overline{p}^g \equiv \widehat{p}^g_1 + \widehat{p}^g_2$. For my arguments below, modifications of the test $\widehat{T}$ that affect only the first two test results will be important; that is, consider tests from

$$\widehat{\mathcal{T}}(\widehat{T}) \equiv \{(\check{p}^b, \check{p}^g) \in \mathcal{T} | \check{p}^b_1 + \check{p}^b_2 = \overline{p}^b, \check{p}^g_1 + \check{p}^g_2 = \overline{p}^g \text{ and } \forall \sigma \geq 3 \forall \omega \in \{b, g\} : \check{p}^\omega_\sigma = \widehat{p}^\omega_\sigma\}.$$

I explain now how I can rewrite the principal's objective function and the threshold agent's participation constraint in the problem $\mathsf{TD}(s)$ for tests from $\widehat{\mathcal{T}}(\widehat{T})$ in a convenient way.

Consider first the principal's objective function. For the transformation of this function it will prove useful to introduce two definitions and to state three properties. Define $\widehat{\theta}_Y \equiv \theta_Y(s)\overline{p}^g / (\theta_Y(s)\overline{p}^g + (1 - \theta_Y(s))\overline{p}^b)$ and $C_P \equiv \sum_{\sigma > 2} p^{\theta_Y(s)}(\widehat{p}^b_\sigma, \widehat{p}^g_\sigma) v(\mu(\widehat{p}^b_\sigma, \widehat{p}^g_\sigma; \theta_Y(s)))$. The three properties are then as follows. First,

$$\begin{aligned}
\mu(\check{p}^b_\sigma, \check{p}^g_\sigma; \theta_Y(s)) &= \frac{\theta_Y(s)\check{p}^g_\sigma}{\theta_Y(s)\check{p}^g_\sigma + (1 - \theta_Y(s))\check{p}^b_\sigma} \\
&= \frac{\frac{\theta_Y(s)\overline{p}^g}{\theta_Y(s)\overline{p}^g + (1-\theta_Y(s))\overline{p}^b} \frac{\check{p}^g_\sigma}{\overline{p}^g}}{\frac{\theta_Y(s)\overline{p}^g}{\theta_Y(s)\overline{p}^g + (1-\theta_Y(s))\overline{p}^b} \frac{\check{p}^g_\sigma}{\overline{p}^g} + \frac{(1-\theta_Y(s))\overline{p}^b}{\theta_Y(s)\overline{p}^g + (1-\theta_Y(s))\overline{p}^b} \frac{\check{p}^b_\sigma}{\overline{p}^b}} \\
&= \mu(\check{p}^b_\sigma / \overline{p}^b, \check{p}^g_\sigma / \overline{p}^g; \widehat{\theta}_Y).
\end{aligned} \tag{A.9}$$

Second,

$$\begin{aligned}\frac{p^{\theta_Y(s)}(\check{p}^b_\sigma,\check{p}^g_\sigma)}{p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)} &= \frac{\theta_Y(s)\check{p}^g_\sigma+(1-\theta_Y(s))\check{p}^b_\sigma}{\theta_Y(s)\overline{p}^g+(1-\theta_Y(s))\overline{p}^b}\\ &= \frac{\theta_Y(s)\overline{p}^g}{\theta_Y(s)\overline{p}^g+(1-\theta_Y(s))\overline{p}^b}\frac{\check{p}^g_\sigma}{\overline{p}^g}+\frac{(1-\theta_Y(s))\overline{p}^b}{\theta_Y(s)\overline{p}^g+(1-\theta_Y(s))\overline{p}^b}\frac{\check{p}^b_\sigma}{\overline{p}^b}\\ &= p^{\widehat{\theta}_Y}(\check{p}^b_\sigma/\overline{p}^b,\check{p}^g_\sigma/\overline{p}^g). \qquad (A.10)\end{aligned}$$

Third,

$$p^{\widehat{\theta}_Y}(\check{p}^b_1/\overline{p}^b,\check{p}^g_1/\overline{p}^g) = 1-p^{\widehat{\theta}_Y}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g). \qquad (A.11)$$

This allows me to rewrite the principal's objective function in the following way:

$$\begin{aligned}&p^{\theta_Y(s)}(\check{p}^b_1,\check{p}^g_1)v(\mu(\check{p}^b_1,\check{p}^g_1;\theta_Y(s)))+p^{\theta_Y(s)}(\check{p}^b_2,\check{p}^g_2)v(\mu(\check{p}^b_2,\check{p}^g_2;\theta_Y(s)))+C_P\\ =\;& p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)\left[\frac{p^{\theta_Y(s)}(\check{p}^b_1,\check{p}^g_1)}{p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)}v(\mu(\check{p}^b_1,\check{p}^g_1;\theta_Y(s)))+\frac{p^{\theta_Y(s)}(\check{p}^b_2,\check{p}^g_2)}{p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)}v(\mu(\check{p}^b_2,\check{p}^g_2;\theta_Y(s)))\right]\\ &+C_P\\ =\;& p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)\Big[(1-p^{\widehat{\theta}_Y}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g))v(\mu(\check{p}^b_1/\overline{p}^b,\check{p}^g_1/\overline{p}^g;\widehat{\theta}_Y))\\ &+p^{\widehat{\theta}_Y}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g)v(\mu(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g;\widehat{\theta}_Y))\Big]+C_P\\ =\;& p^{\theta_Y(s)}(\overline{p}^b,\overline{p}^g)\mathcal{V}_{\mathcal{T}}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g;\widehat{\theta}_Y)+C_P.\end{aligned}$$

The second equality uses the properties (A.9), (A.10) and (A.11). The third equality shows that the principal's objective function is for the considered class of tests a positive linear transformation of $\mathcal{V}_{\mathcal{T}}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g;\widehat{\theta}_Y)$.

