Optimal test design under imperfect private information and voluntary participation*

Frank Rosar[†]
University of Bonn

Elisabeth Schulte
University of Mannheim

This version: April 10, 2012

Abstract

We study the optimal design of a test that issues a public signal about an agent's quality. The agent is imperfectly informed on whether his quality is good or bad. Participation is voluntary and publicly observed. The agent strives for a high quality perception and is averse to information generation, the designer strives for information generation. For a broad class of HARA utility functions on the agent's side, any implementable participation behavior is optimally induced by a pass–fail–test that is not subject to false positives. We demonstrate our results in the context of five applications.

JEL classification: D82, D83

Keywords: test design, asymmetric information, voluntary participation, Bayesian learning, false positives, false negatives, signalling, pooling, certification, grading

*Previous versions of the paper circulated under the titles "Better information generation through inaccurate signaling devices" and "Imperfect private information and the design of information–generating mechanisms". We would like to thank the participants of seminars where this research has been presented, and in particular Marco Battaglini, Georgy Egorov, Hans Peter Grüner, Emir Kamenica, Johannes Koenen, Benny Moldovanu, Tymofiy Mylovanov, Marco Ottaviani, Martin Peitz, Sergei Severinov, Konrad Stahl and Nora Szech for helpful comments and suggestions. The first author gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15. The second author gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through SFB 884.

†Corresponding author: Frank Rosar, e–mail: email@frankrosar.de

1 Introduction

We study the design of a test that issues a public signal about an agent's quality. The agent is imperfectly informed on whether his quality is good or bad. Participation in the test is voluntary and the participation decision is publicly observed. The public's perception of the agent's quality is endogenous and reflects the supposed participation behavior and the test design. Participation exposes the agent to a risk with respect to the perception of his quality. We are interested in situations in which the agent is averse to this risk and prefers a higher quality perception, while the designer of the test benefits from a more accurate quality perception.

The general framework consists of three stages: First, a test is designed. A test maps the agent's quality to a public signal. Second, the agent decides upon participation. He has private information about his quality. Third, the test result and the participation decision are observed by a (set of) receiver(s) who take actions in a receiver game. The agent and the designer care about the outcome of the receiver game. The designer faces two kinds of constraints. She can only indirectly affect decisions by manipulating quality perceptions. In doing so, she is constrained by Bayesian plausibility. Moreover, although the test can generate information about the agent's quality, it requires the agent's voluntary participation. Voluntary participation imposes a constraint on the designer's test choice. The test design problem deals with the interplay of participation incentives and information usage. While other means to govern decisions in the light of asymmetric information have been extensively explored, e.g., in the literature on mechanism design,[1] comparably little is known about test design. Notable exceptions are [18], [13] and [15]. This paper adds to the better understanding of test design.

If the participation stage is absent, we obtain a persuasion problem as studied in [18]. A designer tries to influence a receiver's action and is restricted only by Bayesian plausibility. [13] and [15] study design problems in which an additional constraint on the test choice comes into play. In [13], a monopolistic seller, who is informed whether his product quality is good or bad, chooses a test before pricing his product. A design problem results in which the high quality monopolist chooses a test to affect the pricing game through the induced quality beliefs.[2] A constraint on the test choice is exogenously imposed, only a set of inaccurate pass–fail–tests is admissible.[3] In [15], a certifier designs a grading scheme in order to minimize the mean squared error of the public's estimate of a firm's quality. The firm is perfectly informed about its continuous quality and maximizes the public's estimate of its quality. Applying for certification is voluntary. The participation decision and the certification outcome are observable. A tension between participation incentives and information generation arises from an exogenously imposed participation cost.[4] Exploiting that pooling can strengthen the participation

[1] In the considered framework, a mechanism maps an announcement of the agent directly to a decision. While the first constraint is obsolete, the second constraint is more severe as the designer has to respect revelation incentives besides participation incentives.

[2] As a low quality monopolist can costlessly duplicate the actions of a high quality monopolist, there is no signalling through the test choice or the pricing decision. Attention is restricted to pooling equilibria in which the preferred test of the high quality monopolist is chosen.

[3] The high quality monopolist chooses a test to separate from the low quality monopolist. Without a restriction on the set of admissible tests, an accurate test is optimal.

[4] If no cost of participation is imposed, perfect information can be generated by an accurate grading scheme as the firm is not averse to information generation.

incentives, the authors show that the informativeness of an otherwise accurate grading scheme can be improved by a certain extent of pooling of the worst participating qualities. While the test choice in [13] and [15] is restricted because of an exogenously imposed friction, a restriction arises endogenously from the imperfection of private information, information aversion and voluntary participation of the agent in our paper. Like us, [15] studies the interplay of participation incentives and information generation. The two papers are complementary as they focus on different aspects of pooling. In [15], participation falls to a perfectly informed, information neutral individual and attention is restricted to deterministic grading schemes. Uncertainty on the agent's side does not arise. The question under consideration is *which* qualities should optimally be pooled. In our paper, participation falls to an imperfectly informed, information averse individual. The question is how to optimally impose uncertainty on the agent, that is, *how* to optimally pool a given set of qualities.

We first analyze a reduced–form problem in which the receiver game is not modeled explicitly. The game ends with the receiver's observations and payoffs are determined by the realized quality perception. The agent has a private assessment of the probability that he is good. His utility is increasing in the receiver's quality perception and exhibits hyperbolic absolute risk aversion. Participation in the test is costly for the agent as it comes with a perception risk, caused by the imperfectness of the private information and the informativeness of the test. Participation allows for signalling of favorable private information. Non–participation reveals unfavorable private information. However, the imperfection of private information bounds the associated quality perception from below and precludes unravelling for an accurate test. The participation constraint of the threshold participant weighs up his costs (i.e., the perception risk) and his benefits of participation (i.e., the better expected perception).

There is a continuum of tests that induces the same participation behavior. Different tests induce different distributions of the quality perception. For instance, they differ in the expected perception of the threshold participant. For a higher expected perception of the threshold participant, his participation constraint is "less binding", which allows for the imposition of a higher perception risk on him. The design of a test that induces a certain participation behavior deals with the interplay between the participation constraint and the induced lottery over quality perceptions. We obtain strong results when the designer maximizes information generation as measured by the variance of the induced quality belief distribution or when information aversion on the agent's side and information love on the designer's side yield a zero sum structure. When the agent's risk aversion is neither too strongly increasing nor too strongly decreasing, a pass–fail test which is not subject to false positives is optimal. Surprisingly, the optimal test is not the test for which the expected perception of the threshold participant is highest, but lowest. The reason lies in the interplay of the constraints. On one side, a higher expected perception of the threshold participant allows for making the test more informative. On the other side, a higher expected perception of the threshold participant can only be obtained by a "redistribution of expected perception" through "more pooling of good and bad quality" which makes the inferences less informative. It turns out that the second effect generally dominates. Our strong result regarding the optimal inducement of a given participation behavior makes the problem of finding the optimal participation behavior one–dimensional. For quadratic preferences and uniformly distributed signals, we obtain that a higher participation level is optimally induced by a harder test. A harder test reduces the stigma of failure. It turns out that this reduction makes participation generally more attractive, although it is accompanied by a countervailing effect, the higher probability of failure.

Endowed with our theoretical results for the reduced–form problem, we introduce a non–reduced problem in which we model the receiver game explicitly. The receiver assesses the probability with which the agent is good. Receiver and designer both strive for minimizing weighted squared errors of the receiver's probability assessment. We allow the weights of the different errors to be asymmetric and different for designer and receiver. The agent's utility is increasing in the receiver's probability estimate. We explain how the problem reduces and under which conditions our theoretical results apply. Finally, we discuss non–reduced applications in different economic contexts: the problem to extract information from an interested party, a certification problem, a task assignment problem, an assessment task, and the design of a justice system. Each application reduces to a game as studied in Section 2, but each has a different structure and highlights different ways of how our theoretical results can be applied.

In Section 2 we study the reduced–form problem for which we give a foundation in Section 3. In Section 4 we discuss five non–reduced applications that are instances of the non–reduced game introduced in Section 3. We discuss our contribution in the context of the related literature in Section 5. The final section concludes.

2 The reduced–form problem

In this section we analyze a reduced–form game. Utility functions are belief–dependent and the equilibrium notion is adapted to the reduced–form environment. We derive structural properties of the optimal test design.

2.1 The model

There is an agent whose quality ω is either good, $\omega = g$, or bad, $\omega = b$, a principal who designs a test, and a receiver who observes the test result. We allow for the interpretation that the role of the receiver is assumed by the player who designs the test, or by a different (set of) player(s). We use the male pronoun for the agent and the female pronoun for the principal and the receiver.

A test is a device that generates a signal about the agent's quality. Test $T = (Z, p_b, p_g)$ is characterized by a finite set $\Sigma^Z := \{1, \ldots, Z\}$ of test results σ and for each ω a probability distribution over test results, $p_\omega := (p_\omega^1, \ldots, p_\omega^Z)$. p_ω^σ is the probability that the test result is σ conditional on the quality ω. We restrict attention to informative tests (i.e., $p_b \neq p_g$), and to test results which occur with a strictly positive probability (i.e., $\forall \sigma : p_b^\sigma + p_g^\sigma > 0$). Let $\mathcal{T}$ be the set containing all such tests. Important subclasses are accurate tests, $\mathcal{T}^a := \{(Z, p_b, p_g) \in \mathcal{T} | \forall \sigma \in \Sigma^Z : p_b^\sigma = 0 \text{ or } p_g^\sigma = 0\}$, and binary tests, $\mathcal{T}^2 := \{(Z, p_b, p_g) \in \mathcal{T} | Z = 2\}$.

The agent's own assessment of his quality is imperfect. He assigns probability $q_\theta^g := \theta$ to being good, and $q_\theta^b := 1 - \theta$ to being bad. He infers q_θ^ω form a private signal θ, which is drawn from a cumulative distribution function F. The support of F, Θ, is a closed subset of $(0, 1)$ with minimal element $\underline{\theta}$ and maximal element $\overline{\theta} > \underline{\theta}$. Let $\theta_0 := \int_\Theta \theta \mathrm{d}F(\theta)$. Participation in the test is voluntary. After observing the test design T, the agent either participates ($d = Y$) or does not participate ($d = N$).

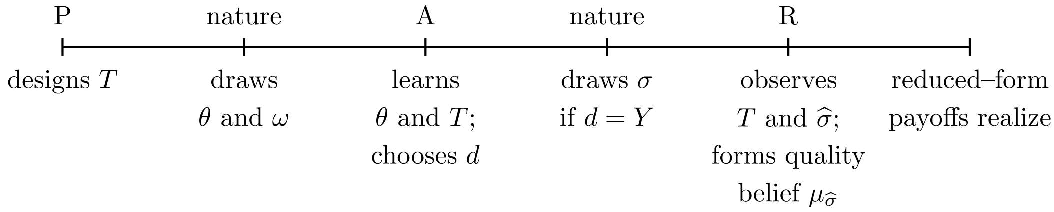


Figure 1: Timeline

If he participates, test result σ is generated with a probability $q_\theta^g p_g^\sigma + q_\theta^b p_b^\sigma$. The participation strategy $x \in \mathcal{X} := \{x : \Theta \to [0,1]\}$ describes the probability of participation for each private signal θ. From an ex ante perspective, participation occurs with probability $P(x) := \int_\Theta x(\theta)\mathrm{d}F(\theta)$. The receiver observes $\widehat{\sigma} = N$ if $d = N$ and the generated test result $\widehat{\sigma} = \sigma$ if $d = Y$.

The reduced–form game ends here and the agent's reduced–form utility is determined by the probability $\mu_{\widehat{\sigma}} \in [0,1]$ that the receiver assigns to the agent being good after observing T and $\widehat{\sigma} \in \widehat{\Sigma}^Z := \{N, 1, \ldots, Z\}$. We refer to $\mu_{\widehat{\sigma}}$ as the receiver's quality belief or as her perception of the agent's quality. Let $\mu := (\mu_N, \mu_1, \ldots, \mu_Z)$. The agent evaluates $\mu_{\widehat{\sigma}}$ at a utility function v_A which exhibits hyperbolic absolute risk aversion (HARA) and maximizes expected utility. For parameters $(c_1, c_2) \in \mathcal{C} := \{(c_1, c_2) | c_1 \in \mathbf{R}, c_2 \in [\max\{-c_1, 0\}, \infty) \text{ if } c_1 \notin \{0,1\}, c_2 \in (0, \infty) \text{ if } c_1 \in \{0,1\}\}$, $v_A : [0,1] \to \mathbf{R}$ is defined by

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

For $(c_1, c_2) \in \mathcal{C}$, v_A is strictly increasing and strictly concave on $(0,1)$. The Arrow–Pratt measure of absolute risk aversion is $r_a(\mu_{\widehat{\sigma}}) := -\frac{v_A''(\mu_{\widehat{\sigma}})}{v_A'(\mu_{\widehat{\sigma}})} = \frac{1}{c_1\mu_{\widehat{\sigma}} + c_2}$.[5] The timeline is illustrated in Figure 1.

The receiver's updating procedure is conducted in two steps. First, information about the agent's private signal as revealed by his participation decision is processed. Then, information about his quality as revealed by the test result is processed. Let x be the supposed participation strategy. Conditional on the agent's participation decision d, the receiver believes that the agent is good with probability $\mu_d(x)$:

$$\mu_Y(x) := \begin{cases} \overline{\theta} & \text{if } P(x) = 0 \\ \frac{\int_\Theta x(\theta) q_\theta^g \mathrm{d}F(\theta)}{P(x)} & \text{if } P(x) > 0 \end{cases}, \quad \mu_N(x) := \begin{cases} \frac{\int_\Theta (1-x(\theta)) q_\theta^g \mathrm{d}F(\theta)}{1-P(x)} & \text{if } P(x) < 1 \\ \underline{\theta} & \text{if } P(x) = 1 \end{cases}$$

The quality beliefs associated to participation and non–participation are determined by Bayes' Law whenever possible.[6] If $d = Y$, the test design is used to interpret the realized test result. Given $\mu_Y(x)$

[5] Special cases are quadratic utility ($c_1 = -1$), cubic utility ($c_1 = -\frac{1}{2}$), exponential/CARA utility ($c_1 = 0$), logarithmic utility ($c_1 = 1$) and CRRA utility ($c_2 = 0$). The parameter space is illustrated in Figure 5 in Appendix A.

[6] Our assumptions on the quality beliefs for which Bayes' Law does not apply can be interpreted as follows: If the agent chooses an action that is supposed to occur with probability zero, the receiver believes that he has the signal that gives him the strongest incentive to choose this action. For any test and for any quality beliefs that are consistent with some participation behavior, the agent's incentive to take the test is the stronger, the larger his private signal (see Lemma A 1 in the proof to Proposition 1).

obtained in the first step, Bayes' Law is applicable.[7] The receiver's quality belief is $\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma) := \frac{\mu_Y(x)p_g^\sigma}{\mu_Y(x)p_g^\sigma + (1-\mu_Y(x))p_b^\sigma}$. Let $\mu(x,T) := (\mu_N(x), \mu_1(\mu_Y(x), p_b^1, p_g^1), \ldots, \mu_Z(\mu_Y(x), p_b^Z, p_g^Z))$.

$(x, \mu) \in \mathcal{X} \times [0,1]^{Z+1}$ is a reduced–form equilibrium for test T if (I) for given μ, $x(\theta)$ is a solution to

$$\max_{x_\theta \in [0,1]} x_\theta V_\theta^Y(T,\mu) + (1-x_\theta)v_A(\mu_N) \text{ with } V_\theta^Y(T,\mu) := \sum_{\omega,\sigma} q_\theta^\omega p_\omega^\sigma v_A(\mu_\sigma) \tag{1}$$

for any $\theta \in \Theta$, and (II) for given x, $\mu = \mu(x,T)$. $\mathcal{E}(T)$ denotes the set of equilibria for test T. Participation strategy x is induced by test T if $(x, \mu(x,T)) \in \mathcal{E}(T)$. x is inducible if it is induced by some test $T \in \mathcal{T}$.

The principal evaluates $\mu_{\widehat{\sigma}}$ at a reduced–form utility function $v_P : [0,1] \to \mathbf{R}$ which is strictly convex on $(0,1)$ and continuous. She chooses a test $T \in \mathcal{T}$ and an equilibrium $(x,\mu) \in \mathcal{E}(T)$ of the game induced by test T to maximize her expected reduced–form utility

$$\mathcal{V}_P(T,x,\mu) := \int_\Theta \left[x(\theta) \left(\sum_{\omega,\sigma} q_\theta^\omega p_\omega^\sigma v_P(\mu_\sigma) \right) + (1-x(\theta))v_P(\mu_N) \right] \mathrm{d}F(\theta).$$

On the principal's preference for learning

From an ex ante perspective, any subsequent updating induces a quality belief lottery with expected quality belief θ_0. If an informative signal about the agent's quality becomes available, the induced quality belief distribution is transformed by a mean–preserving spread. The value of any additional information is strictly positive for the principal if v_P is strictly convex. This relation establishes the link between the convexity of v_P and the principal's preference for information generation, i.e., learning.

Informativeness is measured in terms of the principal's expected utility. It derives from the quality belief lottery induced by a test and the assigned equilibrium. If the choice of the respective equilibrium is clear, we say that a test T' is more informative than a test T'' if the principal attains a higher expected utility under the quality belief lottery induced by T' than under that induced by T''.[8] Note that, if T' is more informative than T'' in the sense of [4], it is more informative in our sense.

For a fixed participation strategy x, we define the concept of accuracy which induces a partial order over tests inducing participation strategy x. If a quality belief lottery associated to a test T' (and equilibrium $(x, \mu(x,T'))$) is obtained from a lottery associated to a test T'' (and equilibrium $(x, \mu(x,T''))$) by a mean–preserving spread, we call T' more accurate than T''. Since v_P is strictly convex, the principal strictly prefers more accurate tests. I.e., a more accurate test is also more informative. The converse holds not. Since v_A is strictly concave, from an ex ante point of view, the agent strictly prefers less accurate tests.

[7] $\Theta \subset (0,1)$ implies $\mu_Y(x) \in (0,1)$ for any $x \in \mathcal{X}$. This together with $p_g^\sigma + p_b^\sigma > 0$ for any σ implies that any test result occurs with a positive probability.

[8] [15] measure the informativeness of the test in terms of the mean squared error of the public's quality estimate. In our setting with a binary quality variable this is equivalent to our non–reduced specification in Section 3 with $\gamma_P = \gamma_R = 1$.

2.2 Equilibrium characterization

In this subsection we derive equilibrium properties for a given test and we characterize the participation strategies that are part of an equilibrium for some test.[9]

A threshold strategy is described by a threshold signal $t \in \Theta$ and the participation probability $r \in [0,1]$ for $\theta = t$. We call $B = (t,r) \in \Theta \times [0,1]$ a (threshold) participation behavior. The threshold strategy associated with B is $x^B : \Theta \to [0,1]$ with $x^B(\theta) = 0$ if $\theta < t$, $x^B(\theta) = r$ if $\theta = t$ and $x^B(\theta) = 1$ if $\theta > t$. For brevity, if the agent has the threshold signal, we call him the threshold participant.