Consider now the threshold agent's participation constraint. When I define $\widehat{s}\equiv s\overline{p}^g/(s\overline{p}^g+(1-s)\overline{p}^b)$ and $C_A\equiv\sum_{\sigma>2}p^s(\widehat{p}^b_\sigma,\widehat{p}^g_\sigma)u(\mu(\widehat{p}^b_\sigma,\widehat{p}^g_\sigma;\theta_Y(s)))$, I obtain by an analogous reasoning that I can rewrite the threshold agent's expected payoff in the following way:

$$\begin{aligned}&p^s(\check{p}^b_1,\check{p}^g_1)u(\mu(\check{p}^b_1,\check{p}^g_1;\theta_Y(s)))+p^s(\check{p}^b_2,\check{p}^g_2)u(\mu(\check{p}^b_2,\check{p}^g_2;\theta_Y(s)))+C_A\\ =\;& p^s(\overline{p}^b,\overline{p}^g)\mathcal{U}^{\widehat{s}}_{\mathcal{T}}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g;\widehat{\theta}_Y)+C_A.\end{aligned}$$

This allows me to write the threshold agent's participation constraint as $\mathcal{U}^{\widehat{s}}_{\mathcal{T}}(\check{p}^b_2/\overline{p}^b,\check{p}^g_2/\overline{p}^g;\widehat{\theta}_Y)=\widehat{\overline{u}}$ with $\widehat{\overline{u}}\equiv(u(\theta_N(s))-C_A)/p^s(\overline{p}^b,\overline{p}^g)$.

Define $p^b_\sigma\equiv\check{p}^b_\sigma/\overline{p}^b$ and $p^g_\sigma\equiv\check{p}^g_\sigma/\overline{p}^g$, and consider $\check{p}^b_\sigma/\overline{p}^b<\check{p}^g_\sigma/\overline{p}^g$ such that the quality perception associated to the first test result is smaller than that associated to the second one. Choosing a test $(\check{p}^b,\check{p}^g)\in\widehat{\mathcal{T}}(\widehat{T})$ with $\check{p}^b_\sigma/\overline{p}^b<\check{p}^g_\sigma/\overline{p}^g$ in the problem $\mathsf{TD}(s)$ is essentially the same as choosing $(p^b_2,p^g_2)\in[0,1]^2$ with $p^b_2<p^g_2$ in the problem $\mathsf{BTD}(\widehat{s},\widehat{\theta}_Y,\widehat{\overline{u}})$. Moreover, by Lemma 2, this is essentially the same as choosing a quality perception pair $(\mu_1,\mu_2)\in\mathcal{Q}(\widehat{\theta}_Y)$ in the problem $\mathsf{QPD}(\widehat{s},\widehat{\theta}_Y,\widehat{\overline{u}})$. The subsequent lemma summarizes what this proof has established so far:

Lemma A4 *Suppose that $s\in\mathcal{S}\backslash\mathcal{S}_a$, that $\widehat{T}\in\mathcal{T}_Z$ with $Z>2$, and that $T',T''\in\widehat{\mathcal{T}}(\widehat{T})$. Let $\widehat{\theta}_Y$, $\widehat{s}$ and $\widehat{\overline{u}}$ derive from s and $\widehat{T}$ as described above. Moreover, denote the quality perception pair associated to the first two test results of the test T' (resp. T'') by (μ'_1,μ'_2) (resp. (μ''_1,μ''_2)) and suppose that $\mu'_1<\mu'_2$ (resp. $\mu''_1<\mu''_2$). (a) T' satisfies the participation constraint in the problem*

TD(s) if, and only if, $\mathcal{U}_{\mathcal{Q}}^{\widehat{s}}(\mu_1', \mu_2'; \widehat{\theta}_Y) = \widehat{\overline{u}}$. *(b) The principal's objective function in the problem TD(s) is strictly higher for* T' *than for* T'' *if, and only if,* $\mathcal{V}_{\mathcal{Q}}(\mu_1', \mu_2'; \widehat{\theta}_Y) > \mathcal{V}_{\mathcal{Q}}(\mu_1'', \mu_2''; \widehat{\theta}_Y)$. *(c)* $(\mu_1', \mu_2') \in \mathcal{Q}(\widehat{\theta}_Y)$. *Moreover, for any* $(\mu_1, \mu_2) \in \mathcal{Q}(\widehat{\theta}_Y)$ *there exists a unique test* $T \in \widehat{\mathcal{T}}(\widehat{T})$ *that implies these quality perceptions.*

The merging of test results. Note that $\mu(p_{\sigma'}^b, p_{\sigma'}^g; \theta_Y(s)) = \mu(p_{\sigma''}^b, p_{\sigma''}^g; \theta_Y(s)) = \mu^\circ$ implies that $\mu(p_{\sigma'}^b + p_{\sigma''}^b, p_{\sigma'}^g + p_{\sigma''}^g; \theta_Y(s)) = \mu^\circ$. Merging the test results σ' and σ'' means to construct a new test that generates a single test result whenever the old test would generate either the test result σ' or the test result σ''. When I merge two test results that imply the same quality perceptions, neither the quality perception distribution faced by the principal nor the distribution faced by the threshold agent are affected. Thus, this kind of merging has no effect on the expected payoff of the principal and of the threshold agent.

The main argument. Take any test $T_Z \in \mathcal{T}_Z$ with $Z > 2$ that satisfies the constraint in the problem $\mathsf{TD}(s)$ as given. I construct now a binary test that does also satisfy this constraint and that is weakly better for the principal.

Step 1: "merge". Construct a test $T_{Z'} \in \mathcal{T}_{Z'}$ with $Z' \leq Z$ by merging all test results that imply the same quality perceptions. By the reasoning above, this transformation does neither affect the participation constraint nor the objective function in the problem $\mathsf{TD}(s)$. If $Z' > 2$, set $\widehat{T}$ equal to $T_{Z'}$ and go to Step 2. Otherwise, set $\widehat{T}''$ equal to $T_{Z'}$ and go to Step 3.

Step 2: "transform and merge". Define $\widehat{\mu}_\sigma \equiv \mu(\widehat{p}_\sigma^b, \widehat{p}_\sigma^g; \theta_Y(s))$ for the given test $\widehat{T} = (\widehat{p}^b, \widehat{p}^g)$. By Step 1, all test results imply different quality perceptions. Assume without loss of generality that $\widehat{\mu}_1 < \widehat{\mu}_2 < \widehat{\mu}_3 < \ldots$. Let $\widehat{\theta}_Y$, $\widehat{s}$ and $\widehat{\overline{u}}$ derive from s and $\widehat{T}$ as described in the first part of this proof. It can easily be verified that $\widehat{\mu}_1 < \widehat{\theta}_Y < \widehat{\mu}_2$ and that $\widehat{s} < \widehat{\theta}_Y$. By the fact that the test $\widehat{T}$ satisfies the constraint in the problem $\mathsf{TD}(s)$ and Lemma A4 (a), $\mathcal{U}_{\mathcal{Q}}^{\widehat{s}}(\widehat{\mu}_1, \widehat{\mu}_2; \widehat{\theta}_Y) = \widehat{\overline{u}}$. By Lemma 4 (c), there exists a unique $\mu_1'' \in (\widehat{\mu}_1, \widehat{\theta}_Y)$ such that $(\mu_1'', \widehat{\mu}_3) \in \overline{\mathcal{Q}}(\widehat{\mu}_1, \widehat{\mu}_2; \widehat{\theta}_Y)$ and a unique $\mu_1''' \in (\widehat{\mu}_1, \mu_1'')$ such that $\mathcal{U}_{\mathcal{Q}}^{\widehat{s}}(\mu_1''', \widehat{\mu}_3; \widehat{\theta}_Y) = \mathcal{U}_{\mathcal{Q}}^{\widehat{s}}(\widehat{\mu}_1, \widehat{\mu}_2; \widehat{\theta}_Y)$. By Lemma A4 (c), there exists a test $\widehat{T}' \in \widehat{\mathcal{T}}(\widehat{T})$ that implies the quality perceptions μ_1''' and $\widehat{\mu}_3$ for the first two test results. By construction and Lemma A4 (a), also the test $\widehat{T}'$ satisfies the constraint in the problem $\mathsf{TD}(s)$. Furthermore, note that $\mathcal{V}_{\mathcal{Q}}(\widehat{\mu}_1, \widehat{\mu}_2; \widehat{\theta}_Y) \leq \mathcal{V}_{\mathcal{Q}}(\mu_1'', \widehat{\mu}_3; \widehat{\theta}_Y)$ by Lemma 4 (b) and that $\mathcal{V}_{\mathcal{Q}}(\mu_1'', \widehat{\mu}_3; \widehat{\theta}_Y) < \mathcal{V}_{\mathcal{Q}}(\mu_1''', \widehat{\mu}_3; \widehat{\theta}_Y)$ by Lemma 3 (c). Thus, it follows from Lemma A4 (b) that the objective function in the problem $\mathsf{TD}(s)$ is strictly higher for the test $\widehat{T}'$ than for the test $\widehat{T}$. Finally, construct a test $\widehat{T}''$ that generates one test result less than the test $\widehat{T}'$ by merging the test results $\sigma = 2$ and $\sigma = 3$, which generate the same quality perception $\widehat{\mu}_3$. As explained above, this does neither affect the participation constraint nor the objective function in the problem $\mathsf{TD}(s)$.

Step 3: "iterate". Repeat Step 2 with $\widehat{T}''$ assuming the role of $\widehat{T}$ until I obtain a test $\widehat{T}''$ that generates (at most) two test results. By construction, this test satisfies the constraint in the problem $\mathsf{TD}(s)$ and it is at least weakly better for the principal than the initial test T_Z.

This three-step procedure establishes that any non-binary test is weakly dominated by some binary test. Hence, if the problem $\mathsf{BTD}(s)$ possesses a solution, it must also solve the problem

TD(s). By Proposition 3 (b), such a solution exists and it is as described in Proposition 3 (b) with $\theta_Y = \theta_Y(s)$ and $\overline{u} = u(\theta_N(s))$. In fact, when I apply the reasoning in Proposition 3 to the binary test that I obtained from the three-step-procedure here, I obtain the optimal binary test as a test that is better than the test T_Z with that I started here. q.e.d.

Proof of Corollary 1.

(a) In the proof of Propositions 3 and 4, I transform any non-optimal test in a sequence of steps into the optimal test $T^{\text{NFP}}(\rho(s))$ (as characterized in Proposition 4). Since each step does either leave the expected perception of the threshold agent constant or decreases it by the reasoning in Lemma 3 (b), I obtain the result. (Note that Lemma 3 (b) applies also to the transformations in Proposition 4 since for any test $T \in \widehat{\mathcal{T}}(\widehat{T})$, by a reasoning analogous to that in the first part of the proof of Proposition 4, the threshold agent's expected perception corresponds to a positive linear transformation of the term $\overline{\mu}^{\widehat{s}}(\mu_1, \mu_2; \widehat{\theta}_Y)$ where μ_1 and μ_2 describe the first two quality perceptions implied by the test T.)