Proposition 1 *Fix any $T \in \mathcal{T}$. (a) $\mathcal{E}(T)$ is non–empty. (b) $(x,\mu) \in \mathcal{E}(T)$ if and only if there exists B such that $\mu = \mu(x^B, T)$ and such that $x^B(\theta)$ is a maximizer of (1) for any $\theta \in \Theta$.*

The proof of (b) makes use of the monotonicity of V_θ^Y in θ. The intuition is the following: For any given test and for any supposed participation strategy, test results that are relatively more often obtained when the agent is good yield higher quality beliefs than test results that are relatively more often obtained when the agent is bad. Conversely, the agent is more likely to get test results that are associated with high quality beliefs when his private signal is higher, i.e. when it is more likely that he is good. The agent's utility associated with non–participation does not depend on his private signal Thus, his participation incentive is the stronger, the higher his private signal (see Lemma A 1 in the proof to the proposition). Only threshold participation strategies can be part of an equilibrium.

$\mathcal{B} := \{B \in \Theta \times [0,1] | \exists T \in \mathcal{T} : (x^B, \mu(x^B,T)) \in \mathcal{E}(T)\}$ is the set of inducible participation behavior. The relation of $\mathcal{B}$ to three further sets is useful: First, participation behavior that is inducible with a binding participation constraint, $\mathcal{B}^b := \{B = (t,r) \in \Theta \times [0,1] | \exists T \in \mathcal{T} : (x^B, \mu(x^B,T)) \in \mathcal{E}(T)$ and $V_t^Y(T, \mu(x^B,T)) = v_A(\mu_N(x^B))\}$. Second, participation behavior that is inducible by an accurate test, $\mathcal{B}^a := \{B \in \Theta \times [0,1] | \exists T \in \mathcal{T}^a : (x^B, \mu(x^B,T)) \in \mathcal{E}(T)\}$. Third, participation behavior that is inducible by a binary test, $\mathcal{B}^2 := \{B \in \Theta \times [0,1] | \exists T \in \mathcal{T}^2 : (x^B, \mu(x^B,T)) \in \mathcal{E}(T)\}$.

Proposition 2 *(a) $\mathcal{B}^b = \{B = (t,r) \in \Theta \times [0,1] | t v_A(1) + (1-t) v_A(0) \leq v_A(\mu_N(x^B))\}$. (b) $\mathcal{B} = \mathcal{B}^a \cup \mathcal{B}^b$. (c) $\mathcal{B} = \mathcal{B}^2$.*

(a) gives a simple characterization of the set $\mathcal{B}^b$. The intuition is as follows: Fix some participation behavior B and consider the quality beliefs $\mu(x^B, T)$. If the threshold participant has a strict incentive to participate if $T \in \mathcal{T}^a$, it is impossible to construct a test that makes him indifferent between participation and non–participation. Any $T \notin \mathcal{T}^a$ yields an even stronger incentive to participate for the threshold participant (see Lemma A 2 in the proof to the proposition).[10] By contrast, if the threshold participant does not have a strict incentive to participate if $T \in \mathcal{T}^a$, a test exists which makes him indifferent. By assumption, $V_t^Y(T', \mu(x^B, T')) \leq v_A(\mu_N(x^B))$ for $T' \in \mathcal{T}^a$. Further, the threshold

[9] In Appendix B we present equilibrium properties for tests in two benchmark cases (the case with uninformative and with perfectly informative private information) and relate them to the case studied here.

[10] An inaccurate test pools good and bad quality on some test result(s) and hence allows the threshold participant a certain extent of pooling with on average better quality. Moreover, for a given participation behavior, the quality belief lottery induced by a less accurate test is spread out less. Both effects make participation more attractive for the threshold participant.

participant has a strict incentive to participate in any sufficiently inaccurate test as this allows him to pool with higher private information at almost no cost. That is, $V_t^Y(T'', \mu(x^B, T'')) \approx v_A(\mu_Y(x^B)) > v_A(\mu_N(x^B))$ for T'' sufficiently inaccurate. Because of a continuity property, there exists a binary, informative but inaccurate test for which the participation constraint is binding for $\theta = t$. Since the agent's expected utility from participation is monotonic in his private signal, this renders participation behavior B inducible with a binding participation constraint.

(b) establishes a relation between the sets $\mathcal{B}$, $\mathcal{B}^a$ and $\mathcal{B}^b$. If a participation behavior is inducible with a non–binding participation constraint, it is possible to make the test more accurate until the participation constraint is binding or the test is accurate. Consequently, any inducible participation behavior is also inducible with a binding participation constraint or by an accurate test.

(c) establishes the equality of sets $\mathcal{B}$ and $\mathcal{B}^2$. By (b), any inducible participation behavior is either inducible by an accurate test or with a binding participation constraint. Any accurate test can be replaced by a binary accurate test without affecting participation incentives, since the quality is binary. Any test that yields a binding participation constraint can be replaced by a binary test that yields a binding participation constraint as argued above.

2.3 Optimal inducement of a given participation behavior

The design of the test has ample effects on learning. First, the induced participation behavior determines the extent of indirect learning about the agent's quality. The participation decision allows for inferences about the agent's private signal, and is hence indirectly informative about the agent's quality. Second, the induced participation behavior determines the probability of observing a test result, i.e., the probability of direct learning. Third, for a fixed participation behavior, the test design determines the informativeness of the test results. In this subsection we concentrate on the third effect, holding the first two effects fixed. We analyze the optimal inducement of a given participation behavior $B \in \mathcal{B}$. The principal chooses a test T from $\mathcal{T}_B := \{T \in \mathcal{T} | (x^B, \mu(x^B, T)) \in \mathcal{E}(T)\}$ to maximize $\mathcal{V}_P(T, x^B, \mu(x^B, T))$.

Binary tests $T \in \mathcal{T}^2$ play an important role in our analysis of optimal test design. They can be interpreted as pass–fail–tests where $\sigma = 2$ means "pass" and $\sigma = 1$ means "fail". A binary test is completely characterized by the probabilities $p_g^2 \in (0, 1]$ and $p_b^2 \in [0, p_g^2)$ with which a good and a bad agent passes, respectively. If $p_b^2 = 0$ and $p_g^2 = 1$, the test is accurate. The test result perfectly reveals the agent's quality. If $p_b^2 = 0$, the test is not subject to false positives as a bad agent never passes. Passing the test perfectly reveals that the agent is good. If $p_g^2 = 0$, the test is not subject to false negatives as a good agent never fails. Failing the test perfectly reveals that the agent is bad. If $p_b^2 > 0$ and $p_g^2 < 1$, the test is subject to false positives and to false negatives.

We identify clear–cut effects of the test design on the agent's ad interim expected utility (Lemma 2) and on the variance of the induced quality belief distribution (Lemma 3). The principal tends to like a high variance (because she is information loving) and she tends to like a low expected utility of the agent (because the agent is in contrast to her information averse). Whether the effect of the test design on the principal's expected utility is as clear–cut as for the agent depends on the specifics of her utility function v_P. Proposition 3 presents optimal test design results for two simple classes of utility functions v_P for which the effects are clear–cut.

Proposition 3 *Fix $B \in \mathcal{B}$. (a) Suppose that the principal maximizes a non–positive function of the agent's ad interim expected utility. If $c_1 \leq \frac{1}{2}$, a binary test which is not subject to false positives is optimal. If $c_1 \geq \frac{1}{2}$, a binary test which is not subject to false negatives is optimal. (b) Suppose that the principal maximizes the variance of the induced quality belief distribution. If $c_1 \geq -1$, a binary test which is not subject to false positives is optimal.*

Under the supposition in (a), the structure of the optimal test depends on c_1 but not on c_2, Θ, F and B. c_1 determines whether v_A exhibits increasing absolute risk aversion ($c_1 < 0$), constant absolute risk aversion ($c_1 = 0$) or decreasing absolute risk aversion ($c_1 > 0$). Higher absolute values of c_1 are associated with stronger changes of absolute risk aversion. If the agent's absolute risk aversion is not too strongly decreasing, a binary test which is not subject to false positives is optimal. Otherwise, a binary test which is not subject to false negatives is optimal. Under the supposition in (b), the result depends on the third derivative of v_A as $c_1 \geq -1$ is equivalent to $v_A'''(\mu_{\widehat{\sigma}}) \geq 0$ for HARA utility functions. Again, c_2, Θ, F and B do not affect the structure of the optimal test. If the third derivative of v_A is non–negative, a test which is not subject to false positives is optimal.

Proposition 3 follows directly from the subsequently introduced Lemma 1, 2 and 3 by imposing the specific structure of the considered classes of utility functions. The intuition for the proposition is provided in the discussion of the three lemma from which it follows. Proposition 3 directly applies to the non–reduced problems presented in Subsections 4.1 and 4.4. The proposition is not directly applicable to the remaining applications in Section 4, where the optimal test design follows from Lemma 2 (Subsection 4.5), from Lemma 3 (Subsection 4.3) and from a combination of both lemmas (Subsection 4.2).

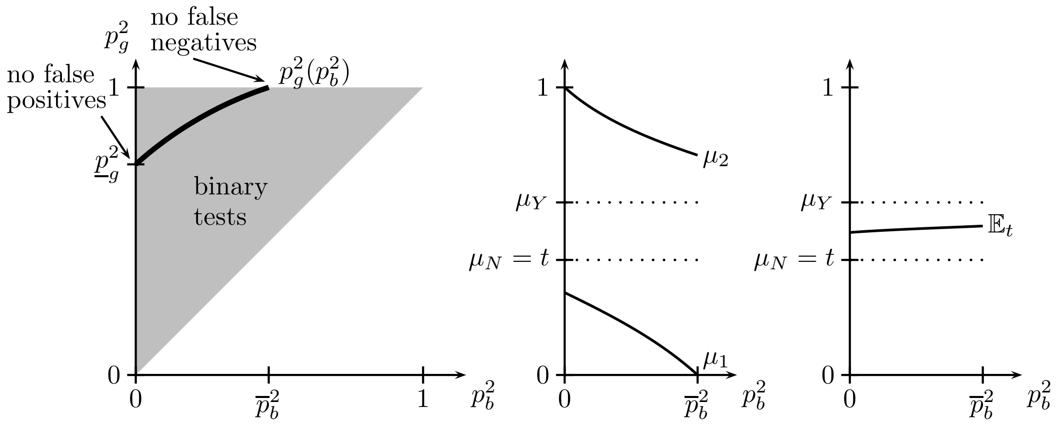
From Proposition 2 (b), we know that the participation behavior B can either be induced by an accurate test or with a binding participation constraint. The following lemma establishes that restricting attention to these kinds of inducement is without loss of generality.

Lemma 1 *Fix $B = (t, r) \in \mathcal{B}$. (a) If $B \in \mathcal{B}^a$, the optimal test inducing B is accurate. (b) If $B \in \mathcal{B} \backslash \mathcal{B}^a$, $\mathcal{T}_B^b := \{T \in \mathcal{T} | (x^B, \mu(x^B, T)) \in \mathcal{E}(T)$ and $V_t^Y(T, \mu(x^B, T)) = v_A(\mu_N(x^B))\}$ contains an optimal test.*

If the participation behavior B can be induced by an accurate test, such a test is optimal as it is more accurate than any other test inducing B. In particular, the binary accurate test is optimal. As such a test is neither subject to false positives nor to false negatives, the statement in Proposition 3 is true for $B \in \mathcal{B}^a$. If B cannot be induced by an accurate test, it is optimal to have the participation constraint binding. The proof is by contradiction. For any test T that induces B with a slack participation constraint, it is possible to modify the test such that it still induces B, but is more accurate. So, $\mathcal{T}_B^b := \{T \in \mathcal{T} | (x^B, \mu(x^B, T)) \in \mathcal{E}(T)$ and $V_t^Y(T, \mu(x^B, T)) = v_A(\mu_N(x^B))\}$ contains an optimal test.

It remains to consider participation behavior $B \in \mathcal{B} \backslash \mathcal{B}^a$ and to identify the optimal test in $\mathcal{T}_B^b$. We first consider the effect of the test design on the agent's ad interim expected utility. Since we keep the participation behavior fixed and consider tests for which the participation constraint is binding, the test design affects the agent's ad interim expected utility only for private signals $\theta > t$.

Lemma 2 *Fix $B = (t, r) \in \mathcal{B} \backslash \mathcal{B}^a$ and consider tests from $\mathcal{T}_B^b$. Consider any $\theta > t$. If $c_1 < \frac{1}{2}$, $V_\theta^Y(T, \mu(x^B, T))$ is minimized by a binary test which is not subject to false positives. If $c_1 = \frac{1}{2}$,*



(a) Tests $\mathcal{T}_B^b \cap \mathcal{T}^2$ (b) Quality beliefs (c) Expected perception

Figure 2: Binary tests inducing B with a binding participation constraint and properties thereof [figures are for $\theta \sim U[\frac{2}{5}, \frac{4}{5}]$, $t = \frac{2}{5}$, $v_A(\mu_{\widehat{\sigma}}) := -(1-\mu_{\widehat{\sigma}})^2$]

$V_\theta^Y(T, \mu(x^B, T))$ is not affected by the test design. If $c_1 > \frac{1}{2}$, $V_\theta^Y(T, \mu(x^B, T))$ is minimized by a binary test which is not subject to false negatives.

We first give an intuition for the case that the agent's utility is quadratic (i.e., $c_1 = -1$) and the principal chooses among binary tests (i.e., $T \in \mathcal{T}_B^b \cap \mathcal{T}^2$), then we explain how the result extends. From the viewpoint of the agent with private signal θ, his participation induces a quality belief lottery with expected value $\mathbb{E}_\theta := \sum_\sigma [\theta p_g^\sigma + (1-\theta) p_b^\sigma] \mu_\sigma$ and variance $\mathbb{V}_\theta := \sum_\sigma [\theta p_g^\sigma + (1-\theta) p_b^\sigma](\mu_\sigma - \mathbb{E}_\theta)^2$. As v_A is quadratic, the agent's ad interim expected utility is completely determined by the expected perception $\mathbb{E}_\theta$ (+) and the variance of perception $\mathbb{V}_\theta$ (–). The participation constraint requires that $-\frac{1}{2}(c_2 - \mathbb{E}_t)^2 - \frac{1}{2}\mathbb{V}_t$ equals a constant. There is a continuum of binary tests that induce the participation behavior B with a binding participation constraint. More formally, $\mathcal{T}_B^b \cap \mathcal{T}^2 = \{(2, (1-p_b^2, p_b^2), (1-p_g^2, p_g^2)) | p_b^2 \in [0, \overline{p}_b^2]$ and $p_g^2 = p_g^2(p_b^2)\}$ for some $\overline{p}_b^2 \in (0,1)$ and some strictly increasing function $p_g^2 : [0, \overline{p}_b^2] \to (0,1]$. See Figure 2(a) for an illustration. The test choice implies a trade–off between the occurrence of false positives and false negatives. Choosing a test with a higher p_b^2 has three consequences: First, false positives are more frequent and false negatives are less frequent. Second, the quality beliefs associated to both test results decrease (see Figure 2(b)).[11] Third, the expected perception of the threshold participant increases (see Figure 2(c)).[12]

Due to the third consequence, we can view the test choice as a choice of the expected perception of the threshold participant $\mathbb{E}_t$. Key to the result is how this choice affects the expected perception

[11] To keep the participation incentive for the threshold participant intact, a more severe stigma of failure must be compensated with a higher probability with which a bad agent passes the test. This deteriorates also the quality belief associated with passing the test. A formal proof of the statement proceeds along the lines of the proof of Lemma 5 in [31].

[12] A rough intuition is obtained by looking at the polar cases: If $p_b^2 = 0$, a good agent is correctly identified with a certain probability, otherwise he is pooled with *all* bad agents. If $p_b^2 = \overline{p}_b^2$, a bad agent is correctly identified with a certain probability, otherwise he is pooled with *all* good agents. With the second kind of pooling, a bad agent is on average pooled with "more" good agents. He is on average perceived as better. Consequently, an agent who is relatively likely to be bad, e.g., the agent with the threshold signal, is on average perceived as better with the second kind of pooling, i.e. when p_b^σ is higher. A formal proof is analogous to the proof of Lemma 6 in [31].

$\mathbb{E}_\theta$ and the variance of perception $\mathbb{V}_\theta$ for $\theta > t$. Consider first the effect on $\mathbb{E}_\theta$. Using that quality beliefs obey Bayes' Law, we can write $\mathbb{E}_\theta = \mu_Y + (\theta - \mu_Y)(\mathbb{E}_g - \mathbb{E}_b)$ where $\mathbb{E}_\omega := p_\omega^2 \mu_2 + (1 - p_\omega^2)\mu_1$ is the expected perception of the agent conditional on his quality being ω.[13] An increase in $\mathbb{E}_t$ requires a "redistribution of expected perception" from $\omega = g$ to $\omega = b$. This generates an increase of expected perception when it is relatively likely that $\omega = b$ ($\theta < \mu_Y$) and a decrease of expected perception when it is relatively likely that $\omega = g$ ($\theta > \mu_Y$).

Effect 1: $\mathbb{E}_t \uparrow \Rightarrow \mathbb{E}_\theta \uparrow$ *if* $\theta < \mu_Y$ *and* $\mathbb{E}_\theta \downarrow$ *if* $\theta > \mu_Y$.

There are two effects on the variance $\mathbb{V}_\theta = \mathbb{V}_t + [(\mu_1 + \mu_2) - (\mathbb{E}_t + \mathbb{E}_\theta)](\mathbb{E}_\theta - \mathbb{E}_t)$.[14] The first effect is as follows: An increase in $\mathbb{E}_t$ relaxes the participation constraint. To restore indifference of the threshold participant, the test has to impose a higher perception risk on him. $\mathbb{V}_t$ increases. As $\mathbb{V}_t$ feeds directly into $\mathbb{V}_\theta$, this suggests that the agent faces a quality belief lottery with a higher variance for all $\theta > t$.

Effect 2: $\mathbb{E}_t \uparrow \Rightarrow$ *first summand of* $\mathbb{V}_\theta \uparrow$.

However, there is a "distributional effect" which is opposite. $\mathbb{V}_\theta$ is a convex combination of the two squared errors $(\mu_1 - \mathbb{E}_\theta)^2$ and $(\mu_2 - \mathbb{E}_\theta)^2$. A small increase of $\mathbb{E}_t$ is accompanied by a small change of $\mathbb{E}_\theta$ and a strong decrease of both quality beliefs, μ_1 and μ_2. $(\mu_1 - \mathbb{E}_\theta)^2$ increases while $(\mu_2 - \mathbb{E}_\theta)^2$ decreases. The decreasing error term is relatively more likely for private signals $\theta > t$ than for private signal t while the converse is true for the increasing error term. This suggests that $\mathbb{V}_t$ and $\mathbb{V}_\theta$ move apart.

Effect 3: $\mathbb{E}_t \uparrow \Rightarrow$ *second summand of* $\mathbb{V}_\theta \downarrow$.

Summing up, the agent's ad interim expected utility is unambiguously affected positively by Effect 3 and negatively by Effect 2. The direction of Effect 1 depends on the agent's private signal. Surprisingly, the third, positive effect generally dominates the other two effects. Heuristically, a small change in $\mathbb{E}_t$ causes a small change in $\mathbb{E}_\theta$ (follows from Bayes' Law) and a small change in $\mathbb{V}_t$ (follows from the participation constraint). Effects 1 and 2 are weak. By contrast, the changes in actual quality beliefs μ_1 and μ_2 are comparably large. Effect 3 is strong. It follows that for any private signal $\theta > t$ the agent's ad interim expected utility is minimized by choosing $\mathbb{E}_t$ as small as possible. As the test with the smallest possible $\mathbb{E}_t$ is not subject to false positives (see Figure 2), we obtain our result.