(b) It follows from Lemma 3 (d) that all binary tests that induce s can be ordered according to whether both quality perceptions are higher or lower. Thus, when (μ_1', μ_2') and $(\mu_1'', 1)$ with $\mu_2' < 1$ are the quality perception pairs implied by tests that do both induce s, then $\mu_1' < \mu_1''$. The result follows from this and the fact that some test that induces a quality perception pair $(\mu_1, 1)$ is optimal by Proposition 3 (b). q.e.d.

Proof of Lemma 5.

(a) By differentiating $y(\mu) = \alpha^{\text{FN}}\mu/(\alpha^{\text{FN}}\mu + (1-\mu)\alpha^{\text{FP}})$, I obtain $y'(\mu) = \alpha^{\text{FN}}\alpha^{\text{FP}}/(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))^2$, $y''(\mu) = -2\alpha^{\text{FN}}\alpha^{\text{FP}}(\alpha^{\text{FN}}-\alpha^{\text{FP}})/(\alpha^{\text{FN}}\mu+(1-\mu)\alpha^{\text{FP}})^3$, and $y'''(\mu) = 6\alpha^{\text{FN}}\alpha^{\text{FP}}(\alpha^{\text{FN}}-\alpha^{\text{FP}})^2/(\alpha^{\text{FN}}\mu + (1-\mu)\alpha^{\text{FP}})^4$. The inequalities in the lemma follow directly from this.

(b) By differentiating $v(\mu) = -\alpha^{\text{FN}}\alpha^{\text{FP}}\mu(1-\mu)/(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))$, I obtain $v'(\mu) = \alpha^{\text{FN}}\alpha^{\text{FP}}(\alpha^{\text{FN}}\mu^2 - \alpha^{\text{FP}}(1-\mu)^2)/(\alpha^{\text{FN}}\mu+\alpha^{\text{FP}}(1-\mu))^2$, $v''(\mu) = 2(\alpha^{\text{FN}}\alpha^{\text{FP}})^2/(\alpha^{\text{FN}}\mu+\alpha^{\text{FP}}(1-\mu))^3$, and $v'''(\mu) = 6(\alpha^{\text{FN}}\alpha^{\text{FP}})^2(\alpha^{\text{FP}} - \alpha^{\text{FN}})/(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))^4$. The inequalities in the lemma follow directly from this.

(c) By differentiating $u = \widehat{u} \circ y$, I obtain

$$\begin{aligned} u'(\mu) &= \widehat{u}'(y(\mu))y'(\mu), \\ u''(\mu) &= \widehat{u}''(y(\mu))(y'(\mu))^2 + \widehat{u}'(y(\mu))y''(\mu), \text{ and} \\ u'''(\mu) &= \widehat{u}'''(y(\mu))(y'(\mu))^3 + 3\widehat{u}''(y(\mu))y'(\mu)y''(\mu) + \widehat{u}'(y(\mu))y'''(\mu). \end{aligned}$$

$u' > 0$ follows from the assumption that $\widehat{u}' > 0$ and the fact that $y' > 0$ by (a).

Consider first $\alpha^{\text{FP}} < \alpha^{\text{FN}}$. My assumptions on $\widehat{u}$ and (a) imply then that each summand of u'' is negative and that each summand of u''' is positive. Thus, $u'' < 0$ and $u''' > 0$.

Consider next $\alpha^{\text{FP}} > \alpha^{\text{FN}}$. Moreover, suppose that $\widehat{u}$ is a HARA utility function (see Section 1 of the supplementary material); this means that there exist parameters c_1 and c_2 such that $-\widehat{u}''(\mu)/\widehat{u}'(\mu) = 1/(c_1\mu + c_2)$. Moreover, it follows from the proof of Part (a) that $y''(\mu)/(y'(\mu))^2 = 2(\alpha^{\text{FP}} - \alpha^{\text{FN}})(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))/(\alpha^{\text{FN}}\alpha^{\text{FP}})$. This allows me to write

$$\begin{aligned} u''(\mu) &= \widehat{u}''(y(\mu))(y'(\mu))^2 \left[1 + \frac{\widehat{u}'(y(\mu))}{\widehat{u}''(y(\mu))}\frac{y''(\mu)}{(y'(\mu))^2}\right] \\ &= \widehat{u}''(y(\mu))(y'(\mu))^2 \left[1 - (c_1\mu + c_2)\frac{2(\alpha^{\text{FP}} - \alpha^{\text{FN}})(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))}{\alpha^{\text{FN}}\alpha^{\text{FP}}}\right]. \end{aligned}$$

The expression in front of the brackets is strictly negative by the assumption that $\widehat{u}'' < 0$. Thus, to identify conditions under which $u'' < 0$ ($u'' > 0$), I need to argue that the expression in brackets is strictly positive (strictly negative). For any given $\alpha^{\text{FN}}, \alpha^{\text{FP}} > 0$, sufficient for $u'' < 0$ is that c_1 and c_2 are sufficiently close to zero. This is, in particular, the case when $c_1 = 0$ and c_2 is sufficiently close to zero (=CARA utility with sufficiently sufficiently strong risk-aversion), and when $c_2 = 0$ and c_1 is sufficiently close to zero (=CRRA utility with sufficiently strong risk-aversion). On the other hand, if $c_1 = 0$ and c_2 is sufficiently large (=CARA utility with sufficiently weak risk-aversion), I have $u'' > 0$.