We now explain how the result extends. We start with the optimal binary test for general HARA utility functions. Consider the set of binary tests inducing B. The crucial property is that, comparing any two tests in this set, the one associated with a higher μ_1 also induces a higher μ_2. If the agent's preferences exhibit strongly increasing absolute risk aversion (i.e., if c_1 is low), he prefers the quality belief lottery with relatively lower outcomes, and vice versa for strongly decreasing absolute risk aversion (i.e., for high c_1). We know that for quadratic utility ($c_1 = -1$), the agent's ad interim expected utility is minimized by a binary test which induces the highest possible quality beliefs. For lower values of c_1, the

[13]Bayes' Law implies $\mathbb{E}_{\mu_Y} = \mu_Y$ for any test. By the definition of $\mathbb{E}_\omega$, $\mathbb{E}_g - \mathbb{E}_b = (p_g^2 - p_b^2)(\mu_2 - \mu_1)$. Using this, $\mathbb{E}_\theta = \mathbb{E}_{\mu_Y} + (\theta - \mu_Y)(p_g^2 - p_b^2)(\mu_2 - \mu_1) = \mu_Y + (\theta - \mu_Y)(\mathbb{E}_g - \mathbb{E}_b)$.

[14]We have $\mathbb{V}_\theta = [\theta p_g^2 + (1-\theta)p_b^2]\mu_2^2 + [\theta(1-p_g^2) + (1-\theta)(1-p_b^2)]\mu_1^2 - \mathbb{E}_\theta^2 = \mathbb{V}_t + (\theta - t)(p_g^2 - p_g^1)[\mu_2^2 - \mu_1^2] + (\mathbb{E}_t^2 - \mathbb{E}_\theta^2)$. Using $\mathbb{E}_\theta - \mathbb{E}_t = (\theta - t)(p_g^2 - p_b^2)(\mu_2 - \mu_1)$, $\mathbb{V}_\theta = \mathbb{V}_t + [(\mu_1 + \mu_2) - (\mathbb{E}_t + \mathbb{E}_\theta)](\mathbb{E}_\theta - \mathbb{E}_t)$.

agent's preferences are even more pronounced, and the same type of test is optimal. As c_1 increases, the quality belief lotteries with high outcomes become relatively more attractive for the agent. At $c_1 = \frac{1}{2}$, the agent's preferences over the set of tests reverses. Our result regarding the optimal test design jumps from one extreme (highest possible quality beliefs/no false positives) to the other (lowest possible quality beliefs/no false negatives).

It remains to argue why the test that minimizes the agent's ad interim expected utility is binary. We give the argument for $c_1 < \frac{1}{2}$, the argument for $c_1 > \frac{1}{2}$ is analogous. The crucial property is that for any binary test inducing B with $\mu_1 < \mu_2 < 1$ we can modify the test such that B is still induced, both quality beliefs increase and the agent's ad interim expected utility decreases. For any test with the property that $\mu_{\sigma'} < \mu_{\sigma''} < 1$ for some test results σ' and σ'', we can use a similar argument. By considering modifications of this test such that only the conditional probabilities associated to test results σ' and σ'' change, the design problem becomes basically binary. A test for which two quality beliefs are strictly smaller than one cannot be optimal. A binary test must be optimal.

Finally, we consider the effect of the test design on the moments of the induced quality belief distribution, $\mathbb{M}^k(T,x) := \int_\Theta \left[x(\theta)\left(\sum_{\omega,\sigma} q_\theta^\omega p_\omega^\sigma \mu_\sigma(x,T)^k\right) + (1-x(\theta))\mu_N(x)^k\right] \mathrm{d}F(\theta)$.

Lemma 3 *Fix $B = (t,r) \in \mathcal{B}\backslash\mathcal{B}^a$ and consider $T \in \mathcal{T}_B^b$. If $c_1 \geq -1$, $\mathbb{M}^2(T,x)$ and $\mathbb{M}^3(T,x)$ are both maximized by a binary test which is not subject to false positives.*

The first moment is not affected by the test design because of Bayes' Law. Maximization of the second moment is equivalent to maximization of the ex ante variance. If the agent's utility function is quadratic ($c_1 = -1$), maximization of the variance and minimization of the agent's ex ante expected utility are equivalent. For any $\theta \leq t$ the agent's ad interim expected utility is not affected by the test design and for any $\theta > t$ his ad interim expected utility is minimized by the same test. This test also minimizes the agent's ex ante expected utility. Hence, for $c_1 = -1$ the result regarding the second moment follows directly from Lemma 2. It extends to $c_1 > -1$. Interestingly, for $c_1 > \frac{1}{2}$, for any private signal $\theta > t$, the agent prefers the test from $\mathcal{T}_B^b$ that is associated with the highest variance.[15] The variance is highest for the test that induces the highest possible quality beliefs. As the agent's utility function exhibits decreasing absolute risk aversion, it is possible that his preference for high quality beliefs dominates his aversion against a high variance.

The second moment result extends to the third moment. High quality beliefs are relatively more important for the third than for the second moment. As the second moment is maximized by a test for which both quality beliefs are highest, the result carries over.

2.4 Optimal participation behavior in the quadratic–quadratic case

Having understood how to optimally induce inducible participation behavior, we now address the questions (i) which participation behavior is inducible, (ii) how to induce more participation, and (iii) what is the link between the relative importance of indirect learning, as reflected in the optimal participation

[15]The statement about what the agent likes *most* follows from the proof to Lemma 2.

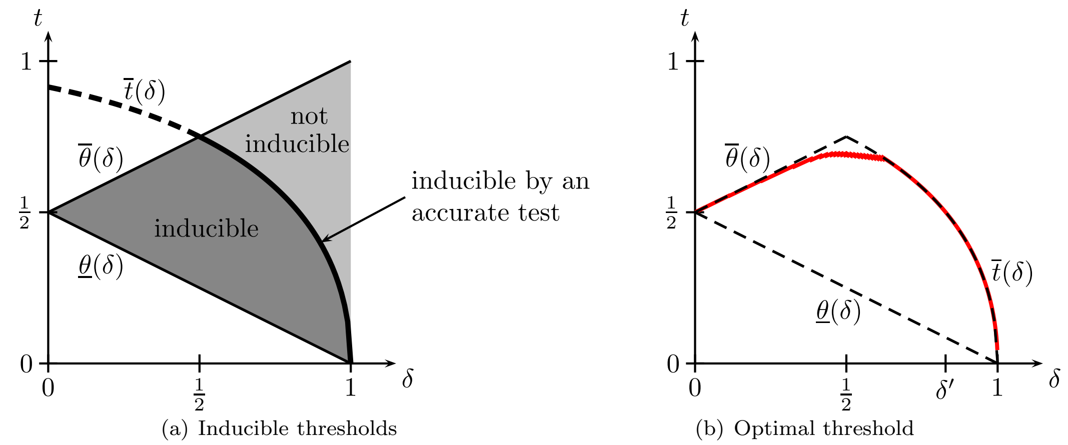


Figure 3: Participation thresholds $[\theta \sim U[\frac{1}{2}(1-\delta), \frac{1}{2}(1+\delta)]$, $v_A(\mu_{\widehat{\sigma}}) := -\frac{1}{2}(1-\mu_{\widehat{\sigma}})^2]$

behavior, and the accuracy of the agent's private information. Our answers to these questions are summarized in Results 1, 2 and 3. The formal derivation of the statements in this subsection is relegated to the appendix.

To get cleaner effects, we impose additional structure: First, the agent's utility is quadratic. That is, $v_A(\mu_{\widehat{\sigma}}) = -\frac{1}{2}(c_2 - \mu_{\widehat{\sigma}})^2$ with $c_2 \geq 1$. Second, the principal's utility is also quadratic. From Proposition 3 we know that any participation behavior is optimally induced by a binary test which is not subject to false positives.[16] We thus consider tests $T(p_f) := (2, (1, 0), (p_f, 1-p_f))$ with $p_f \in [0, 1)$, which can be interpreted as pass–fail–tests. The agent always fails when he is bad and he fails with probability p_f when he is good. By Lemma 1, the optimal test is accurate if $B \in \mathcal{B}^a$ and the participation constraint is binding for the optimal test if $B \in \mathcal{B} \backslash \mathcal{B}^a$. For any $B \in \mathcal{B}$, this pins down a unique optimal p_f. The principal's design problem reduces to the problem of choosing a participation behavior $B \in \mathcal{B}$. Third, we assume that the agent's private signal is uniformly distributed around $\frac{1}{2}$.[17] That is, $\theta \sim U[\underline{\theta}(\delta), \overline{\theta}(\delta)]$ with $\delta \in (0, 1)$, $\underline{\theta}(\delta) := \frac{1}{2}(1-\delta)$ and $\overline{\theta}(\delta) := \frac{1}{2}(1+\delta)$. Note that an increase in δ transforms the distribution of the agent's private signal by a mean–preserving spread. The agent's signal is more accurate for higher values of δ. As the distribution is continuous, the mixing probability r is not important.[18] Without loss of generality, we can restrict attention to participation behaviors $B = (t, 1)$ with $t \in \Theta$ and $t < \overline{\theta}$.

From Proposition 2 and the distribution of private information assumed here, we obtain a characterization of the set of inducible participation thresholds. The characterization depends on the function $\overline{t}(\delta) := -(2c_2 - 2 + \underline{\theta}(\delta)) + \sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)}$. $\overline{t}(\delta)$ is continuous and strictly decreasing. As $\lim_{\delta \downarrow 0} \overline{t}(\delta) > \lim_{\delta \downarrow 0} \overline{\theta}(\delta)$ and $\lim_{\delta \uparrow 1} \overline{t}(\delta) = \lim_{\delta \uparrow 1} \underline{\theta}(\delta)$, there exists a unique intersection of $\overline{t}(\delta)$ and $\overline{\theta}(\delta)$. The intersection is at $\delta = \frac{1}{4c_2 - 2}$. Moreover, $\overline{t}(\delta) > \underline{\theta}(\delta)$.

[16] In this specification, the principal seeks to maximize the variance of the quality belief lottery, and her utility is a negative function of the agent's utility. Both parts of Proposition 3 apply.

[17] In a working paper version of this paper ([31]), we analyze the problem for arbitrary binary distributions and obtain analogous results.

[18] First, if participation threshold $t < \overline{\theta}$ is inducible with some mixing probability $r \in [0, 1)$, it is also inducible with mixing probability $r = 1$. Second, the principal's expected utility does not depend on the mixing probability.

Result 1: *If $\delta \in (0, \frac{1}{4c_2-2})$, any $t \in [\underline{\theta}(\delta), \overline{\theta}(\delta)]$ is inducible. No t is inducible by an accurate test. If $\delta \in [\frac{1}{4c_2-2}, 1)$, any $t \in [\underline{\theta}(\delta), \overline{t}(\delta)]$ is inducible, any $t \in (\overline{t}(\delta), \overline{\theta}(\delta)]$ is not inducible. Only $t = \overline{t}(\delta)$ is inducible by an accurate test.*

Figure 3(a) displays the set of inducible participation thresholds for $c_2 = 1$. If the threshold participant decides not to participate, his utility is $v_A(\mathbf{E}[\theta|\theta < t])$. His utility is bounded from above by $v_A(\frac{1}{2})$. If he participates in an accurate test, he obtains expected utility $tv_A(1) + (1-t)v_A(0)$. As the accurate test is the least attractive test for the threshold participant, his expected utility is bounded from below by this amount, which is close to $tv_A(1)$ for t sufficiently close to 1. As $tv_A(1) > v_A(\frac{1}{2})$ for sufficiently high t, the presumed threshold participant has a strict incentive to participate in any test. Hence, for sufficiently high t, participation threshold t is not inducible.[19] Whether "sufficiently high" signals exist depends on δ. $\overline{t}(\delta)$ describes the probability of being good, t, that makes the agent indifferent between the perfect revelation of his quality and the quality belief $\mu_N = \mathbf{E}[\theta|\theta < t]$. If the accuracy of the agent's signal is low, there is no "sufficiently high" private signal (i.e., $\overline{\theta}(\delta) < \overline{t}(\delta)$). Any participation threshold can be induced. If the accuracy of the agent's private signal is high, sufficiently high private signals exist (i.e., $\overline{\theta}(\delta) > \overline{t}(\delta)$). Only participation thresholds $t \leq \overline{t}(\delta)$ can be induced.

Applying the Implicit Function Theorem to the participation constraint tells us how to adjust the test to foster participation.

Result 2: *Higher participation ($t \downarrow$) is induced by a harder test ($p_f \uparrow$).*

The harder a test is to pass, the lower is the stigma of failure. Holding the participation behavior fixed, making the test harder to pass benefits the threshold participant, who is most likely to fail, and gives him a strict incentive to participate. The participation threshold must adjust until indifference is reinstalled. The direction of adjustment is not obvious as the average quality in the pool of participants and in the pool of non–participants change into the same direction. For uniformly distributed signals and quadratic preferences, the threshold participant's participation incentive decreases as participation increases. Consequently, participation increases until indifference is reinstalled.

Finally, we address the issue of optimal participation and the importance of indirect learning. For the considered class of distributions, for any given accuracy of private information δ, learning through private information is better the closer the participation threshold is to $\frac{1}{2}$. If the participation threshold lies very close to $\underline{\theta}(\delta)$ or to $\overline{\theta}(\delta)$, there is almost no learning through private information.

The red solid line in Figure 3(b) displays the optimal participation threshold in our graphical example.[20] If the accuracy of the agent's private signal is low, any participation threshold is inducible. As signals are spread out in a small interval around $\frac{1}{2}$, the worst possible quality belief associated to non–participation is quite high, and the agent's signalling incentive is weak. Fostering participation is costly (in terms of test accuracy). The principal prefers little participation in a relatively accurate test to higher participation in a much less accurate test. The optimal participation threshold lies close to $\overline{\theta}(\delta)$. If the accuracy of the agent's private signal is high, signalling incentives are strong as beliefs

[19] If the agent has a strict incentive to participate when $\theta = t$, he also has a strict incentive to participate for private signals just below t. As the distribution is continuous, such signals exist, contradicting that the participation threshold is t.

[20] We compute the optimal participation threshold numerically using Maple.

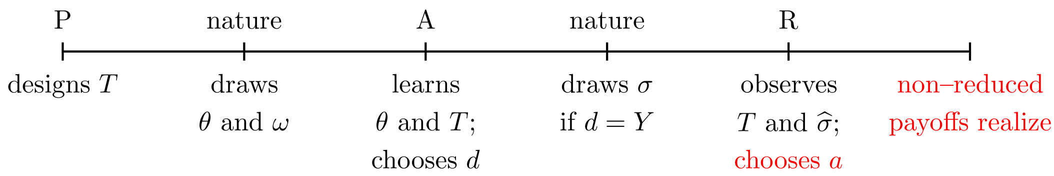


Figure 4: Modified timeline

associated to non–participation are low. An unravelling effect kicks in. Although the imperfection of private information precludes full unravelling for an accurate test, there is almost full unravelling for $\underline{\theta}(\delta)$ close to zero. Only participation thresholds close to $\underline{\theta}(\delta)$ are inducible. An accurate test and therewith the highest inducible participation threshold is optimal. Summing up, for a low accuracy of the private information, there is not much indirect learning as the optimal threshold lies close to $\overline{\theta}(\delta)$. For a high accuracy of the private information, there is not much indirect learning as it lies close to $\underline{\theta}(\delta)$. By a continuity property, there exist intermediate levels of accuracy where indirect learning works quite well (e.g., $\delta = \delta'$ in Figure 3(b)). We obtain the following comparative statics result.

Result 3: *A more accurate private signal ($\delta \uparrow$) may lead to less learning through private information.*

3 A non–reduced problem

In this section we introduce a non–reduced problem and an adapted equilibrium notion with standard properties. We show how the problem reduces and how the results from Section 2 can be applied to infer optimal test design results. Rather than aiming at maximum generality, we set up our non-reduced model such that it allows for a number of interpretations and it has nice analytical properties (see Section 4).

3.1 The model

The non–reduced model differs from the reduced–form model in two aspects. First, the reduced–form game is augmented by a receiver game which is played after the receiver learns $\widehat{\sigma} \in \widehat{\Sigma}^Z$. The receiver game leads to an outcome $a \in [0,1]$. For the ease of exposition, we consider a single receiver who takes action $a \in [0,1]$, the extension to more complex receiver games is straightforward. The receiver's strategy $y \in \mathcal{Y} := \{y : \widehat{\Sigma}^Z \to [0,1]\}$ assigns an action $a = y(\widehat{\sigma})$ to each observation $\widehat{\sigma}$. Second, utility functions are non–reduced and depend on a and ω. The receiver's utility is $\widetilde{v}_R(a,\omega) := -(a-1)^2$ if $\omega = g$ and $\widetilde{v}_R(a,\omega) := -\gamma_R(a-0)^2$ if $\omega = b$ with $\gamma_R \in (0,\infty)$. The principal's utility is $\widetilde{v}_P(a,\omega) := -(a-1)^2$ if $\omega = g$ and $\widetilde{v}_P(a,\omega) := -\gamma_P(a-0)^2$ if $\omega = b$ with $\gamma_P \in [\frac{1}{2}\gamma_R, 2\gamma_R]$. The agent's utility function $\widetilde{v}_A(a)$ is twice continuously differentiable and is strictly increasing on $(0,1)$. The modified timeline is displayed in Figure 4.

Each observation $\widehat{\sigma} \in \widehat{\Sigma}^Z$ constitutes an information set of the receiver. A system of beliefs is described by a joint distribution over private signals and quality realizations for any information set. We denote the probability with which the agent's private signal is believed to be weakly smaller than

θ and his quality is believed to be ω when the observation is $\widehat{\sigma}$ by $F^{\omega}_{\widehat{\sigma}}(\theta)$. If $P(x) > 0$, Bayes' Law is applicable to compute $F^{\omega}_{\widehat{\sigma}}$ for any $\widehat{\sigma} \in \Sigma^Z$. If $P(x) < 1$, Bayes' Law is applicable to compute F^{ω}_N. If Bayes' Law is applicable,

$$F^{\omega}_{\widehat{\sigma}}(\tilde{\theta}) = \frac{\text{Prob}\{\theta \leq \tilde{\theta} \text{ and } \omega|\widehat{\sigma}\}}{\text{Prob}\{\widehat{\sigma}\}} = \begin{cases} \frac{\int_{\Theta\cap[0,\tilde{\theta}]} q^{\omega}_{\theta} p^{\sigma}_{\omega} x(\theta) \mathrm{d}F(\theta)}{\int_{\Theta}[q^g_{\theta} p^{\sigma}_g + q^b_{\theta} p^{\sigma}_b] x(\theta) \mathrm{d}F(\theta)} & \text{if } \widehat{\sigma} = \sigma \in \Sigma^Z \\ \frac{\int_{\Theta\cap[0,\tilde{\theta}]} q^{\omega}_{\theta}(1-x(\theta)) \mathrm{d}F(\theta)}{\int_{\Theta}(1-x(\theta)) \mathrm{d}F(\theta)} & \text{if } \widehat{\sigma} = N \end{cases}. \tag{2}$$

A system of beliefs $\{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z}$ is admissible if there exist sequences x^Y_n and x^N_n in $\mathcal{X}$ such that (i) for any n, $P(x^Y_n) > 0$ and $P(x^N_n) < 1$, (ii) for any $\sigma \in \Sigma^Z$, F^{ω}_{σ} is obtained as the limit of (2) with $x = x^Y_n$, and (iii) F^{ω}_N is obtained as the limit of (2) with $x = x^N_n$. We denote the set containing all admissible systems of beliefs by $\mathcal{F}$.