It remains to argue that $u''' > 0$ for CARA and CRRA with sufficiently strong risk-aversion. Since $y'''(\mu)/(3y'(\mu)y''(\mu)) = (\alpha^{\text{FP}} - \alpha^{\text{FN}})(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))/(\alpha^{\text{FN}}\alpha^{\text{FP}})$, I can write

$$\begin{aligned} u'''(\mu) &= \widehat{u}'''(y(\mu))(y'(\mu))^3 \left[1 + 3\frac{\widehat{u}''(y(\mu))}{\widehat{u}'''(y(\mu))}\frac{y''(\mu)}{(y'(\mu))^2}\right] + \widehat{u}'(y(\mu))y'''(\mu) \\ &= \widehat{u}'''(y(\mu))(y'(\mu))^3 \left[1 - 3\left(-\frac{\widehat{u}''(y(\mu))}{\widehat{u}'''(y(\mu))}\right)\frac{2(\alpha^{\text{FP}} - \alpha^{\text{FN}})(\alpha^{\text{FN}}\mu + \alpha^{\text{FP}}(1-\mu))}{\alpha^{\text{FN}}\alpha^{\text{FP}}}\right] \\ &\quad + \widehat{u}'(y(\mu))y'''(\mu) \end{aligned}$$

Since $\widehat{u}'''(y(\mu))(y'(\mu))^3$ is weakly positive and since $\widehat{u}'(y(\mu))y'''(\mu)$ is strictly positive by the assumptions on $\widehat{u}$ and (a), it suffices to argue that the expression in brackets is weakly positive. I have $-\widehat{u}''(\mu)/\widehat{u}'''(\mu) = c_2$ for CARA utility and $-\widehat{u}''(\mu)/\widehat{u}'''(\mu) = c_1\mu/(1+c_1)$ for CRRA utility. For any given $\alpha^{\text{FN}}, \alpha^{\text{FP}} > 0$, it follows directly from this that sufficient for what I have to show is that c_2 (resp. c_1) is sufficiently close to zero. This yields the result. q.e.d.

Proof of Proposition 5.

(a) Suppose that $\alpha^{\text{FP}} > \alpha^{\text{FN}}$. By Lemma 5 (b) and (c), all the assumptions of my reduced-form model in Section 3 are satisfied when $\widehat{u}$ is either a CARA or a CRRA utility function that exhibits sufficiently strong risk-aversion. Thus, the first part of (a) follows directly from Proposition 4. If $\widehat{u}$ is instead a CARA utility function that exhibits sufficiently weak risk-aversion, the principal and the agent are both risk-loving by Lemma 5 (b) and (c). Thus, any accurate test leads to full participation and perfect information generation. Choosing the accurate test $T^{\text{NFP}}(1)$ is clearly optimal.

(b) Suppose that $\alpha^{\text{FP}} < \alpha^{\text{FN}}$. By Lemma 5 (b) and (c), I have $u' > 0$, $u'' < 0$, $u''' > 0$ and $v'' > 0$ as in my reduced-form model, but the assumption that $v''' > 0$ is violated. I use $v''' \geq 0$ only in the proof to Lemma 4 (b) in the analysis of the quality perception design problem $\mathsf{QPD}(s, \theta_Y, \overline{u})$. It is part of a condition that ensures that as I move to the north-east of any iso-expected perception curve $\overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$, the principal becomes weakly better off and the threshold agent becomes strictly better off. What I need in the derivation of the optimal test structure (i.e., in Propositions 3 and 4) is that there exists for any quality perception pair $(\mu_1', \mu_2') \in \mathcal{Q}(\theta_Y)$ an increasing curve of quality perception pairs on $\mathcal{Q}(\theta_Y)$ through (μ_1', μ_2') such that this property holds true. I use now the specific structure of the here considered problem to show that such a curve exists even though the assumption $v''' > 0$ is violated. More specifically, I will show that the curve that keeps the expected estimate from the principal's perspective constant (instead of the expected perception from the threshold agent's perspective) does the job.

Define

$$\overline{y}^{\theta}(\mu_1, \mu_2; \theta_Y) \equiv y(\mu_1) + p^{\theta}(\beta^b(\mu_1, \mu_2; \theta_Y), \beta^g(\mu_1, \mu_2; \theta_Y)) \cdot (y(\mu_2) - y(\mu_1)) \tag{A.12}$$

and

$$\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y) \equiv \{(\mu_1, \mu_2) \in \mathcal{Q}(\theta_Y) | \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) = \overline{y}^{\theta_Y}(\mu_1', \mu_2'; \theta_Y)\}.$$

For the binary test that induces the quality perception pair (μ_1, μ_2), $\overline{y}^{\theta}(\mu_1, \mu_2; \theta_Y)$ describes the expected estimate from the viewpoint of somebody, who believes that the agent is good with probability θ; $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$ describes the set of all quality perception pairs that imply the same expected estimate from the principal's perspective as the quality perception pair (μ_1', μ_2').

The first property that I need to show is that the principal does not get worse off as I move to the north-east of the graph $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$. The following lemma shows that she is just indifferent among all quality perception pairs on this graph:

Lemma A5 *If $(\mu_1'', \mu_2'') \in \overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$, then $\mathcal{V}_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y) = \mathcal{V}_{\mathcal{Q}}(\mu_1'', \mu_2''; \theta_Y)$.*

Proof. By using the structure of $y(\mu)$, it can easily be verified that

$$\mu y(\mu) = \gamma_1 \mu - \gamma_2 y(\mu) \tag{A.13}$$

with $\gamma_1 \equiv \alpha^{\mathrm{FN}}/(\alpha^{\mathrm{FN}} - \alpha^{\mathrm{FP}})$ and $\gamma_2 \equiv \alpha^{\mathrm{FP}}/(\alpha^{\mathrm{FN}} - \alpha^{\mathrm{FP}})$.