$(x, y, \{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{F}$ is an equilibrium for test $T \in \mathcal{T}$ if (I') for given y, $x(\theta)$ solves the problem $\max_{x_{\theta}\in[0,1]} x_{\theta} \left(\sum_{\omega,\sigma} q^{\omega}_{\theta} p^{\sigma}_{\omega} \widetilde{v}_A(y(\sigma))\right) + (1 - x_{\theta})\widetilde{v}_A(y(N))$ for any $\theta \in \Theta$, (II') for given $\{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z}$, $y(\widehat{\sigma})$ solves the problem $\max_{a\in[0,1]} \sum_{\omega} \int_{\Theta} \widetilde{v}_R(a, \omega) \mathrm{d}F^{\omega}_{\widehat{\sigma}}(\theta)$ for any $\widehat{\sigma}$, and (III') $\{(F^g_{\sigma}, F^b_{\sigma})\}_{\sigma\in\Sigma^Z}$ is determined by Bayes' Law when $P(x) > 0$ and (F^g_N, F^b_N) is determined by Bayes' Law when $P(x) < 1$. (I') and (II') are sequential rationality requirements. (III') pins down beliefs for observations $\widehat{\sigma}$ that occur with a positive probability in equilibrium. Moreover, implicit in the definition of $\mathcal{F}$ is a weak consistency requirement. Beliefs after participation and non–participation must be consistent with some (possibly different) participation behavior.

The principal chooses a test $T \in \mathcal{T}$ and an equilibrium $(x, y, \{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z})$ of the game induced by T to maximize her expected utility

$$\widetilde{\mathcal{V}}_P(T, x, y) := \int_{\Theta} \left[x(\theta) \left(\sum_{\omega,\sigma} q^{\omega}_{\theta} p^{\sigma}_{\omega} \widetilde{v}_P(y(\sigma), \omega)\right) + (1 - x(\theta)) \left(\sum_{\omega} q^{\omega}_{\theta} \widetilde{v}_P(y(N), \omega)\right)\right] \mathrm{d}F(\theta).$$

3.2 Reduction and inference of the optimal test and equilibrium

The problem is reduced by solving for the receiver's optimal action. Her optimal action depends on the probability that she assigns to the agent being good, $\mu_{\widehat{\sigma}} = \int_{\Theta} \mathrm{d}F^g_{\widehat{\sigma}}(\theta)$. If she believes that the agent is good with probability $\mu_{\widehat{\sigma}}$, her optimal action is $a = \frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}}+(1-\mu_{\widehat{\sigma}})\gamma_R}$, and the agent's utility is

$$v_A(\mu_{\widehat{\sigma}}) \quad := \quad \widetilde{v}_A\left(\frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}} + (1 - \mu_{\widehat{\sigma}})\gamma_R}\right). \tag{3}$$

The principal's expected utility conditional on the receiver's assessment of the probability that the agent is good $\mu_{\widehat{\sigma}}$ and conditional on her assessment being correct, is

$$\begin{aligned} v_P(\mu_{\widehat{\sigma}}) \quad &:= \quad \mu_{\widehat{\sigma}} \widetilde{v}_P\left(\frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}} + (1 - \mu_{\widehat{\sigma}})\gamma_R}, g\right) + (1 - \mu_{\widehat{\sigma}})\widetilde{v}_P\left(\frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}} + (1 - \mu_{\widehat{\sigma}})\gamma_R}, b\right) \\ &= \quad \frac{-\mu_{\widehat{\sigma}}(1 - \mu_{\widehat{\sigma}})}{(\mu_{\widehat{\sigma}} + (1 - \mu_{\widehat{\sigma}})\gamma_R)^2} (\mu_{\widehat{\sigma}}\gamma_P + (1 - \mu_{\widehat{\sigma}})\gamma^2_R). \end{aligned} \tag{4}$$

For any $\mu_{\widehat{\sigma}} \in (0, 1)$, $v'_A(\mu_{\widehat{\sigma}}) > 0$ and $v''_P(\mu_{\widehat{\sigma}}) > 0$. If $\widetilde{v}_A$ and γ_R are such that $-\frac{v''_A(\mu_{\widehat{\sigma}})}{v'_A(\mu_{\widehat{\sigma}})} = \frac{1}{c_1\mu_{\widehat{\sigma}}+c_2}$ for some $(c_1, c_2) \in \mathcal{C}$, then the distributional assumptions of the non–reduced model together with utility functions v_A and v_P specify a reduced–form model as introduced in Subsection 2.1. We obtain the following result:

Proposition 4 *Let $\widetilde{v}_A$ and γ_R be such that v_A as defined in (3) is a HARA utility function. Consider the non–reduced problem and the reduced–form problem that is specified by utility functions (3) and (4), and the same distribution of information as in the non–reduced problem. A test is optimal for the reduced problem if and only if it is optimal for the non–reduced problem.*

An important consequence of the result is that the restriction that we imposed in Subsection 2.1 on quality beliefs associated with zero probability events is without loss of generality. The proof consists of two parts. First, for any fixed test and for any equilibrium of the reduced–form game, we construct an equilibrium of the non–reduced game that attains the same expected utility for the principal. Second, for any fixed test and for any equilibrium of the non–reduced game that induces participation with a positive probability, we construct an equilibrium of the reduced–form game that attains the same expected utility for the principal. As inducing participation with probability zero can not be optimal for the principal, we obtain the proposition.

4 Applications

In this section we discuss five applications. Each application highlights a specific interpretation of the non–reduced problem introduced in the preceding section and induces a reduced–form problem with a specific structure. We indicate the structure of the induced reduced–form utility functions of the agent and the principal in the brackets of the subsection headings. Each subsection starts with the assumptions on the non–reduced utility functions. Then we give an interpretation of the problem and we demonstrate how the problem reduces and how our theoretical results can be applied.

4.1 Extracting information from an interested party (quadratic/quadratic)

$\gamma_R = 1$. $\gamma_P = 1$. $\widetilde{v}_A(a) = -\frac{1}{2}(1-a)^2$.

The first application can be interpreted as a problem between an uninformed decision maker (who assumes the roles of the principal and the receiver) and a better informed party with partisan interests (the agent). The decision maker chooses $a \in [0,1]$. There are two states of the world, $\omega = g$ and $\omega = b$. The decision maker prefers $a = 1$ in the first state and $a = 0$ in the second state, the interested party always prefers $a = 1$. Decisions in–between are compromises between the extreme decisions. The interested party is in possession of decision–relevant data that allows him an imperfect assessment of the probability that $\omega = g$. He knows that the probability of state $\omega = g$ is θ, while the decision maker could infer the true state if she was given access to the data. The test is a request to investigate the data in a particular way. For instance, she might ask an assessor to search a certain part of the data base for a certain time for indicators that the true state is $\omega = g$. The interested party can accept or decline the request.

The receiver's optimal action is $a = \mu_{\widehat{\sigma}}$ such that $v_P(\mu_{\widehat{\sigma}}) = -\mu_{\widehat{\sigma}}(1-\mu_{\widehat{\sigma}})$ and $v_A(\mu_{\widehat{\sigma}}) = -\frac{1}{2}(1-\mu_{\widehat{\sigma}})^2$ (= HARA utility with $c_1 = -1$ and $c_2 = 1$). As both reduced–form utility functions are quadratic, a binary test which is not subject to false positives is optimal by either Proposition 3 (a) or (b). The optimal participation behavior results from Subsection 2.4 apply.

It is optimal for the decision maker to request a search for evidence that the decision preferred by the interested party is also optimal for her.

4.2 Certification (cubic/cubic)

The receiver is a market that pays a price a to the agent that equals the quality belief. $\gamma_P \in [\frac{1}{2}, 2]$. $\widetilde{v}_A(a) = -\frac{1}{12}(1-a)^3$.

The agent is the seller of an object. The object's quality is either good or bad. For instance, the object might be a painting which is either authentic or forged, or a technology which is useful or useless. The receiver is a market that pays the expected quality as a price to the seller. The seller is only interested in generating revenue and he is risk averse in money. He has imperfect private information about the object's quality and can voluntarily apply for certification. The principal is a monopolistic certifier who is installed by the market such that she is interested in the accuracy of the market's inference independently of whether the seller applies for certification or not. She wants the market to infer the quality correctly. We allow for the possibility that she is biased in one or the other direction. If $\gamma_P > 1$, she suffers relatively more from too high inferences. If $\gamma_P < 1$, she suffers relatively more from too low inferences. The test describes the certification technology.

The receiver's optimal action is $a = \mu_{\widehat{\sigma}}$ such that $v_P(\mu_{\widehat{\sigma}}) = -\mu_{\widehat{\sigma}}(1-\mu_{\widehat{\sigma}})[(1-\mu_{\widehat{\sigma}}) + \gamma_P \mu_{\widehat{\sigma}}]$ and $v_A(\mu_{\widehat{\sigma}}) = -\frac{1}{12}(1-\mu_{\widehat{\sigma}})^3$ (= HARA utility with $c_1 = -\frac{1}{2}$ and $c_2 = \frac{1}{2}$). v_A exhibits positive skewness. v_P is strictly convex on $(0,1)$. We distinguish two cases: Case (i): $\gamma_P \geq 1$. We have $v_P(\mu_{\widehat{\sigma}}) = -(1-\gamma_P)\mu_{\widehat{\sigma}}^3 + (2-\gamma_P)\mu_{\widehat{\sigma}}^2 - \mu_{\widehat{\sigma}}$. Because of Bayes' Law, the expected value of $\mu_{\widehat{\sigma}}$ cannot be affected by the test design. Since the coefficients of $\mu_{\widehat{\sigma}}^3$ and of $\mu_{\widehat{\sigma}}^2$ are both non–negative, it follows from Lemma 3 that a test which is not subject to false positives is optimal. Case (ii): $\gamma_P < 1$. The coefficient of $\mu_{\widehat{\sigma}}^3$ is in this case strictly negative while the coefficient of $\mu_{\widehat{\sigma}}^2$ is strictly positive. There are countervailing effects. The compound effect of the test design does not follow directly from Lemma 3. We can rewrite the principal's reduced–form utility as $v_P(\mu_{\widehat{\sigma}}) = -12(1-\gamma_P)v_A(\mu_{\widehat{\sigma}}) + (2\gamma_P - 1)\mu_{\widehat{\sigma}}^2 + ((2-3\gamma_P)\mu_{\widehat{\sigma}} - (1-\gamma_P))$. By Lemma 2, the expected value of the first summand is maximized by a binary test which is not subject to false positives. By Lemma 3, the expected value of the second summand is maximized by a binary test which is not subject to false positives. By Bayes' Law, the expected value of the last summand is constant. In sum, a binary test which is not subject to false positives is optimal.

A binary certification technology that either awards a certificate or not is optimal. A bad product is never certified. A good product is sometimes not certified. The probability of false negatives is chosen to induce the optimal level of application for certification. The direction of the certifier's bias γ_P does not affect the structure of the optimal certification technology.

4.3 Job assignment (HARA/cubic)

$\gamma_R = 1$. $\gamma_P \in [1, 2]$. $\widetilde{v}_A(a)$ *is HARA utility with* $c_1 \in [-1, \infty)$.

There is a worker (the agent), a manager (the receiver) and the owner of a firm (the principal). The manager assigns a job $a \in [0,1]$ to the worker. Jobs are a combination of routine tasks and cognitive tasks. The lower a, the higher the share of routine tasks. An exogenously given tariff determines the worker's payment $t_w(a) = a$ if he is assigned to job a. The worker is risk averse in money and he maximizes his expected utility. He evaluates his payment at a HARA utility function which exhibits positive skewness. The worker is of type $\omega = g$ or $\omega = b$. Type g is able to develop ideas, type b is not able to develop ideas. The worker has imperfect private information about his type. Type b accomplishes the assigned job according to the job description. Type g develops an idea of an alternative way to accomplish the task. It is impossible to prevent that a worker carries out his idea, which gives him a warm glow. If the manager assigns job a to the worker, she bears a private supervision cost $c_S(a) = a^2$. Moreover, the worker's initiative facilitates the supervisor's coordination tasks if he works on a job $a > \frac{1}{2}$, and makes coordination more difficult for jobs $a < \frac{1}{2}$. The manager's private coordination cost is $c_C(a, \omega) = 1 - 2a$ if $\omega = g$, and $c_C(a, \omega) = 0$ if $\omega = b$. The manager receives a fixed payment t_m. She assigns a job a to the worker so as to minimize the expected value of her cost $c_S(a) + c_C(a, \omega) = -\widetilde{v}_R(a, \omega)$. The owner of the firm installs a test, but it is at the worker's discretion to participate. The owner's revenue is $R(a, \omega) = Y + a - \gamma_P a^2$, if $\omega = b$, and $R(a, \omega) = Y + 3a - a^2 - 1$, if $\omega = g$. The owner's aim is to maximize expected profits, i.e., the expected value of revenues minus payments $R(a, \omega) - (t_w(a) + t_m) = Y + \widetilde{v}_P(a, \omega)$.

The receiver's optimal action is $a = \mu_{\widehat{\sigma}}$ such that $v_A(\mu_{\widehat{\sigma}}) = \widetilde{v}_A(\mu_{\widehat{\sigma}})$ and $v_P(\mu_{\widehat{\sigma}}) = Y - \mu_{\widehat{\sigma}}(1 - \mu_{\widehat{\sigma}})[(1 - \mu_{\widehat{\sigma}}) + \gamma_P \mu_{\widehat{\sigma}}]$. v_A exhibits HARA with $c_1 > -1$. v_P is cubic, strictly convex on $(0, 1)$ and exhibits positive skewness. By Lemma 3, a binary test which is not subject to false positives is optimal for the principal. By Lemma 2, the agent rather prefers a binary test which is not subject to false negatives if $c_1 < \frac{1}{2}$, while the preferences of the principal and the agent are aligned if $c_1 \geq \frac{1}{2}$.

The optimal test can be implemented, e.g., by exposing the agent to a problem which he can solve only if he is able to develop ideas. The occurrence of false negatives (i.e., failure to solve the problem by a worker who is able to develop ideas) can be controlled with the time given to solve the problem.

4.4 Assessment with asymmetric error costs (HARA with $c_1 = \frac{1}{2}$/non–cubic)

$\gamma_R \neq 1$. $\gamma_P = \gamma_R$. $\widetilde{v}_A(a) = a$.

There is a student (the agent) who is either good or bad. He requests a letter of recommendation from a professor (who is the principal and also the receiver). The professor's action is a statement about the probability that the student is good. The student strives for a favorable statement. The professor is interested in her reputation for making accurate statements, but she is either biased towards careful statements (if $\gamma_R > 1$) or towards exaggeration (if $\gamma_R < 1$). That is, while her optimal assessment is $a = 1$ if the agent is good and $a = 0$ if the agent is bad, she suffers asymmetrically from too high or too low statements. The student has an imperfect assessment of his quality and the professor can learn about the student's quality by talking to him. The test describes the professor's questioning practice. For instance, she may ask a tough question, which the student can only answer correctly if he is good, or she may ask an easy question. It is at the student's discretion to give a risky answer which is informative about his quality or to evade the professor's question.

The receiver's optimal action is $a = \frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}} + (1-\mu_{\widehat{\sigma}})\gamma_R}$ such that $v_P(\mu_{\widehat{\sigma}}) = \frac{-\gamma_R \mu_{\widehat{\sigma}}(1-\mu_{\widehat{\sigma}})}{\mu_{\widehat{\sigma}} + (1-\mu_{\widehat{\sigma}})\gamma_R}$ and $v_A(\mu_{\widehat{\sigma}}) = \frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}} + (1-\mu_{\widehat{\sigma}})\gamma_R}$. If $\gamma_R > 1$, v_A is strictly increasing on $(0, 1)$ and strictly convex on $(0, 1)$. If $\gamma_R < 1$, v_A exhibits HARA with parameters $c_1 = \frac{1}{2}$ and $c_2 = \frac{1}{2}\frac{\gamma_R}{1-\gamma_R}$. The principal's reduced–form utility is strictly convex for any γ_R. We distinguish two cases: Case (i): $\gamma_R > 1$. The principal and the agent are both information loving. Full participation can be induced by an accurate test. Perfect learning of the agent's quality is possible. Case (ii): $\gamma_R < 1$. The principal is information loving while the agent is information averse. Fix any $B \in \mathcal{B}$ and consider tests in $\mathcal{T}_B$. Since v_A exhibits HARA with $c_1 = \frac{1}{2}$, his expected utility is not affected by the test design (see Lemma 2). If $\gamma_R = \frac{1}{2}$, $v_P(\mu_{\widehat{\sigma}}) = \mu_{\widehat{\sigma}} - v_A(\mu_{\widehat{\sigma}})$. In this case it follows directly from our theoretical results how the principal's expected utility is affected.[21] The expected value of $\mu_{\widehat{\sigma}}$ is not affected by the test design because of Bayes' Law. The expected value $v_A(\mu_{\widehat{\sigma}})$ is not affected by Lemma 2. It follows that the test design neither affects the agent's nor the principal's expected utility.

If $\gamma_R > 1$ (i.e., the professor is biased towards being careful), it is optimal for her to pose a question which the student is capable to answer correctly if and only if he is good. The student always tries to give an informative answer. If $\gamma_R = \frac{1}{2}$ (i.e., the professor is biased towards exaggeration), perfect learning of the student's quality is not possible. Posing a single binary question is optimal. The question does not perfectly reveal the student's quality and/or the student may try to evade the question. Whether the question is easy or hard does not matter. It is only important that there is the right degree of inaccuracy to induce the desired answering behavior.

4.5 Justice system (HARA with $c_1 = \frac{1}{2}$/non–cubic)

For our last application, our theoretical results simplify the problem. The simplified problem is solved explicitly under a distributional assumption.

$\gamma_R = \frac{1}{2}$. $\gamma_P = 1$. $\widetilde{v}_A(a) = a$. $\theta \sim U[\frac{1}{4}, \frac{3}{4}]$.

A judge (the receiver) decides about the fate of a defendant (the agent). The judge chooses $a \in [0, 1]$. $a = 1$ is an acquittal and $a = 0$ is the highest possible sentence. The lower a, the higher the sentence. The defendant prefers to be acquitted and he is risk averse in the sentence. The judge prefers an acquittal if the defendant is innocent and the highest possible sentence if he is guilty. Facing uncertainty, the judge has an attitude "in dubio pro reo", and a bias towards low sentences. An unbiased legislature (the principal) determines the rules of the interrogation procedure in court. The defendant has imperfect private information about his guilt. He may know what he has done, but he can only imperfectly assess the legal consequences of his actions. He can participate in the interrogation procedure or make use of his right to remain silent.

The receiver's optimal action is $a = \frac{2\mu_{\widehat{\sigma}}}{1+\mu_{\widehat{\sigma}}}$ such that $v_A(\mu_{\widehat{\sigma}}) = \frac{2\mu_{\widehat{\sigma}}}{1+\mu_{\widehat{\sigma}}}$ and $v_P(\mu_{\widehat{\sigma}}) := \frac{-\mu_{\widehat{\sigma}}(1-\mu_{\widehat{\sigma}})(3\mu_{\widehat{\sigma}}+1)}{(1+\mu_{\widehat{\sigma}})^2}$. v_A exhibits HARA with parameters $c_1 = \frac{1}{2}$ and $c_2 = 1$. Fix any inducible participation behavior. We can write $v_P(\mu_{\widehat{\sigma}}) = 3\mu_{\widehat{\sigma}} - 2v_A(\mu_{\widehat{\sigma}}) - v_A(\mu_{\widehat{\sigma}})^2$. The expected value of $\mu_{\widehat{\sigma}}$ is not affected by the test design because of Bayes' Law. Since the agent's utility function exhibits HARA with $c_1 = \frac{1}{2}$, the expected value of $v_A(\mu_{\widehat{\sigma}})$ is also not affected by the test design by Lemma 2. It follows that the principal strives for minimizing the expected value of $v_A(\mu_{\widehat{\sigma}})^2$. For continuous distributions, only the threshold signal t matters. By the distributional assumption, $\mu_N = \frac{1}{8} + \frac{1}{2}t$ and $\mu_Y = \frac{3}{8} + \frac{1}{2}t$. By

[21] If $\gamma_R \neq \frac{1}{2}$, we need also to analyze how $\mu_{\widehat{\sigma}} v_A(\mu_{\widehat{\sigma}})$ is affected by the test design.

Proposition 2 (a), any $t \in [\frac{1}{4}, \frac{1}{8}(\sqrt{33}-1)]$ is inducible with a binding participation constraint. Only non–participation is inducible by an accurate test. Restriction to the case where the participation constraint is binding is without loss of generality. Consider all binary tests inducing the fixed participation threshold $t \in [\frac{1}{4}, \frac{1}{8}(\sqrt{33}-1)]$ with a binding participation constraint. If the quality beliefs are μ_1 and μ_2, it follows from Bayes' Law that they are induced by test $(2,(1-p_b^2, p_b^2),(1-p_g^2, p_g^2))$ with $p_b^2 = \frac{1-\mu_2}{1-\mu_Y}\frac{\mu_Y-\mu_1}{\mu_2-\mu_1}$ and $p_g^2 = \frac{\mu_2}{\mu_Y}\frac{\mu_Y-\mu_1}{\mu_2-\mu_1}$. Using this, the participation constraint depends only on quality beliefs and on t. It follows from this constraint how μ_2 depends on μ_1 and on t, say $\mu_2(\mu_1, t)$.[22] Using this, we obtain how the expected value of $v_A(\mu_{\widehat{\sigma}})^2$ depends on μ_1 and on t.[23] Since the expected value of $v_A(\mu_{\widehat{\sigma}})^2$ is for any t strictly decreasing in μ_1, the principal's expected utility is the larger the larger μ_1. The optimal binary test is not subject to false positives.

It is optimal for the legislature to rather admit questions that aim at unburdening the defendant in court than questions that aim at burdening him.

5 Discussion

Our analysis aims to improve the understanding of the limits to information generation and the optimal way of generating information. It is complementary to the analysis in [18], [15] and [13], which have been discussed in the introduction.

In our model, test results are hard information, but their availability and their interpretation are endogenous. Imperfect private information precludes complete unravelling as in [14], [24] and [26]. The interplay between the accuracy of public signals and the interpretation of actions also plays a role in the analysis in [38], [7] and [1]. With [35] and [30] we share the insight that endogenous participation also determines the value of non–participation.

Our analysis adds insights to the study of several applications which imply information generation, such as obtaining an education, matching workers to jobs, applying for certification, and evidence provision in front of a court.

[33] focuses on the signalling role of education, i.e., the transmission of information that is already known to the signalling individual. [34] and [38] point to the role of the schooling choice as a costly signal and an information generating device at the same time, and to the fact that the precision of the test determines the signalling cost. Acknowledging both sources of information, we identify the test with maximal overall informativeness.[24]

[20], [5] and [37] study information generation incentives prior to matching a worker to a task. The information generation technology is given exogenously and involves an exogenously given cost. Our

[22] More specifically, $\mu_2(\mu_1, t) = \frac{1}{8}\frac{-104t\mu_1-96\mu_1+128t^2\mu_1+128t^3\mu_1+103t-64t^4-28t^2-112t^3+66}{16t\mu_1+32t^2\mu_1-42\mu_1+13t+12-16t^2-16t^3}$.

[23] More specifically, $\mathbf{E}[v_A(\mu_{\widehat{\sigma}})^2] = 4[2048t^6\mu_1^2 - 1280t^4\mu_1^2 + 1280t^3\mu_1^1 - 4536t^2\mu_1^2 + 1008t\mu_1^2 + 1584\mu_1^2 - 1024t^7\mu_1 + 1280t^6\mu_1 - 1408t^5\mu_1 + 96t^4\mu_1 + 8188t^3\mu_1 - 7563t^2\mu_1 - 2394t\mu_1 + 2880\mu_1 - 1024t^7 + 256t^6 + 384t^5 - 2848t^4 + 2204t^3 + 3345t^2 - 1323t - 1134]/[(3t-54+48t^2+16t^3)(24t\mu_1-432\mu_1+384t^2\mu_1+128t^3\mu_1+207t-64t^4-156t^2-240t^3+162)(\mu_1+1)]$

[24] [27] point out that schools with the intention to maximize the average placement of their students have an incentive to issue noisy reports of the student's ability if the mapping of the student's expected ability to job desirability is concave. In our setting with voluntary participation, a school may have to issue a noisy report even if the goal is to maximize informativeness.

analysis suggests an approach to an endogenous information generation technology and an endogenous cost of information generation.

The endogenous cost of information generation is a central element of the literature on certification.[25] Certifiers are intermediaries with the ability to observe a product's quality and to provide hard information about it. In this literature, the certifier's information endowment is typically exogenous, and the price of his service is endogenous. Our analysis instead focuses on the effects of the test design on the participation behavior and the feedback effects on the test's informativeness. Noisy signals are a means to extract and to generate more information, whereas certifiers may strategically commit to noisy reports to extract more profit from the sellers (e.g., [19]).

The literature on certification has focused on the case of perfect private information (e.g., [19]) or symmetric information (e.g., [10]). For many products, it seems reasonable that the producer is not perfectly informed about the quality of his product, but that he has an informational advantage. In the light of imperfect private information, the content of the information that a certifier would release is not perfectly foreseeable. Aversion towards this kind of uncertainty is relevant for the certifier's optimal disclosure policy. A renegotiation-proof contract leads to perfect revelation of the certifier's information if the choice of hiring a certifier is observable ([11]). If the certifier cannot commit to a disclosure rule, he may instead strategically choose his information endowment. This calls for introducing an endogenous information acquisition technology, i.e., a test, into the analysis. [8] show that (exogenously) imperfect certification induces higher participation. Our analysis offers insights how the precision of the information that becomes available to the certifier can be used to optimally control the participation behavior.[26]

The generation of hard information has also been studied in the context of evidence provision to a court. [32] and [23] consider litigation expenditures by the defendant, and by the prosecutor, respectively, while [39] studies a model in which both sides invest in evidence.[27] [29] compare evidence collection in inquisitorial and adversarial systems. In all these contributions, the evidence production technology is exogenous. Our analysis takes the view on evidence provision from a constitutional stage at which the technology derives endogenously from the rules imposed on admissible evidence (such as the "Federal Rules of Evidence" applicable to proceedings in U.S. courts). We offer insights into how such rules affect the information available at court.

A defendant typically has an explicit right to remain silent, and the inference of his guilt from non–cooperation in information generation is not admissible. Evidence of guilt is needed for a conviction. However, facing residual uncertainty, society tends to stigmatize an individual under the suspect of having committed a crime. Even though a formal punishment through a court may not be possible in

[25] For a survey, see [9].

[26] Note that imperfections can have adverse effects on product quality. [21] show that the pooling of bad and good quality by an imperfect testing device can induce sellers of high quality products to exit the market.

[27] The analysis in [2] implies an interesting complementarity between both parties' information provision. In their model, the court's action is binary, and the defendant is interested in maximizing the probability that the court's posterior exceeds a certain threshold. If the court's prior is favorable and the prosecutor's signal is sufficiently noisy, even an innocent defendant may be averse towards providing additional (noisy) information because of the risk that the posterior falls below the threshold. The more accurate the prosecutor's information provision, the better the information provision incentive for the defendant.

the absence of hard evidence, an informal punishment may be induced. Imperfect, private information about the own guilt is plausible in many situations. Consider, e.g., a defendant who is accused of sexual harassment. Although he may know what he has done, he may be uncertain about which of his acts were potentially offensive. The rules for evidence provision in front of a court impact on both, the court's ability to reach correct verdicts, and the level of trust and cooperation in society. Moreover, we anticipate an interesting interrelation between the publicity of a case, its transparency, and the defendant's preferences for information generation.

In our model, the receiver unambiguously benefits from additional information, as it allows her to make better decisions, while the agent does not react to new information. The agent's preferences over the receiver's information derive from his preferences over the receiver's action. An interesting next step would be to study voluntary information generation in a setting with strategic interactions, i.e., in which both, the ex ante less informed party and the better informed party take actions after additional information becomes available. [28] show that public revelation of information that is affiliated to the buyer's valuation is beneficial for a monopolist.[28] The buyer loses part of his information rent,[29] but he may benefit from efficiency gains. His preferences with respect to information generation are ambiguous form an ex ante point of view. To the best of our knowledge, papers that study information generation in mechanism design settings have so far assumed that the information structure can unilaterally be imposed. Under the assumption that the buyer's agreement is not required for generating information, the optimal accuracy of the affiliated signal in [28] is not a very interesting object of study, as it is maximal. It would be interesting to study their model under the assumption of voluntary participation. Looking at the buyer's preferences for information from an ad interim point of view, the seller may be able to exploit a signalling motive on the buyer's side and to induce voluntary participation even if he is information averse.[30]

There is evidence that information aversion[31] is an obstacle to information generation in the context of medical tests. [36] present a study in which pregnant women were offered either a nuchal translucency measurement or a maternal serum test. 47 and 72 percent of them, respectively, decided not take the test. Being afraid of unfavorable results was one of the major reasons stated for declining the test. Likewise, the empirical evidence on participation behavior in genetic tests cited in [17] suggests that some patients are information averse. [6] show that the optimal AIDS–test with information averse patients is imperfect, in order to limit the anxiety involved with the test, but informative enough to induce a socially desirable matching pattern. Our analysis allows for conclusions about the optimal accuracy of a medical test, if the aim is to optimize the information the physician has about his patient. We do not take a stand on whether this is a socially desirable objective, as it ignores the patient's right for privacy. This is an ethical question beyond the scope of our contribution.

---

[28] [25] show that public revelation of affiliated information is beneficial for the seller in auction settings.

[29] If information generation gives rise to private information, the effect on the buyers information rent is opposite in the auction settings with private values, as studied in [3] and in [12].

[30] In a setting of trade with state–contingent claims, all parties may be averse to information generation due to the "Hirshleifer effect" ([16]) that information has on the volume of trade.

[31] An example for a setting in which the seller of a product with uncertain quality is information averse is provided in [22]. See also [10].

6 Conclusion

This paper studies optimal test design under voluntary participation of an imperfectly informed, information averse individual when the designer benefits from information. In order to optimally induce a desired participation level, the test may have to issue false negatives, but it does not issue false positives. We have demonstrated that our reduced–from specification can be derived in a variety of settings. In our analysis, a restriction on information generation derives from an imperfection of private information in conjunction with information aversion. We think that the presence of both is plausible in many interesting economic problems.

A Parameter space of HARA utility

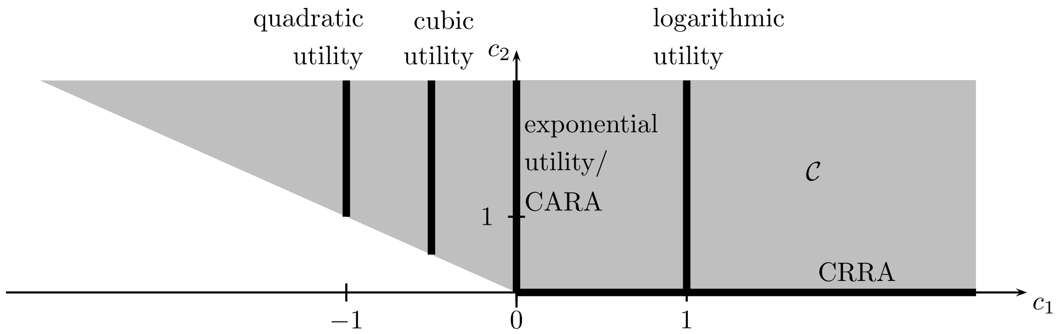


Figure 5: Parameter space of HARA utility and special cases

B Relation to benchmark cases

It is useful to relate the case with an informative but imperfect private signal (i.e., $\Theta \subset (0, 1)$ and $\underline{\theta} < \overline{\theta}$) to two benchmark cases wich stand outside our modeling framework, the case that the agent's private signal is uninformative (i.e., $\Theta = \{\theta_0\}$ and $\theta_0 \notin \{0, 1\}$) and the case that it is perfectly informative about his quality (i.e., $\Theta = \{0, 1\}$).

With an uninformative private signal, the agent does not have a signalling motive. The sole effect of participation in a test is that additional, unbiased information is generated and learnt by the receiver. The agent's information aversion gives him a strict incentive not to participate in any informative test. Learning about the agent's quality is impossible without an informational advantage on the agent's side. In the second polar case, if the agent perfectly knows his quality, adverse beliefs lead to unravelling. A full participation equilibrium exists for any test. An accurate test allows to learn the agent's quality perfectly.

Proposition 5 *(a) Fix any $T \in \mathcal{T}$. If $\Theta = \{\theta_0\}$ and $\theta_0 \notin \{0, 1\}$, there exists a unique equilibrium. There is no participation and no learning. (b) If $\Theta = \{0, 1\}$, an accurate test is optimal. There exists a continuum of equilibria which all perfectly reveal the agent's quality.*

Proof. (a) Consider $\Theta = \{\theta_0\}$ and any test $T \in \mathcal{T}$. For any $x \in \mathcal{X}$, $\mu_Y(x) = \mu_N(x) = \theta_0$. By participating, the agent incurs a non–degenerate quality belief lottery with expected quality belief $\sum_\sigma(\theta_0 p_g^\sigma + (1-\theta_0)p_b^\sigma)\mu_\sigma(\theta_0, p_b^\sigma, p_g^\sigma) = \sum_\sigma \theta_0 p_g^\sigma = \theta_0$. By not participating, the quality belief is $\mu_N(x) = \theta_0$ for sure. The result follows from strict concavity of v_A.

(b) Consider $\Theta = \{0,1\}$ and the accurate test $(2,(1,0),(0,1))$. (x,μ) with $\mu = (0,0,1)$, $x(1) = 1$ and any $x(0) \in [0,1]$ specifies an equilibrium which perfectly reveals the agent's quality. By Lemma A 3 (a) in the proof to Lemma 1, a test which perfectly reveals the agent's quality, i.e., where the induced quality belief distribution has support $\{0,1\}$, is optimal for the principal. q.e.d.

With an informative but imperfect private signal, learning about the agent's quality is possible but imperfect. Learning is possible because the presence of superior private information gives rise to a signalling motive. The test serves as signalling device. If the test is sufficiently inaccurate, risk aversion does not have much bite such that signalling costs are low. The agent participates with a positive probability and information is generated. Learning is imperfect because the imperfectness of private information bounds the quality belief associated to non–participation from below, and therewith limits the agent's signalling motive. Perfect learning requires an accurate test and full participation. However, at least when the agent has the lowest possible private signal, he prefers the receiver's inference that he has the lowest private signal to the perfect revelation of his quality. Perfect learning is impossible because there is either less than full participation or the test is inaccurate.

C Proofs

Proof to Proposition 1

The proof uses the following lemma.

Lemma A 1 $\forall T \in \mathcal{T} \forall x \in \mathcal{X} : V_\theta^Y(T, \mu(x,T))$ *is strictly increasing in* θ.

Proof. We have $V_\theta^Y(T,\mu(x,T)) = \sum_\sigma(\theta p_g^\sigma + (1-\theta)p_b^\sigma)v_A(\mu_Y(x), p_b^\sigma, p_g^\sigma)$ such that

$$\begin{aligned}
& \frac{\mathrm{d}}{\mathrm{d}\theta} V_\theta^Y(T,\mu(x,T)) \\
= & \sum_\sigma (p_g^\sigma - p_b^\sigma) v_A(\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma)) \\
= & \sum_{\sigma'} (p_g^{\sigma'} \sum_{\sigma''} p_b^{\sigma''} - p_b^{\sigma'} \sum_{\sigma''} p_g^{\sigma''}) v_A(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) \\
= & \sum_{\sigma'} \sum_{\sigma'' \neq \sigma'} (p_g^{\sigma'} p_b^{\sigma''} - p_b^{\sigma'} p_g^{\sigma''}) v_A(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) \\
= & \sum_{\sigma'} \sum_{\sigma'' > \sigma'} (p_g^{\sigma'} p_b^{\sigma''} - p_b^{\sigma'} p_g^{\sigma''})(v_A(\mu_{\sigma'}(\mu_Y(x), p_b^{\sigma'}, p_g^{\sigma'})) - v_A(\mu_{\sigma''}(\mu_Y(x), p_b^{\sigma''}, p_g^{\sigma''})))
\end{aligned}$$

$\mu_{\sigma'} < (\leq) \mu_{\sigma''}$ is equivalent to $p_g^{\sigma'} p_b^{\sigma''} - p_b^{\sigma'} p_g^{\sigma''} < (\leq) 0$. It follows that each summand is weakly positive. By the assumption that the test is informative, at least one summand is strictly positive. Hence, $\frac{\mathrm{d}}{\mathrm{d}\theta} V_\theta^Y(T,\mu(x,T)) > 0$. q.e.d.

(a) For $w \in [0,1]$ define

$$\tau(w) := \begin{cases} \min\{F^{-1}(w)\} & \text{if } w \in F(\Theta) \\ \min\{F^{-1}(\min\{\widehat{w} \in F(\Theta) | \widehat{w} \geq w\})\} & \text{if } w \notin F(\Theta) \end{cases}$$

and

$$\rho(w) := \begin{cases} 1 & \text{if } w \in F(\Theta) \\ \frac{w - \sup\{\widehat{w} \in F(\Theta) \cup \{0\} | \widehat{w} \leq w\}}{\min\{\widehat{w} \in F(\Theta) | \widehat{w} \geq w\} - \sup\{\widehat{w} \in F(\Theta) \cup \{0\} | \widehat{w} \leq w\}} & \text{if } w \notin F(\Theta) \end{cases}.$$

Further, define $B : [0,1] \to \Theta \times [0,1]$ by $B(w) := (\tau(w), \rho(w))$. By construction, $B(0) = (\underline{\theta}, 1)$ if there is no point of mass at $\underline{\theta}$, $B(0) = (\underline{\theta}, 0)$ if there is a point of mass at $\underline{\theta}$, $B(1) = (\overline{\theta}, 1)$ and $\forall w \in [0,1] : P(x^{B(w)}) = 1 - w$. $\mu_Y(x^{B(w)})$ and $\mu_N(x^{B(w)})$ are continuous in w.