When I use the definition of $y(\mu)$, I can write $v(\mu) = -\alpha^{\mathrm{FP}}(y(\mu) - \mu y(\mu))$. By using the relationship in (A.13) and simplifying then, this becomes $v(\mu) = -\alpha^{\mathrm{FN}}\alpha^{\mathrm{FP}}/(\alpha^{\mathrm{FN}} - \alpha^{\mathrm{FP}}) \cdot (y(\mu) - \mu)$. Since this expression is linear in $y(\mu)$ and in μ, it follows immediately from this that the principal's expected payoff depends only through $\overline{y}^{\theta_Y}(\mu_1'', \mu_2''; \theta_Y)$ and through $\overline{\mu}^{\theta_Y}(\mu_1'', \mu_2''; \theta_Y)$ on the quality perception pair (μ_1'', μ_2''). Since the latter term is constant by Lemma 3 (a) and since the former is constant by the supposition of this lemma, I obtain what I have to show. q.e.d.

Next I will prove an auxiliary property that will turn out to be useful for showing how the threshold agent is affected by movements on $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$: if I keep the expected estimate from the principal's perspective constant, it is also constant from the threshold agent's perspective.

Lemma A6 *If $(\mu_1'', \mu_2'') \in \overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$, then $\overline{y}^{\theta}(\mu_1', \mu_2'; \theta_Y) = \overline{y}^{\theta}(\mu_1'', \mu_2''; \theta_Y)$ for all $\theta \in [0, 1]$.*

Proof. Consider first the expected estimate from the perspective of somebody who believes that the agent is good:

$$\begin{aligned}
\overline{y}^{1}(\mu_1, \mu_2; \theta_Y) &= y(\mu_1) + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}\frac{\mu_2}{\theta_Y}(y(\mu_2) - y(\mu_1)) \\
&= \left(1 - \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}\frac{\mu_2}{\theta_Y}\right) y(\mu_1) + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}\frac{\mu_2}{\theta_Y} y(\mu_2) \\
&= \frac{1}{\theta_Y}\left[\frac{\mu_2 - \theta_Y}{\mu_2 - \mu_1}\left(\frac{\mu_2 - \mu_1}{\mu_2 - \theta_Y}\theta_Y - \frac{\theta_Y - \mu_1}{\mu_2 - \theta_Y}\mu_2\right) y(\mu_1) + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}\mu_2 y(\mu_2)\right] \\
&= \frac{1}{\theta_Y}\left[\frac{\mu_2 - \theta_Y}{\mu_2 - \mu_1}\mu_1 y(\mu_1) + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}\mu_2 y(\mu_2)\right] \\
&= \frac{1}{\theta_Y}\left[\frac{\mu_2 - \theta_Y}{\mu_2 - \mu_1}[\gamma_1\mu_1 - \gamma_2 y(\mu_1)] + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1}[\gamma_1\mu_2 - \gamma_2 y(\mu_2)]\right] \\
&= \frac{1}{\theta_Y}\left[\gamma_1 \overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - \gamma_2 \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)\right] \\
&= \gamma_1 - \frac{\gamma_2}{\theta_Y}\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)
\end{aligned}$$

The transformations arise as follows: The first equality follows from using $\theta = 1$ and the definition of $\beta^g(\mu_1, \mu_2; \theta_Y)$ in (A.12). The second and the third equality follow from rearranging. The fourth equality follows from using that the expression in round brackets simplifies to μ_1. The fifth equality follows from using (A.13). The sixth equality follows from using the definition of $\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)$ and of $\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)$. The last equality follows from using that $\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) = \theta_Y$ by Lemma 3 (a).

The transformations show that if (μ_1, μ_2) is modified such that $\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)$ stays constant, then $\overline{y}^{1}(\mu_1, \mu_2; \theta_Y)$ stays constant. Since $\overline{y}^{\theta}(\mu_1, \mu_2; \theta_Y)$ is linear in θ, this implies also for any other θ that $\overline{y}^{\theta}(\mu_1, \mu_2; \theta_Y)$ stays constant. q.e.d.

To see how the threshold agent is affected by movements on the graph $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$, I need to derive some properties of this graph:

Lemma A7 *(a) $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$ is described by a curve $(\mu_1, \mu_2(\mu_1))$ where $\mu_2(\mu_1)$ is strictly increasing and continuous differentiable with*

$$\mu_2'(\mu_1) = \frac{y(\mu_2) - \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)}{\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - y(\mu_1)} \cdot \left. \frac{y'(\mu_1) - \frac{y(\mu_2)-y(\mu_1)}{\mu_2-\mu_1}}{\frac{y(\mu_2)-y(\mu_1)}{\mu_2-\mu_1} - y'(\mu_2)} \right|_{\mu_2=\mu_2(\mu_1)}.$$

(b) In the considered case with $\alpha^{FP} < \alpha^{FN}$, $\mu_2'(\mu_1) > 0$.

Proof. (a) By using that $p^{\theta_Y}(\beta^b(\mu_1, \mu_2; \theta_Y), \beta^g(\mu_1, \mu_2; \theta_Y)) = (\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - \mu_1)/(\mu_2 - \mu_1)$ by (4) and that $\overline{\mu}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) = \theta_Y$ by Lemma 3 (a), I can rewrite (A.12) for $\theta = \theta_Y$ as

$$\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) = y(\mu_1) + \frac{\theta_Y - \mu_1}{\mu_2 - \mu_1} \cdot (y(\mu_2) - y(\mu_1)).$$

By differentiating this expression partially, I obtain

$$\begin{aligned} \frac{\partial}{\partial \mu_1} \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) &= \frac{\mu_2 - \theta_Y}{\mu_2 - \mu_1} \left[y'(\mu_1) - \frac{y(\mu_2) - y(\mu_1)}{\mu_2 - \mu_1} \right] \text{ and} \\ \frac{\partial}{\partial \mu_2} \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) &= -\frac{\theta_Y - \mu_1}{\mu_2 - \mu_1} \left[\frac{y(\mu_2) - y(\mu_1)}{\mu_2 - \mu_1} - y'(\mu_2) \right]. \end{aligned}$$