We consider three cases: Case (i): If $V^Y_{\tau(0)}(T, \mu(x^{B(0)}, T)) - v_A(\mu_N(x^{B(0)})) \geq 0$, it is optimal for the agent to participate when his signal is $\tau(0) = \underline{\theta}$. By Lemma A 1, he has a strict incentive to participate for any signal $\theta > \underline{\theta}$. Hence, $(x^{B(0)}, \mu(x^{B(0)}, T)) \in \mathcal{E}(T)$. Case (ii): If $V^Y_{\tau(1)}(T, \mu(x^{B(1)}, T)) - v_A(\mu_N(x^{B(1)})) \leq 0$, it is optimal for the agent not to participate when his signal is $\tau(1) = \overline{\theta}$. By Lemma A 1, he strictly prefers not to participate for any signal $\theta < \overline{\theta}$. Hence, $(x^{B(1)}, \mu(x^{B(1)}, T)) \in \mathcal{E}(T)$. Case (iii): Suppose neither the condition in (i) nor that in (ii) holds. Since $\mu_Y(x^{B(w)})$ and $\mu_N(x^{B(w)})$ are continuous in w, $V^Y_{\tau(w)}(T, \mu(x^{B(w)}, T)) - v_A(\mu_N(x^{B(w)}))$ is continuous in w. By the Intermediate Value Theorem, there exists $w' \in (0,1)$ such that $V^Y_{\tau(w')}(T, \mu(x^{B(w')}, T)) - v_A(\mu_N(x^{B(w')})) = 0$. It follows from Lemma A 1, that the agent has a strict incentive not to participate if $\theta < \tau(w')$ and a strict incentive to participate if $\theta > \tau(w')$. It follows $(x^{B(w')}, \mu(x^{B(w')}, T)) \in \mathcal{E}(T)$.

(b) "$\Rightarrow$" Suppose $(x, \mu) \in \mathcal{E}(T)$. By (II), $\mu = \mu(x, T)$. Lemma A 1 implies that (I) can only be satisfied for threshold strategies. "$\Leftarrow$" Suppose there exists B such that $\mu = \mu(x^B, T)$ and $x^B(\theta)$ is a maximizer of (1) for any $\theta \in \Theta$. (I) and (II) follow immediately.

Proof to Proposition 2

The proof uses the following lemma.

Lemma A 2 $\forall B = (t, r) \in \Theta \times [0,1] \forall \theta \leq t \forall T \in \mathcal{T} : V^Y_\theta(T, \mu(x^B, T)) \geq \theta v_A(1) + (1-\theta) v_A(0)$.

Proof. Using concavity of v_A, we obtain

$$\begin{aligned} V^Y_\theta(T, \mu(x^B, T)) &= \sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega v_A(\mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g)) \\ &\geq \sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega [\mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g) v_A(1) + (1 - \mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g)) v_A(0)] \\ &= \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega \mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g) \right) v_A(1) \\ &\quad + \left(1 - \sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega \mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g) \right) v_A(0). \end{aligned}$$

Since v_A is increasing, a sufficient condition for the statement is that the first bracketed expression is larger than θ when $\theta \leq t$. We have

$$\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega \mu_\sigma(\mu_Y(x^B), p^\sigma_b, p^\sigma_g) - \theta$$

$$
\begin{aligned}
&= \sum_{\omega,\sigma} q_\theta^\omega p_\omega^\sigma \frac{\mu_Y(x^B)p_g^\sigma}{\mu_Y(x^B)p_g^\sigma + (1-\mu_Y(x^B))p_b^\sigma} - \sum_\sigma \theta p_g^\sigma \\
&= \sum_\sigma \frac{\theta p_g^\sigma + (1-\theta)p_b^\sigma}{\mu_Y(x^B)p_g^\sigma + (1-\mu_Y(x^B))p_b^\sigma}\mu_Y(x^B)p_g^\sigma - \sum_\sigma \frac{\mu_Y(x^B)p_g^\sigma + (1-\mu_Y(x^B))p_b^\sigma}{\mu_Y(x^B)p_g^\sigma + (1-\mu_Y(x^B))p_b^\sigma}\theta p_g^\sigma \\
&= \sum_\sigma \frac{\mu_Y(x^B)-\theta}{\mu_Y(x^B)p_g^\sigma + (1-\mu_Y(x^B))p_b^\sigma} p_b^\sigma p_g^\sigma .
\end{aligned}
$$

Since $\mu_Y(x^B) \geq t$ for any B, the expression is positive for any $\theta \leq t$. q.e.d.

(a) "$\Rightarrow$" Suppose $B \in \mathcal{B}^b$. There exists $T \in \mathcal{T}$ such that $(x^B, \mu(x^B,T)) \in \mathcal{E}(T)$ and $V_t^Y(T, \mu(x^B,T)) = v_A(\mu_N(x^B))$. Using Lemma A 2 with $\theta = t$, $tv_A(1) + (1-t)v_A(0) \leq v_A(\mu_N(x^B))$. Hence, $B \in \{B = (t,r) \in \Theta \times [0,1] | tv_A(1) + (1-t)v_A(0) \leq v_A(\mu_N(x^B))\}$. "$\Leftarrow$" Suppose $B \in \{B = (t,r) \in \Theta \times [0,1] | tv_A(1) + (1-t)v_A(0) \leq v_A(\mu_N(x^B))\}$. For $w \in (0,1]$ define $T(w) := (2, (1,0), (1-w, w))$. $V_t^Y(T(w), \mu(x^B, T(w)))$ is continuous in w. We have $\lim_{w\to 0} V_t^Y(T(w), \mu(x^B, T(w))) - v_A(\mu_N(x^B)) = v_A(\mu_Y(x^B)) - v_A(\mu_N(x^B))$. This expression is strictly positive for any B as v_A is strictly increasing. Moreover, $V_t^Y(T(1), \mu(x^B, T(1))) - v_A(\mu_N(x^B)) = tv_A(1) + (1-t)v_A(0) - v_A(\mu_N(x^B))$. This is non–positive by assumption. By the Intermediate Value Theorem, there exists $w' \in (0,1]$ such that the participation constraint is binding. By Lemma A 1, $(x^B, \mu(x^B, T(w'))) \in \mathcal{E}(T(w'))$.

(b) "$\Rightarrow$" Suppose $B \in \mathcal{B}$. If $B \in \mathcal{B}^b$, we are done. Suppose $B \notin \mathcal{B}^b$. By (a), $tv_A(1) + (1-t)v_A(0) > v_A(\mu_N(x^B))$. That is, when the agent's signal is $\theta = t$, he has a strict incentive to participate in an accurate test. By Lemma A 1, he also has a strict incentive to participate when $\theta > t$. It remains to argue that he has no incentive to participate when $\theta < t$. Because $B \in \mathcal{B}$, there exists T such that for all $\theta < t$, $V_\theta^Y(T, \mu(x^B,T)) \leq v_A(\mu_N(x^B))$, and for all $\theta \geq t$, $V_\theta^Y(T, \mu(x^B,T)) \geq v_A(\mu_N(x^B))$. By Lemma A 2, $V_\theta^Y(T, \mu(x^B,T)) \geq \theta v_A(1) + (1-\theta)v_A(0)$ for all $\theta < t$. It follows that the agent cannot have an incentive to participate in an accurate test when $\theta < t$. Hence, $B \in \mathcal{B}^a$. "$\Leftarrow$" Suppose $B \in \mathcal{B}^a \cup \mathcal{B}^b$. By the definition of $\mathcal{B}^a$ and $\mathcal{B}^b$, $B \in \mathcal{B}$.

(c) "$\Rightarrow$" Suppose $B \in \mathcal{B}$. If $B \in \mathcal{B}^a$, B can also be induced by a binary test as quality is binary. Suppose $B \in \mathcal{B}\backslash\mathcal{B}^a$. By (b), $B \in \mathcal{B}^b$. In the proof of (a) we constructed a binary test that induces B. "$\Leftarrow$" Suppose $B \in \mathcal{B}^2$. $B \in \mathcal{B}$ follows from the definition of $\mathcal{B}^2$.

Proof to Proposition 3

If $B \in \mathcal{B}^a$, an accurate test is optimal by Lemma 1. In particular, a binary accurate test is optimal. The statements in (a) and (b) are trivially true as an accurate test is neither subject to false positives nor to false negatives. Consider $B \in \mathcal{B}\backslash\mathcal{B}^a$. (a) Suppose the principal strives for maximizing a non–positive function of the agent's ad interim expected utility. The result follows directly from Lemma 2. (b) Suppose the principal strives for maximizing the variance. Her ex ante expected utility depends only on the first moment and the second moment of the induced quality belief distribution. Since v_P is convex, the second moment affects her expected utility positively. Since the first moment is fixed because of Bayes' Law, the principal strives for maximizing the second moment. By Lemma 3, a binary test which is not subject to false positives is optimal when $c_1 \geq -1$.

Proof to Lemma 1

In the proofs to Lemma 1, 2 and 3, we make use of a property which eases the exposition: A test T is minimal if no two signals are associated to the same quality belief (i.e., $\mu_{\sigma'}(x^B, p_b^{\sigma'}, p_g^{\sigma'}) \neq \mu_{\sigma''}(x^B, p_b^{\sigma''}, p_g^{\sigma''})$ for any $\sigma' \neq \sigma''$). A test T is ordered if test results are ordered according to inference (i.e., $\mu_1(x^B, p_b^1, p_g^1) \leq \ldots \leq \mu_Z(x^B, p_b^Z, p_g^Z)$). Since neither merging test results associated with the same quality belief nor permuting names of test results affect the payoffs of the principal and the agent, there exists an optimal minimal and ordered test. Restricting attention to minimal and ordered tests is without loss of generality.

We start the proof with a lemma:

Lemma A 3 *Compare two quality belief lotteries with the same expected quality belief. (a) If the support of the first lottery is $M_1 = \{m_1, m_2\}$ with $m_1 < m_2$ and the support of the second lottery is $M_2 \subset [m_1, m_2]$, the principal prefers the first quality belief lottery. (b) Suppose that both quality belief lotteries have Z possible outcomes and suppose that $Z-2$ outcomes are associated with the same quality beliefs and probabilities. Denote the two remaining quality beliefs of the first lottery by m_1' and m_2' with $m_1' < m_2'$ and of the second lottery by m_1'' and m_2'' with $m_1'' < m_2''$. If $m_1' \leq m_1''$ and $m_2' \geq m_2''$ with one inequality being strict, the principal strictly prefers the first lottery.*

Proof. In both cases, the first quality belief lottery can be obtained by a mean–preserving spread from the second quality belief lottery. The two results follow directly from the convexity assumption on v_P. q.e.d.

(a) When comparing tests which induce the same participation behavior, only the quality belief lotteries that are induced conditional on participation matter for the comparison. If the test is accurate, the support of the induced quality belief distribution (conditional on participation) is $\{0,1\}$. For any other test inducing the same participation behavior, the induced quality belief lottery (conditional on participation) has a support which is a subset of $[0,1]$. By Lemma A 3 (a), the principal prefers the accurate test.

(b) Suppose $B = (t,r) \in \mathcal{B} \backslash \mathcal{B}^a$ and suppose $T = (Z, p_b, p_g)$ is the optimal minimal and ordered test inducing B. Moreover, suppose $V_t^Y(T, \mu(x^B, T)) > v_A(\mu_N(x^B))$. We first construct a test $\widehat{T}(\epsilon)$ that differs from T only by the redistribution of probability ϵ from one test result to another such that $V_\theta^Y(\widehat{T}(\epsilon), \mu(x^B, \widehat{T}(\epsilon)))$ locally decreases in ϵ for any $\theta \leq t$. We distinguish two cases.

Case (i): $\exists \sigma' : p_b^{\sigma'} \in (0,1)$. Then, $\exists \sigma'' \neq \sigma' : p_b^{\sigma''} \in (0,1)$. Suppose without loss of generality $\mu_{\sigma'}(\mu_Y(x^B), p_b^{\sigma'}, p_g^{\sigma'}) < \mu_{\sigma''}(\mu_Y(x^B), p_b^{\sigma''}, p_g^{\sigma''})$. Construct a test $\widehat{T}(\epsilon)$ from test T by increasing $p_b^{\sigma'}$ by ϵ and by decreasing $p_b^{\sigma''}$ by ϵ. We have

$$\begin{aligned} V_\theta^Y(\widehat{T}(\epsilon), \mu(x^B, \widehat{T}(\epsilon))) \quad = \quad & ((1-\theta)(p_b^{\sigma'} + \epsilon) + \theta p_g^{\sigma'}) v_A(\mu_{\sigma'}(\mu_Y(x^B), p_b^{\sigma'} + \epsilon, p_g^{\sigma'})) \\ & + ((1-\theta)(p_b^{\sigma''} - \epsilon) + \theta p_g^{\sigma''}) v_A(\mu_{\sigma''}(\mu_Y(x^B), p_b^{\sigma''} - \epsilon, p_g^{\sigma''})) \\ & + \sum_{\sigma \neq \sigma', \sigma''} ((1-\theta) p_b^\sigma + \theta p_g^\sigma) v_A(\mu_\sigma(\mu_Y(x^B), p_b^\sigma, p_g^\sigma)) \end{aligned}$$

and

$$\left.\frac{\mathrm{d}V_\theta^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon)))}{\mathrm{d}\epsilon}\right|_{\epsilon=0}$$

$$\begin{aligned} = \quad & (1-\theta)v_A(\mu_{\sigma'}(\cdot)) - ((1-\theta)p_b^{\sigma'}+\theta p_g^{\sigma'})\frac{(1-\mu_Y(x^B))\mu_{\sigma'}(\cdot)}{(1-\mu_Y(x^B))p_b^{\sigma'}+\mu_Y(x^B)p_g^{\sigma'}}v_A'(\mu_{\sigma'}(\cdot)) \\ & -(1-\theta)v_A(\mu_{\sigma''}(\cdot)) + ((1-\theta)p_b^{\sigma''}+\theta p_g^{\sigma''})\frac{(1-\mu_Y(x^B))\mu_{\sigma''}(\cdot)}{(1-\mu_Y(x^B))p_b^{\sigma''}+\mu_Y(x^B)p_g^{\sigma''}}v_A'(\mu_{\sigma''}(\cdot)). \end{aligned}$$

Using $\frac{(1-\theta)p_b^\sigma+\theta p_g^\sigma}{(1-\mu_Y(x^B))p_b^\sigma+\mu_Y(x^B)p_g^\sigma} = \frac{1}{1-\mu_Y(x^B)}\left((1-\theta)-(1-\frac{\theta}{\mu_Y(x^B)})\mu_\sigma(\mu_Y(x^B),p_b^\sigma,p_g^\sigma)\right)$, we get

$$\begin{aligned} \ldots \quad = \quad & (1-\theta)v_A(\mu_{\sigma'}(\cdot)) - \left((1-\theta)-\left(1-\frac{\theta}{\mu_Y(x^B)}\right)\mu_{\sigma'}(\cdot)\right)\mu_{\sigma'}(\cdot)v_A'(\mu_{\sigma'}(\cdot)) \\ & -(1-\theta)v_A(\mu_{\sigma''}(\cdot)) + \left((1-\theta)-\left(1-\frac{\theta}{\mu_Y(x^B)}\right)\mu_{\sigma''}(\cdot)\right)\mu_{\sigma''}(\cdot)v_A'(\mu_{\sigma''}(\cdot)). \end{aligned}$$

This expression is strictly negative if the auxiliary function

$$\xi(m) \quad := \quad (1-\theta)\left(v_A(m)-mv_A'(m)\right) + \left(1-\frac{\theta}{\mu_Y(x^B)}\right)m^2v_A'(m)$$

is strictly increasing in m on $(0,1)$. We have

$$\begin{aligned} \xi'(m) \quad & = \quad -(1-\theta)mv_A''(m) + \left(1-\frac{\theta}{\mu_Y(x^B)}\right)(2mv_A'(m)+m^2v_A''(m)) \\ & = \quad \left(1-\frac{\theta}{\mu_Y(x^B)}\right)2mv_A'(m) - \left((1-\theta)-\left(1-\frac{\theta}{\mu_Y(x^B)}\right)m\right)mv_A''(m). \end{aligned}$$

The first expression is strictly positive for $m\in(0,1)$ as $\theta<\mu_Y(x^B)$ for $\theta\le t$ and as v_A is strictly increasing on $(0,1)$. The second expression is strictly positive for $m\in(0,1)$ as $1-\theta > 1-\frac{\theta}{\mu_Y(x^B)} > \left(1-\frac{\theta}{\mu_Y(x^B)}\right)m$ for $\theta\le t$ and $m\in(0,1)$, and as v_A is strictly concave on $(0,1)$. This establishes that $V_\theta^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon)))$ is strictly decreasing in ϵ at $\epsilon=0$ for any $\theta\le t$.

Case (ii): $\forall\sigma: p_b^\sigma\in\{0,1\}$. Since the test is minimal and ordered, $Z=2$, $p_b^1=1$, $p_b^2=0$. Since $T\notin\mathcal{T}^a$, $p_g^1,p_g^2\in(0,1)$. Construct a test $\widehat{T}(\epsilon)$ from test T by increasing p_g^2 by ϵ and by decreasing p_g^1 by ϵ. We have

$$V_\theta^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon))) = ((1-\theta)+\theta(p_g^1-\epsilon))v_A(\mu_1(\mu_Y(x^B),1,p_g^1-\epsilon)) + \theta(p_g^2+\epsilon)v_A(1)$$

and

$$\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}V_\theta^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon)))\right|_{\epsilon=0}$$

$$= \quad \theta(v_A(1)-v_A(\mu_1(\cdot))) - (\theta p_g^1+(1-\theta))\frac{\mu_Y(x^B)}{\mu_Y(x^B)p_g^1+(1-\mu_Y(x^B))}(1-\mu_1(\cdot))v_A'(\mu_1(\cdot)).$$

This expression is strictly negative as $v_A(1)-v_A(\mu_1(\cdot)) < (1-\mu_1(\cdot))v_A'(\mu_1(\cdot))$ by strict concavity of v_A and as for any $\theta\le t$, $\theta\le\mu_Y(x^B)$ and $\frac{\theta p_g^1+(1-\theta)}{\mu_Y(x^B)p_g^1+(1-\mu_Y(x^B))}\ge 1$.

Consequently, for a strictly positive but sufficiently small ϵ, $V_t^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon))) \ge v_A(\mu_N(x^B))$. I.e., the agent still has an incentive to take the adjusted test when $\theta=t$. Moreover, by Lemma A 1, the agent still has an incentive to take the adjusted test when $\theta>t$. Finally, by $\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon}V_\theta^Y(\widehat{T}(\epsilon),\mu(x^B,\widehat{T}(\epsilon)))\right|_{\epsilon=0} <$

0 for any $\theta < t$, he still has no incentive to take the adjusted test when $\theta < t$. Hence, also test $\widehat{T}(\epsilon)$ induces participation behavior x^B. The induced quality belief lotteries differ only with respect to two quality beliefs (and the respective probabilities), the smaller one is in both cases weakly smaller for test $\widehat{T}(\epsilon)$ and the larger one is in both cases weakly larger for test $\widehat{T}(\epsilon)$. A strict difference obtains for at least one of the two cases. By Lemma A 3 (b), the principal strictly prefers the quality belief lottery induced by the adjusted test. Hence, a test which induces participation behavior B with a non–binding participation constraint cannot be optimal.

Proof to Lemma 2

Define $H : [0,1]^2 \to \mathbf{R}$ by $(m', m'') \mapsto [v_A(m'') - v_A(m')]^2 - (m'' - m')^2 v_A'(m') v_A'(m'')$. In Part I of the proof, we derive a result which is more general than Lemma 2 and which is stated in Lemma A 4. The result depends on the sign of $H(m', m'')$, but does not require v_A to exhibit HARA. In Part II of the proof we relate the cases in Lemma A 4 to HARA utility functions.

Lemma A 4 *Fix $B = (t, r) \in \mathcal{B}\backslash\mathcal{B}^a$ and consider tests from $\mathcal{T}_B^b$. (a) If $H(m', m'') > 0$ for all $m', m'' \in [0,1]$ with $m' \neq m''$, the agent's ad interim expected utility for $\theta > t$ is minimized by a binary test which is not subject to false positives and maximized by a binary test which is not subject to false negatives. (b) If $H(m', m'') < 0$ for all $m', m'' \in [0,1]$ with $m' \neq m''$, the agent's ad interim expected utility for $\theta > t$ is minimized by a binary test which is not subject to false negatives and maximized by a binary test which is not subject to false positives. (c) If $H(m', m'') = 0$ for all $m', m'' \in [0,1]$, the agent's ad interim expected utility for $\theta > t$ is not affected by the test design.*

Part I. Fix $B = (t, r) \in \mathcal{B}\backslash\mathcal{B}^a$. Suppose $T = (Z, p_b, p_g) \in \mathcal{T}$ is a minimal and ordered test such that $(x^B, \mu) \in \mathcal{E}(T)$ with $V_t^Y(T, \mu) = v_A(\mu_N)$. Equilibrium requires $\mu = \mu(x^B, T)$. Since $B \notin \mathcal{B}^a$, either two of the induced quality beliefs are strictly smaller than one or two of the induced quality beliefs are strictly larger than zero.

Case (i): Suppose $\exists \sigma', \sigma'' \in \Sigma^Z : \mu_{\sigma'}(\mu_Y(x^B), p_b^{\sigma'}, p_g^{\sigma'}) < \mu_{\sigma''}(\mu_Y(x^B), p_b^{\sigma''}, p_g^{\sigma''}) < 1$. Define $\widehat{T}(\epsilon_b, \epsilon_g)$ from T by choosing $\widehat{p}_b^{\sigma''} = p_b^{\sigma''} - \epsilon_b$, $\widehat{p}_b^{\sigma'} = p_b^{\sigma'} + \epsilon_b$, $\widehat{p}_g^{\sigma''} = p_g^{\sigma''} - \epsilon_g$, $\widehat{p}_g^{\sigma'} = p_g^{\sigma'} + \epsilon_g$ and $\forall \sigma \in \Sigma^Z\backslash\{\sigma', \sigma''\} \forall \omega : \widehat{p}_\omega^\sigma = p_\omega^\sigma$. Note that $\mu_{\sigma''} < 1$ implies $p_b^{\sigma''} > 0$ and thus $p_b^{\sigma'} < 1$. Moreover, $\mu_{\sigma'} < \mu_{\sigma''}$ requires $p_g^{\sigma''} > 0$ which implies $p_g^{\sigma'} < 1$. It follows $p_b^{\sigma''} - \epsilon_b, p_b^{\sigma'} + \epsilon_b, p_g^{\sigma''} - \epsilon_g, p_g^{\sigma'} + \epsilon_g \in (0,1)$ for ϵ_g and ϵ_b strictly positive and sufficiently small. I.e., $\widehat{T}(\epsilon_b, \epsilon_g)$ is indeed a test.

We introduce the following shortcut notation: For $w \in [0,1]$, define $\widehat{p}_w^\sigma := w\widehat{p}_g^\sigma + (1-w)\widehat{p}_b^\sigma$. Note that $\widehat{p}_0^\sigma = \widehat{p}_b^\sigma$ and $\widehat{p}_1^\sigma = \widehat{p}_g^\sigma$ such that we can also write $\widehat{p}_w^\sigma := w\widehat{p}_1^\sigma + (1-w)\widehat{p}_0^\sigma$. For $w \in (0,1)$ define $\widehat{V}_w^Y(\epsilon_b, \epsilon_g) := \widehat{p}_w^{\sigma'} v_A(\mu_{\sigma'}) + \widehat{p}_w^{\sigma''} v_A(\mu_{\sigma''})$ with $\mu_\sigma = \mu_Y \frac{\widehat{p}_1^\sigma}{\widehat{p}_{\mu_Y}^\sigma}$ and with μ_Y short for $\mu_Y(x^B)$. We obtain the following properties which we will use in the course of the proof.

Lemma A 5 *(a)* $\mu_{\sigma''} - \mu_{\sigma'} = \mu_Y(1-\mu_Y)\frac{\widehat{p}_0^{\sigma'}\widehat{p}_1^{\sigma''} - \widehat{p}_0^{\sigma''}\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}\widehat{p}_{\mu_Y}^{\sigma''}}$. *(b)* $\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(0,0) < 0$. *(c)* $\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(0,0) > 0$. *(d)* $\frac{\partial \widehat{V}_s^Y}{\partial \epsilon_b}(\epsilon_b, \epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(\epsilon_b, \epsilon_g) - \frac{\partial \widehat{V}_s^Y}{\partial \epsilon_g}(\epsilon_b, \epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(\epsilon_b, \epsilon_g) = -(s-t)H(\mu_{\sigma'}, \mu_{\sigma''})$.

Proof. (a) We have $\mu_{\sigma''} - \mu_{\sigma'} = \frac{\mu_Y \widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}} - \frac{\mu_Y \widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} = \mu_Y(1-\mu_Y)\frac{\widehat{p}_0^{\sigma'}\widehat{p}_1^{\sigma''} - \widehat{p}_0^{\sigma''}\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}\widehat{p}_{\mu_Y}^{\sigma''}}$.

(b) We have $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} = -\frac{1-\mu_Y}{\widehat{p}^{\sigma'}_{\mu_Y}}\mu_{\sigma'}$ and $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} = \frac{1-\mu_Y}{\widehat{p}^{\sigma''}_{\mu_Y}}\mu_{\sigma''}$. We get

$$\begin{aligned}\frac{\partial \widehat{V}^Y_t}{\partial \epsilon_b}(0,0) &= -(1-t)[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}^{\sigma'}_t \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma'}) + \widehat{p}^{\sigma''}_t \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma''}) \\ &= -(1-t)[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] \\ &\quad -(1-\mu_Y)\frac{\widehat{p}^{\sigma'}_t}{\widehat{p}^{\sigma'}_{\mu_Y}}\mu_{\sigma'} v'_A(\mu_{\sigma'}) + (1-\mu_Y)\frac{\widehat{p}^{\sigma''}_t}{\widehat{p}^{\sigma''}_{\mu_Y}}\mu_{\sigma''} v'_A(\mu_{\sigma''}).\end{aligned}$$

By strict concavity of v_A, $v'_A(\mu_{\sigma'}) > v'_A(\mu_{\sigma''})$ and $[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] > [\mu_{\sigma''} - \mu_{\sigma'}]v'_A(\mu_{\sigma''})$. Hence,

$$\frac{\partial \widehat{V}^Y_t}{\partial \epsilon_b}(0,0) < \left\{-(1-t)[\mu_{\sigma''} - \mu_{\sigma'}] - (1-\mu_Y)\frac{\widehat{p}^{\sigma'}_t}{\widehat{p}^{\sigma'}_{\mu_Y}}\mu_{\sigma'} + (1-\mu_Y)\frac{\widehat{p}^{\sigma''}_t}{\widehat{p}^{\sigma''}_{\mu_Y}}\mu_{\sigma''}\right\} v'_A(\mu_{\sigma''}).$$

Note that $\frac{\widehat{p}^{\sigma'}_t}{\widehat{p}^{\sigma'}_{\mu_Y}} \geq \frac{\widehat{p}^{\sigma''}_t}{\widehat{p}^{\sigma''}_{\mu_Y}} \Leftrightarrow (t\widehat{p}^{\sigma'}_1 + (1-t)\widehat{p}^{\sigma'}_0)(\mu_Y \widehat{p}^{\sigma''}_1 + (1-\mu_Y)\widehat{p}^{\sigma''}_0) \geq (t\widehat{p}^{\sigma''}_1 + (1-t)\widehat{p}^{\sigma''}_0)(\mu_Y \widehat{p}^{\sigma'}_1 + (1-\mu_Y)\widehat{p}^{\sigma'}_0) \Leftrightarrow (\mu_Y - t)(\widehat{p}^{\sigma''}_1 \widehat{p}^{\sigma'}_0 - \widehat{p}^{\sigma'}_1 \widehat{p}^{\sigma''}_0) \geq 0$. The inequality holds since $\mu_Y \geq t$ and since $\widehat{p}^{\sigma''}_1 \widehat{p}^{\sigma'}_0 - \widehat{p}^{\sigma'}_1 \widehat{p}^{\sigma''}_0 > 0$ for $\epsilon_b = \epsilon_g = 0$ by (a) and our initial assumption in case (i). Using this,

$$\begin{aligned}\ldots &\leq \left\{-(1-t) + (1-\mu_Y)\frac{\widehat{p}^{\sigma''}_t}{\widehat{p}^{\sigma''}_{\mu_Y}}\right\}[\mu_{\sigma''} - \mu_{\sigma'}]v'_A(\mu_{\sigma''}) \\ &= \left\{-(1-t)\widehat{p}^{\sigma''}_{\mu_Y} + (1-\mu_Y)\widehat{p}^{\sigma''}_t\right\}\frac{1}{\widehat{p}^{\sigma''}_{\mu_Y}}[\mu_{\sigma''} - \mu_{\sigma'}]v'_A(\mu_{\sigma''}) \\ &= (t-\mu_Y)\frac{\widehat{p}^{\sigma''}_1}{\widehat{p}^{\sigma''}_{\mu_Y}}[\mu_{\sigma''} - \mu_{\sigma'}]v'_A(\mu_{\sigma''}) \\ &\leq 0.\end{aligned}$$

(c) We have $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} = \frac{\mu_Y}{\widehat{p}^{\sigma'}_{\mu_Y}}(1-\mu_{\sigma'})$ and $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} = -\frac{\mu_Y}{\widehat{p}^{\sigma''}_{\mu_Y}}(1-\mu_{\sigma''})$. We get

$$\begin{aligned}\frac{\partial \widehat{V}^Y_t}{\partial \epsilon_g}(0,0) &= -t[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}^{\sigma'}_t \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma'}) + \widehat{p}^{\sigma''}_t \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma''}) \\ &= -t[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] \\ &\quad +\mu_Y \frac{\widehat{p}^{\sigma'}_t}{\widehat{p}^{\sigma'}_{\mu_Y}}(1-\mu_{\sigma'})v'_A(\mu_{\sigma'}) - \mu_Y \frac{\widehat{p}^{\sigma''}_t}{\widehat{p}^{\sigma''}_{\mu_Y}}(1-\mu_{\sigma''})v'_A(\mu_{\sigma''}).\end{aligned}$$

Using $\mu_Y \frac{\widehat{p}^{\sigma}_t}{\widehat{p}^{\sigma}_{\mu_Y}} = t + \frac{\widehat{p}^{\sigma}_0}{\widehat{p}^{\sigma}_{\mu_Y}}(\mu_Y - t)$, we get

$$\begin{aligned}\frac{\partial \widehat{V}^Y_t}{\partial \epsilon_g}(0,0) &= -\left[t v_A(\mu_{\sigma''}) + \left(t + \frac{\widehat{p}^{\sigma''}_0}{\widehat{p}^{\sigma''}_{\mu_Y}}(\mu_Y - t)\right)(1-\mu_{\sigma''})v'_A(\mu_{\sigma''})\right] \\ &\quad + \left[t v_A(\mu_{\sigma'}) + \left(t + \frac{\widehat{p}^{\sigma'}_0}{\widehat{p}^{\sigma'}_{\mu_Y}}(\mu_Y - t)\right)(1-\mu_{\sigma'})v'_A(\mu_{\sigma'})\right].\end{aligned}$$

To show $\frac{\partial \widehat{V}^Y_t}{\partial \epsilon_g}(0,0) > 0$, it suffices to show that $\xi(\mu_\sigma) := t v_A(\mu_\sigma) + (t + \frac{\widehat{p}^{\sigma}_0}{\widehat{p}^{\sigma}_{\mu_Y}}(\mu_Y - t))(1-\mu_\sigma)v'_A(\mu_\sigma)$ is strictly decreasing in μ_σ on $(0,1)$. For any $\mu_\sigma \in (0,1)$, we have

$$\begin{aligned}\xi'(\mu_\sigma) &= t v'_A(\mu_\sigma) - \left(t + \frac{\widehat{p}^{\sigma'}_0}{\widehat{p}^{\sigma'}_{\mu_Y}}(\mu_Y - t)\right)v'_A(\mu_\sigma) + \left(t + \frac{\widehat{p}^{\sigma'}_0}{\widehat{p}^{\sigma'}_{\mu_Y}}(\mu_Y - t)\right)(1-\mu_\sigma)v''_A(\mu_\sigma) \\ &= -\frac{\widehat{p}^{\sigma'}_0}{\widehat{p}^{\sigma'}_{\mu_Y}}(\mu_Y - t)v'_A(\mu_\sigma) + \left(t + \frac{\widehat{p}^{\sigma'}_0}{\widehat{p}^{\sigma'}_{\mu_Y}}(\mu_Y - t)\right)(1-\mu_\sigma)v''_A(\mu_\sigma) \\ &< 0\end{aligned}$$

since $\mu_Y \geq t$, $v'_A > 0$ and $v''_A < 0$.

(d) We have

$$\begin{aligned}
& \frac{\partial \widehat{V}_s^Y}{\partial \epsilon_b}(\epsilon_b, \epsilon_g) \frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(\epsilon_b, \epsilon_g) - \frac{\partial \widehat{V}_s^Y}{\partial \epsilon_g}(\epsilon_b, \epsilon_g) \frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(\epsilon_b, \epsilon_g) \\
= \; & \left[-(1-s)[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}_s^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma'}) + \widehat{p}_s^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma''}) \right] \\
& \cdot \left[-t[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}_t^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma'}) + \widehat{p}_t^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma''}) \right] \\
& - \left[-s[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}_s^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma'}) + \widehat{p}_s^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} v'_A(\mu_{\sigma''}) \right] \\
& \cdot \left[-(1-t)[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] + \widehat{p}_t^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma'}) + \widehat{p}_t^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} v'_A(\mu_{\sigma''}) \right] \\
= \; & -[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})]^2 (s-t) \\
& + [v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] v'_A(\mu_{\sigma'}) \\
& \quad \cdot \left\{ [s\widehat{p}_t^{\sigma'} - t\widehat{p}_s^{\sigma'}] \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} + [(1-t)\widehat{p}_s^{\sigma'} - (1-s)\widehat{p}_t^{\sigma'}] \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} \right\} \\
& + [v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] v'_A(\mu_{\sigma''}) \\
& \quad \cdot \left\{ [s\widehat{p}_t^{\sigma''} - t\widehat{p}_s^{\sigma''}] \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} + [(1-t)\widehat{p}_s^{\sigma''} - (1-s)\widehat{p}_t^{\sigma''}] \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} \right\} \\
& + v'_A(\mu_{\sigma'}) v'_A(\mu_{\sigma''}) \left\{ \widehat{p}_s^{\sigma'} \widehat{p}_t^{\sigma''} - \widehat{p}_t^{\sigma'} \widehat{p}_s^{\sigma''} \right\} \left\{ \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} - \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} \right\}.
\end{aligned}$$

Consider first the first two expression in curly brackets:

$$\begin{aligned}
& [s\widehat{p}_t^{\sigma} - t\widehat{p}_s^{\sigma}] \frac{\mathrm{d}\mu_{\sigma}}{\mathrm{d}\epsilon_b} + [(1-t)\widehat{p}_s^{\sigma} - (1-s)\widehat{p}_t^{\sigma}] \frac{\mathrm{d}\mu_{\sigma}}{\mathrm{d}\epsilon_g} \\
= \; & (s-t) \left\{ \widehat{p}_0^{\sigma} \frac{\mathrm{d}\mu_{\sigma}}{\mathrm{d}\epsilon_b} + \widehat{p}_1^{\sigma} \frac{\mathrm{d}\mu_{\sigma}}{\mathrm{d}\epsilon_g} \right\} \\
= \; & \begin{cases} (s-t)[-\widehat{p}_0^{\sigma'} \mu_Y (1-\mu_Y) \frac{\widehat{p}_1^{\sigma'}}{(\widehat{p}_{\mu_Y}^{\sigma'})^2} + \widehat{p}_1^{\sigma'} \mu_Y (1-\mu_Y) \frac{\widehat{p}_0^{\sigma'}}{(\widehat{p}_{\mu_Y}^{\sigma'})^2}] & \text{if } \sigma = \sigma' \\ (s-t)[\widehat{p}_0^{\sigma''} \mu_Y (1-\mu_Y) \frac{\widehat{p}_1^{\sigma''}}{(\widehat{p}_{\mu_Y}^{\sigma''})^2} - \widehat{p}_1^{\sigma''} \mu_Y (1-\mu_Y) \frac{\widehat{p}_0^{\sigma''}}{(\widehat{p}_{\mu_Y}^{\sigma''})^2}] & \text{if } \sigma = \sigma'' \end{cases} \\
= \; & 0
\end{aligned}$$

Consider now the third expression in curly brackets:

$$\begin{aligned}
\widehat{p}_s^{\sigma'} \widehat{p}_t^{\sigma''} - \widehat{p}_t^{\sigma'} \widehat{p}_s^{\sigma''} &= (s-t)[\widehat{p}_0^{\sigma''} \widehat{p}_1^{\sigma'} - \widehat{p}_0^{\sigma'} \widehat{p}_1^{\sigma''}] \\
&\overset{(a)}{=} -(s-t) \frac{\widehat{p}_{\mu_Y}^{\sigma'} \widehat{p}_{\mu_Y}^{\sigma''}}{\mu_Y(1-\mu_Y)} (\mu_{\sigma''} - \mu_{\sigma'})
\end{aligned}$$

Finally, consider the last expression in curly brackets:

$$\begin{aligned}
& \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} - \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} \\
= \; & \left(-(1-\mu_Y) \frac{\mu_{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} \right) \left(-\mu_Y \frac{1-\mu_{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}} \right) - \left(\mu_Y \frac{1-\mu_{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} \right) \left((1-\mu_Y) \frac{\mu_{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}} \right) \\
= \; & -\frac{\mu_Y(1-\mu_Y)}{\widehat{p}_{\mu_Y}^{\sigma'} \widehat{p}_{\mu_Y}^{\sigma''}} [\mu_{\sigma''} - \mu_{\sigma'}]
\end{aligned}$$

By replacing the curly brackets, we obtain the result:

$$\begin{aligned}
& \frac{\partial \widehat{V}_s^Y}{\partial \epsilon_b}(\epsilon_b,\epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(\epsilon_b,\epsilon_g) - \frac{\partial \widehat{V}_s^Y}{\partial \epsilon_g}(\epsilon_b,\epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(\epsilon_b,\epsilon_g) \\
= \; & -[v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})]^2(s-t) + v_A'(\mu_{\sigma'})v_A'(\mu_{\sigma''})(s-t)[\mu_{\sigma''}-\mu_{\sigma'}]^2 \\
= \; & -(s-t)H(\mu_{\sigma'},\mu_{\sigma''}) \qquad \text{q.e.d.}
\end{aligned}$$

We now consider tests $\widehat{T}(\epsilon_b, \epsilon_g(\epsilon_b))$ where $\epsilon_g(\epsilon_b)$ is chosen such that the participation behavior is kept constant and show how the agent's ad interim expected utility from participation depends on ϵ_b. Note that there exists C_θ which neither depends on ϵ_b nor on ϵ_g such that $V_\theta^Y(\widehat{T}(\epsilon_b,\epsilon_g), \mu(x^B, \widehat{T}(\epsilon_b,\epsilon_g))) = \widehat{V}_\theta^Y(\epsilon_b,\epsilon_g) + C_\theta$. By the Implicit Function Theorem, $\epsilon_g'(\epsilon_b) = -\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(\epsilon_b,\epsilon_g)/\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(\epsilon_b,\epsilon_g)$. By Lemma A 5 (b) and (c), $\epsilon_g'(0) > 0$ such that test $\widehat{T}(\epsilon_b, \epsilon_g(\epsilon_b))$ is indeed a test for sufficiently small but strictly positive ϵ_b. The sign of

$$\frac{\partial \widehat{V}_\theta^Y}{\partial \epsilon_b}(0,0) + \epsilon_g'(0)\frac{\partial \widehat{V}_\theta^Y}{\partial \epsilon_g}(0,0) = \frac{\frac{\partial \widehat{V}_\theta^Y}{\partial \epsilon_b}(0,0)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(0,0) - \frac{\partial \widehat{V}_\theta^Y}{\partial \epsilon_g}(0,0)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(0,0)}{\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(0,0)}$$

determines the direction in which a small increase in ϵ_b changes the agent's ad interim expected utility. By Lemma A 5 (c), the denominator is positive such that the sign of the expression is the sign of the numerator. By Lemma A 5 (d), the numerator is $-(\theta - t)H(\mu_{\sigma'}, \mu_{\sigma''})$.