By Lemma 5 (a), both partial derivatives are continuous and non-zero. Thus, by the Implicit Function Theorem, there exists a continuously differentiable function $\mu_2(\mu_1)$ such that $\overline{y}^{\theta_Y}(\mu_1, \mu_2(\mu_1); \theta_Y)$ is constant. Moreover, the derivative of this function is implicitly defined by

$$\frac{\mathrm{d}\mu_2}{\mathrm{d}\mu_1} = -\frac{\frac{\partial}{\partial \mu_1} \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)}{\frac{\partial}{\partial \mu_2} \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)} = \frac{\frac{\mu_2 - \theta_Y}{\mu_2 - \mu_1} y'(\mu_1) - \frac{y(\mu_2)-y(\mu_1)}{\mu_2-\mu_1}}{\frac{\theta_Y - \mu_1}{\mu_2 - \mu_1} \frac{y(\mu_2)-y(\mu_1)}{\mu_2-\mu_1} - y'(\mu_2)}.$$

Since (A.12) implies that $(\theta_Y - \mu_1)/(\mu_2 - \mu_1) = (\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - y(\mu_1))/(y(\mu_2) - y(\mu_1))$ and that $(\mu_2 - \theta_Y)/(\mu_2 - \mu_1) = (y(\mu_2) - \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y))/(y(\mu_2) - y(\mu_1))$, I obtain the formula in (a).

(b) By Lemma 5 (a), $y(\mu)$ is strictly concave in the considered case with $\alpha^{\mathrm{FP}} < \alpha^{\mathrm{FN}}$. This implies that the second term in the formula for $\mu_2'(\mu_1)$ in (a) is strictly positive. The first term is also strictly positive since $\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)$ is a strict convex combination of $y(\mu_1)$ and $y(\mu_2)$ and since $y(\mu_2) > y(\mu_1)$ by Lemma 5 (a). Hence, $\mu_2'(\mu_1) > 0$. q.e.d.

I have now everything at hand to explain why the agent becomes better off as I move to the northeast of the graph $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$.

Lemma A8 *If $(\mu_1'', \mu_2'') \in \overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$ with $\mu_2' < \mu_2''$, then $\mathcal{U}^s_{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y) < \mathcal{U}^s_{\mathcal{Q}}(\mu_1'', \mu_2''; \theta_Y)$.*

Proof. By using that $p^s(\beta^b(\mu_1, \mu_2; \theta_Y), \beta^g(\mu_1, \mu_2; \theta_Y)) = (\overline{y}^s(\mu_1, \mu_2; \theta_Y) - y(\mu_1))/(y(\mu_2) - y(\mu_1))$ by (A.12) and that $u(\mu) = \widehat{u}(y(\mu))$, I can write

$$\begin{aligned}
\mathcal{U}^s_{\mathcal{Q}}(\mu_1, \mu_2; \theta_Y) &= \widehat{u}(y(\mu_1)) + \frac{\overline{y}^s(\mu_1, \mu_2; \theta_Y) - y(\mu_1)}{y(\mu_2) - y(\mu_1)}(\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))) \\
&= \widehat{u}(y(\mu_1)) + \frac{\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - y(\mu_1)}{y(\mu_2) - y(\mu_1)}(\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))) \\
&\quad -(\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - \overline{y}^s(\mu_1, \mu_2; \theta_Y))\frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}. \qquad \text{(A.14)}
\end{aligned}$$

Sufficient for what I need to show is that the expression in each of the two lines of (A.14) is strictly increasing in μ_1 when I set $\mu_2 = \mu_2(\mu_1)$ (as defined in Lemma A7).

Effect of μ_1 on the second line of (A.14). First note that $\overline{y}^{\theta_Y}(\mu_1, \mu_2(\mu_1); \theta_Y)$ is constant for the considered quality perception pairs by construction and that $\overline{y}^s(\mu_1, \mu_2(\mu_1); \theta_Y)$ is constant by Lemma A6. Thus, μ_1 affects the second line only through the expression $(\widehat{u}(y(\mu_2(\mu_1))) - \widehat{u}(y(\mu_1)))/(y(\mu_2(\mu_1)) - y(\mu_1))$. Since $\overline{y}^{\theta_Y}(\mu_1, \mu_2(\mu_1); \theta_Y) > \overline{y}^s(\mu_1, \mu_2(\mu_1); \theta_Y)$, I need to argue that this expression is strictly decreasing in μ_1. Note next that $(\widehat{u}(y_2) - \widehat{u}(y_1))/(y_2 - y_1)$ is strictly decreasing in y_1 and in y_2 by strict concavity of $\widehat{u}$. Hence, what I need to show follows because $y' > 0$ by Lemma 5 (a) and because $\mu_2'(\mu_1) > 0$ by Lemma A7 (b).

Effect of μ_1 on the first line of (A.14). By taking the total differential of the first line with respect to μ_1, I obtain

$$\begin{aligned}
&y'(\mu_1)\frac{y(\mu_2) - \overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y)}{y(\mu_2) - y(\mu_1)}\left(\widehat{u}'(y(\mu_1)) - \frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}\right) \\
&+\mu_2'(\mu_1)y'(\mu_2)\frac{\overline{y}^{\theta_Y}(\mu_1, \mu_2; \theta_Y) - y(\mu_1)}{y(\mu_2) - y(\mu_1)}\left(\widehat{u}'(y(\mu_2)) - \frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}\right).
\end{aligned}$$

I need to argue that this expression is strictly positive. By using the formula for $\mu_2'(\mu_1)$ from Lemma A7 and by simplifying, I obtain that this is equivalent to showing that

$$\begin{aligned}
\xi(\mu_1, \mu_2) &\equiv \left(\widehat{u}'(y(\mu_1)) - \frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}\right) \\
&\quad +c(\mu_1, \mu_2)\left(\widehat{u}'(y(\mu_2)) - \frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}\right)
\end{aligned}$$

with

$$c(\mu_1, \mu_2) \equiv \frac{y'(\mu_1) - \frac{y(\mu_2) - y(\mu_1)}{\mu_2 - \mu_1}}{\frac{y(\mu_2) - y(\mu_1)}{\mu_2 - \mu_1} - y'(\mu_2)} \cdot \frac{y'(\mu_2)}{y'(\mu_1)}$$

is strictly positive. I have

$$\xi(\mu_1, \mu_2) = \left(\widehat{u}'(y(\mu_1)) + \widehat{u}'(y(\mu_2)) - 2\frac{\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1))}{y(\mu_2) - y(\mu_1)}\right)$$

$$
\begin{aligned}
&+(1-c(\mu_1,\mu_2))\left(\frac{\widehat{u}(y(\mu_2))-\widehat{u}(y(\mu_1))}{y(\mu_2)-y(\mu_1)}-\widehat{u}'(y(\mu_2))\right)\\
>\;&\left(\widehat{u}'(y(\mu_1))+\widehat{u}'(y(\mu_2))-2\frac{\widehat{u}(y(\mu_2))-\widehat{u}(y(\mu_1))}{y(\mu_2)-y(\mu_1)}\right)\\
\geq\;&0
\end{aligned}
$$

The transformations arise as follows: The equality follows from rearranging. The first inequality follows from using that $y(\mu_2) > y(\mu_1)$ by Lemma 5 (a), that $(\widehat{u}(y(\mu_2)) - \widehat{u}(y(\mu_1)))/(y(\mu_2) - y(\mu_1)) - \widehat{u}'(y(\mu_2)) > 0$ by strict concavity of $\widehat{u}$, and the fact that $c(\mu_1, \mu_2) < 1$. To see the last property, I can use the specific structure of $y(\mu)$ to simplify $c(\mu_1, \mu_2)$:

$$
c(\mu_1,\mu_2)=\frac{\alpha^{\mathrm{FN}}\mu_1+\alpha^{\mathrm{FP}}(1-\mu_1)}{\alpha^{\mathrm{FN}}\mu_2+\alpha^{\mathrm{FP}}(1-\mu_2)}.
$$

The supposition that $\alpha^{\mathrm{FP}} < \alpha^{\mathrm{FN}}$ and $\mu_2 > \mu_1$ imply that $c(\mu_1, \mu_2) < 1$. Finally, the last inequality follows from $\widehat{u}''' \geq 0$ and Lemma A2 with $h = \widehat{u}$. q.e.d.

Lemma A5 and A8 show that Propositions 3 and 4 extend to the here considered case where $v''' \geq 0$ is violated when $\overline{\mathcal{Y}}(\mu_1', \mu_2'; \theta_Y)$ assumes the role of $\overline{\mathcal{Q}}(\mu_1', \mu_2'; \theta_Y)$ in Lemma 4 (b). (Note that the technical property in Lemma 4 (c), which is also used, follows from a reasoning that is analogous to that in Lemma 4 (c).) q.e.d.

References

Alonso, R. and O. Câmara (2015). Persuading voters. Mimeo, November 2015.

Alonso, R. and O. Câmara (2016). Bayesian persuasion with heterogeneous priors. Mimeo, April 2016.

Alós-Ferrer, C. and J. Prat (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., V. Kadiyali, and A. Zussman (2012). Putting grades in context. *Journal of Labor Economics 30*, 445–478.

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

Bergemann, D. and M. Pesendorfer (2007). Information structures in optimal auctions. *Journal of Economic Theory 137*, 580–609.

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

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

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

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

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

Dubey, P. and J. Geanakoplos (2010). Grading exams: 100,99,98,...or a,b,c? *Games and Economic Behavior 69*, 72–94.

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

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

Gentzkow, M. and E. Kamenica (2014). Costly persuasion. *American Economic Review: Papers & Proceedings 104*, 457–462.

Gentzkow, M. and E. Kamenica (2016). Competition in persuasion. Mimeo, January 2016.

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

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

Goldstein, I. and Y. Leitner (2015). Stress tests and information disclosure. Mimeo, November 2015.

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

Harbaugh, R. and E. Rasmusen (2014). Coarse grades: Informing the public by withholding information. Mimeo, December 2014.

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., C. J. Bettinger, M. M. Gallaher, N. M. Keller, J. Wilson, and G. J. Mertz (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 M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*, 2590–2615.

Kolotilin, A. (2015). Experimental design to persuade. *Games and Economic Behavior 90*, 215–226.

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

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

Lyter, D. W., R. O. Valdiserri, L. A. Kingsley, W. P. Amoroso, and C. R. Rinaldo (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 A. Postlewaite (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*, 380–391.

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

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

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

Ottaviani, M. and A. Prat (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*, 702–723.

Perez-Richet, E. (2014). Interim bayesian persuasion: First steps. *American Economic Review: Papers & Proceedings 104*, 469–474.

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

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

Economic Review 102, 1–30.

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

Rosar, F. (2014). Test design under voluntary participation and conflicting preferences. Mimeo, July 2014.

Schweizer, N. and N. Szech (2014). Optimal revelation of life-changing information. Mimeo, November 2014.

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

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

Taneva, I. (2016). Information design. Mimeo, February 2016.

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

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

Wang, Y. (2013). Bayesian persuasion with multiple receivers. Mimeo, June 2013.

Weiss, A. (1983). A sorting-cum-learning model of education. *Journal of Political Economy 91*, 420–442.