The numerator is strictly negative for any $\theta > t$ if $H(m', m'') > 0$ for all $m', m'' \in (0,1)$. Under this condition, a test inducing at least two quality beliefs which are strictly smaller than one cannot minimize the ad interim expected utility of the agent when $\theta > t$. If follows that the test minimizing the agent's ad interim expected utility is a binary test which is not subject to false positives. This is the first part of Lemma A 4 (a).

The numerator is strictly positive for any $\theta > t$ if $H(m', m'') < 0$ for all $m', m'' \in (0,1)$. Under this condition, a test inducing at least two quality beliefs which are strictly smaller than one cannot maximize the ad interim expected utility of the agent when $\theta > t$. If follows that the test maximizing the agent's ad interim expected utility is a binary test which is not subject to false positives. This is the second part of Lemma A 4 (b).

Case (ii): Suppose $\exists \sigma', \sigma'' \in \Sigma^Z : 0 < \mu_{\sigma'} < \mu_{\sigma''}$. Case (ii) is like case (i) but with $\epsilon_b, \epsilon_g < 0$. We obtain the first part of Lemma A 4 (b) and the second part of Lemma A 4 (a) as result.

Part (c) of Lemma A 4 follows from the following two properties: First, for any two tests inducing B there exists a sequence of tests inducing B such that any two successive tests differ only in two quality beliefs and the associated probabilities. Second, by Lemma A 5 (d) with $H(m', m'') = 0$, all tests inducing B that differ only by two quality beliefs and the associated probabilities yield the same ad interim expected utility for the agent.

Part II. We now show the implications of Lemma A 4 for HARA utility functions v_A. Consider $m'' > m'$. Since v_A is continuous and strictly increasing on $(0,1)$, $v_A(m'') - v_A(m') > 0$ and $v_A'(m')v_A'(m'') \geq 0$. So, $\text{sgn}(H(m', m'')) = \text{sgn}\left([v_A(m'') - v_A(m')] - (m'' - m')(v_A'(m')v_A'(m''))^{\frac{1}{2}}\right)$. We use notation $\bar{m} := \frac{m''+m'}{2}$ and $\epsilon := \frac{m''-m'}{2}$, $m'' = \bar{m} + \epsilon$ and $m' = \bar{m} - \epsilon$. Define $G_{\bar{m}}(\epsilon) := [v_A(\bar{m} + \epsilon) - v_A(\bar{m} -$

$\epsilon)] - 2\epsilon v_A'(\bar{m}+\epsilon)^{\frac{1}{2}} v_A'(\bar{m}-\epsilon)^{\frac{1}{2}}$ such that $\text{sgn}(H(m',m'')) = \text{sgn}(G_{\bar{m}}(\epsilon))$. We have

$$\begin{aligned}
G_{\bar{m}}(\epsilon) &= G_{\bar{m}}(0) + \int_0^\epsilon G'_{\bar{m}}(e)\text{d}e \\
&= 0 + \int_0^\epsilon \Big[v_A'(\bar{m}+e) + v_A'(\bar{m}-e) - 2v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \\
&\qquad -e\frac{v_A''(\bar{m}+e)}{v_A'(\bar{m}+e)^{\frac{1}{2}}} v_A'(\bar{m}-e)^{\frac{1}{2}} + e v_A'(\bar{m}+e)^{\frac{1}{2}} \frac{v_A''(\bar{m}-e)}{v_A'(\bar{m}-e)^{\frac{1}{2}}} \Big] \text{d}e \\
&= \int_0^\epsilon \Big[(v_A'(\bar{m}+e)^{\frac{1}{2}} - v_A'(\bar{m}-e)^{\frac{1}{2}})^2 \\
&\qquad + e\left(-\frac{v_A''(\bar{m}+e)}{v_A'(\bar{m}+e)} + \frac{v_A''(\bar{m}-e)}{v_A'(\bar{m}-e)} \right) v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \Big] \text{d}e \\
&= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \left[\left(\left(\frac{v_A'(\bar{m}-e)}{v_A'(\bar{m}+e)} \right)^{-\frac{1}{4}} - \left(\frac{v_A'(\bar{m}-e)}{v_A'(\bar{m}+e)} \right)^{\frac{1}{4}} \right)^2 \right. \\
&\qquad \left. + e\left(-\frac{v_A''(\bar{m}+e)}{v_A'(\bar{m}+e)} + \frac{v_A''(\bar{m}-e)}{v_A'(\bar{m}-e)} \right) \right] \text{d}e
\end{aligned}$$

Since v_A exhibits HARA, $-\frac{v_A''(m)}{v_A'(m)} = \frac{1}{c_1 m + c_2} = r_a(m)$. It follows for $c_1 = 0$ that the integrand is positive almost everywhere as $\left(-\frac{v_A''(\bar{m}+e)}{v_A'(\bar{m}+e)} + \frac{v_A''(\bar{m}-e)}{v_A'(\bar{m}-e)} \right) = 0$. Moreover, for $c_1 \in \mathbf{R}\backslash\{0\}$, $\frac{v_A'(\bar{m}-e)}{v_A'(\bar{m}+e)} = \left(\frac{c_1(\bar{m}-e)+c_2}{c_1(\bar{m}+e)+c_2} \right)^{-\frac{1}{c_1}} = \left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{-\frac{1}{c_1}}$. It follows

$$\begin{aligned}
G_{\bar{m}}(\epsilon) &= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \left[\left(\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{\frac{1}{4c_1}} - \left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{-\frac{1}{4c_1}} \right)^2 \right. \\
&\qquad \left. + e\left(r_a(\bar{m}+e) - r_a(\bar{m}-e) \right) \right] \text{d}e.
\end{aligned}$$

By the definition of r_a, $e = \frac{1}{2c_1}\left(\frac{1}{r_a(\bar{m}+e)} - \frac{1}{r_a(\bar{m}-e)}\right)$. Using this, we get

$$\begin{aligned}
G_{\bar{m}}(\epsilon) &= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \left[\left(\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{-\frac{1}{4c_1}} - \left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{\frac{1}{4c_1}} \right)^2 \right. \\
&\qquad \left. - \frac{1}{2c_1} \left(\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{-\frac{1}{2}} - \left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \right)^{\frac{1}{2}} \right)^2 \right] \text{d}e.
\end{aligned}$$

If $c_1 < 0$, all parts of the integrand are positive almost everywhere such that $G_{\bar{m}}(\epsilon) > 0$. $H(m',m'') > 0$ follows. Consider $c_1 > 0$. In this case v_A exhibits decreasing absolute risk aversion such that $\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)} \in (0,1)$. Define $R : (0,1] \times (0,\infty) \to \mathbf{R}$ by $(r,c) \mapsto 2c(r^{-\frac{1}{4c}} - r^{\frac{1}{4c}})^2$. We can write

$$\begin{aligned}
G_{\bar{m}}(\epsilon) &= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \frac{1}{2c_1} \left[R\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)}, c_1 \right) - R\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)}, \frac{1}{2} \right) \right] \text{d}e \\
&= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \frac{1}{2c_1} \left[\int_{\frac{1}{2}}^{c_1} \frac{\partial R}{\partial c}\left(\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)}, c \right) \text{d}c \right] \text{d}e \\
&= \int_0^\epsilon v_A'(\bar{m}+e)^{\frac{1}{2}} v_A'(\bar{m}-e)^{\frac{1}{2}} \frac{1}{2c_1} \left[\int_{\frac{1}{2}}^{c_1} \left[\int_{\frac{r_a(\bar{m}+e)}{r_a(\bar{m}-e)}}^1 -\frac{\partial^2 R}{\partial c \partial r}(r,c)\,\text{d}r + \frac{\partial R}{\partial c}(1,c) \right] \text{d}c \right] \text{d}e.
\end{aligned}$$

It can easily be checked that $\frac{\partial R}{\partial c}(1,c) = 0$ for all c and that $\frac{\partial^2 R}{\partial r \partial c}(r,c) = -\ln(r)\frac{1}{2c}\left(r^{\frac{1}{2c}-1} + r^{-\frac{1}{2c}-1} \right) > 0$. The sign of the integrand of the outer integral depends almost everywhere only on whether $c_1 < \frac{1}{2}$,

$c_1 = \frac{1}{2}$ or $c_1 < \frac{1}{2}$. It follows $G_{\bar{m}}(\epsilon) > 0$ almost everywhere for $c_1 \in (0, \frac{1}{2})$, $G_{\bar{m}}(\epsilon) = 0$ for $c_1 = \frac{1}{2}$ and $G_{\bar{m}}(\epsilon) < 0$ almost everywhere for $c_1 \in (\frac{1}{2}, \infty)$. The sign of $H(m', m'')$ follows accordingly. The result follows directly from Part I and from the dependence of the sign of $H(m', m'')$ on c_1.

Proof to Lemma 3

Part I: $\mathbb{M}^2(T, x^B)$ is maximized by a binary test which is not subject to false positive. Fix $B = (t, r) \in \mathcal{B} \backslash \mathcal{B}^a$ and consider tests from $\mathcal{T}_B^b$. Assume to the contrary that the optimal minimal and ordered test $T \in \mathcal{T}_B^b$ is not a binary test which is not subject to false positives. It follows $\exists \sigma', \sigma'' \in \Sigma^Z$: $\mu_{\sigma'}(\mu_Y(x^B), p_b^{\sigma'}, p_g^{\sigma'}) < \mu_{\sigma''}(\mu_Y(x^B), p_b^{\sigma''}, p_g^{\sigma''}) < 1$. As in the proof to Lemma 2, define $\widehat{T}(\epsilon_b, \epsilon_g)$ from T by choosing $\widehat{p}_b^{\sigma''} = p_b^{\sigma''} - \epsilon_b$, $\widehat{p}_b^{\sigma'} = p_b^{\sigma'} + \epsilon_b$, $\widehat{p}_g^{\sigma''} = p_g^{\sigma''} - \epsilon_g$, $\widehat{p}_g^{\sigma'} = p_g^{\sigma'} + \epsilon_g$ and $\forall \sigma \in \Sigma^Z \backslash \{\sigma', \sigma''\} \forall \omega$: $\widehat{p}_\omega^\sigma = p_\omega^\sigma$. Note that $\mu_{\sigma''} < 1$ implies $p_b^{\sigma''} > 0$ and thus $p_b^{\sigma'} < 1$. Moreover, $\mu_{\sigma'} < \mu_{\sigma''}$ requires $p_g^{\sigma''} > 0$ which implies $p_g^{\sigma'} < 1$. It follows that we have $p_b^{\sigma''} - \epsilon_b, p_b^{\sigma'} + \epsilon_b, p_g^{\sigma''} - \epsilon_g, p_g^{\sigma'} + \epsilon_g \in (0, 1)$ for ϵ_g and ϵ_b strictly positive and sufficiently small. I.e., $\widehat{T}(\epsilon_b, \epsilon_g)$ is indeed a test.

As in the proof to Lemma 2, define $\widehat{p}_w^\sigma := w\widehat{p}_g^\sigma + (1 - w)\widehat{p}_b^\sigma$ for $w \in [0, 1]$. Note that $\widehat{p}_0^\sigma = \widehat{p}_b^\sigma$ and $\widehat{p}_1^\sigma = \widehat{p}_g^\sigma$ such that we can also write $\widehat{p}_w^\sigma := w\widehat{p}_1^\sigma + (1 - w)\widehat{p}_0^\sigma$. For $w \in (0, 1)$ define $M^2(\epsilon_b, \epsilon_g) := \widehat{p}_{\mu_Y}^{\sigma'} \mu_{\sigma'}^2 + \widehat{p}_{\mu_Y}^{\sigma''} \mu_{\sigma''}^2$ with $\mu_\sigma = \mu_Y \frac{\widehat{p}_1^\sigma}{\widehat{p}_{\mu_Y}^\sigma}$ and with μ_Y short for $\mu_Y(x^B)$. We obtain the following properties which we will use in the course of the proof.

Lemma A 6 *(a)* $\mu_{\sigma''} - \mu_{\sigma'} = \mu_Y(1 - \mu_Y) \frac{\widehat{p}_0^{\sigma'} \widehat{p}_1^{\sigma''} - \widehat{p}_0^{\sigma''} \widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'} \widehat{p}_{\mu_Y}^{\sigma''}}$. *(b)* $\frac{\partial M^2}{\partial \epsilon_b}(0, 0) > 0$. *(c)* $\frac{\partial M^2}{\partial \epsilon_g}(0, 0) < 0$. *(d) If* $c_1 \geq -1$, $\frac{1}{2}[v_A'(\mu_{\sigma'}) + v_A'(\mu_{\sigma''})][\mu_{\sigma''} - \mu_{\sigma'}] - [v_A(\mu_{\sigma''}) - v_A(\mu_{\sigma'})] \geq 0$ *for* $\epsilon_b = \epsilon_g = 0$. *(e)* $\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} > \frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}$ *for* $\epsilon_b = \epsilon_g = 0$. *(f) If* $c_1 \geq -1$, $\frac{\partial M^2}{\partial \epsilon_b}(0, 0) \frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(0, 0) - \frac{\partial M^2}{\partial \epsilon_g}(0, 0) \frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(0, 0) \geq 0$ *(with* $\widehat{V}_t^Y$ *as defined in the proof to Lemma 2).*

Proof. (a) This is Lemma A 5 (a).

(b) We have

$$\frac{\partial M^2}{\partial \epsilon_b}(\epsilon_b, \epsilon_g) = -(1 - \mu_Y)[\mu_{\sigma''}^2 - \mu_{\sigma'}^2] + \widehat{p}_{\mu_Y}^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} 2\mu_{\sigma'} + \widehat{p}_{\mu_Y}^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} 2\mu_{\sigma''}.$$

Using $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} = -\frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}} \mu_{\sigma'}$ and $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} = \frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}} \mu_{\sigma''}$, we get

$$\begin{aligned} \frac{\partial M^2}{\partial \epsilon_b}(\epsilon_b, \epsilon_g) &= (1 - \mu_Y)[\mu_{\sigma''}^2 - \mu_{\sigma'}^2] \\ &= \mu_Y(1 - \mu_Y)[\mu_{\sigma''} - \mu_{\sigma'}] \left(\frac{\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} + \frac{\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}} \right). \end{aligned}$$

Since $\mu_{\sigma''} > \mu_{\sigma'}$ for $\epsilon_b = \epsilon_g = 0$, $\frac{\partial M^2}{\partial \epsilon_b}(0, 0) > 0$.

(c) We have

$$\frac{\partial M^2}{\partial \epsilon_g}(\epsilon_b, \epsilon_g) = -\mu_Y[\mu_{\sigma''}^2 - \mu_{\sigma'}^2] + \widehat{p}_t^{\sigma'} \frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} 2\mu_{\sigma'} + \widehat{p}_t^{\sigma''} \frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} 2\mu_{\sigma''}.$$

Using $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} = \frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}}(1 - \mu_{\sigma'})$ and $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} = -\frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}}(1 - \mu_{\sigma''})$, we get

$$\frac{\partial M^2}{\partial \epsilon_g}(\epsilon_b, \epsilon_g) = -\mu_Y[\mu_{\sigma''} - \mu_{\sigma'}](2 - (\mu_{\sigma'} + \mu_{\sigma''}))$$

$$= \quad -\mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}]\left(\frac{\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right).$$

Since $\mu_{\sigma''} > \mu_{\sigma'}$ for $\epsilon_b = \epsilon_g = 0$, $\frac{\partial M^2}{\partial \epsilon_g}(0,0) < 0$.

(d) We have

$$\begin{aligned}
& \frac{1}{2}[v_A'(\mu_{\sigma'})+v_A'(\mu_{\sigma''})][\mu_{\sigma''}-\mu_{\sigma'}]-[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})] \\
= & \int_0^{\mu_{\sigma''}-\mu_{\sigma'}} \frac{\mathrm{d}}{\mathrm{d}\epsilon}\left(\frac{1}{2}[v_A'(\mu_{\sigma'})+v_A'(\mu_{\sigma'}+\epsilon)]\epsilon-[v_A(\mu_{\sigma'}+\epsilon)-v_A(\mu_{\sigma'})]\right)\mathrm{d}\epsilon \\
= & \int_0^{\mu_{\sigma''}-\mu_{\sigma'}} \frac{1}{2}\left(v_A''(\mu_{\sigma'}+\epsilon)\epsilon+v_A'(\mu_{\sigma'})-v_A'(\mu_{\sigma'}+\epsilon)\right)\mathrm{d}\epsilon \\
= & \int_0^{\mu_{\sigma''}-\mu_{\sigma'}} \int_0^{\epsilon} \frac{1}{2}v_A'''(\mu_{\sigma'}+\epsilon')\epsilon'\mathrm{d}\epsilon'\mathrm{d}\epsilon.
\end{aligned}$$

Since $\mu_{\sigma''} > \mu_{\sigma'}$ for $\epsilon_b = \epsilon_g = 0$, this expression is positive if $v_A'''(m) \geq 0$. Since $v_A'''(m) \geq 0$ when v_A exhibits HARA with $c_1 \geq -1$, we obtain the result.

(e) We have

$$\begin{aligned}
\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}} > \frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}} \quad \Leftrightarrow \quad & [t\widehat{p}_1^{\sigma'}+(1-t)\widehat{p}_0^{\sigma'}][\mu_Y\widehat{p}_1^{\sigma''}+(1-\mu_Y)\widehat{p}_0^{\sigma''}] \\
& \geq [t\widehat{p}_1^{\sigma''}+(1-t)\widehat{p}_0^{\sigma''}][\mu_Y\widehat{p}_1^{\sigma'}+(1-\mu_Y)\widehat{p}_0^{\sigma'}] \\
\Leftrightarrow \quad & (\mu_Y-t)[\widehat{p}_0^{\sigma'}\widehat{p}_1^{\sigma''}-\widehat{p}_1^{\sigma'}\widehat{p}_0^{\sigma''}] \geq 0.
\end{aligned}$$

For $\epsilon_b = \epsilon_g = 0$, $\mu_{\sigma''} > \mu_{\sigma'}$. From this, $\mu_Y \geq t$ and part (a), it follows that the inequality holds.

(f) Using the formulas from part (b) and (c),

$$\begin{aligned}
& \frac{\partial M^2}{\partial \epsilon_b}(\epsilon_b,\epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_g}(\epsilon_b,\epsilon_g)-\frac{\partial M^2}{\partial \epsilon_g}(\epsilon_b,\epsilon_g)\frac{\partial \widehat{V}_t^Y}{\partial \epsilon_b}(\epsilon_b,\epsilon_g) \\
= \quad & \mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}]\left(\frac{\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right) \\
& \cdot\left[-t[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})]+\widehat{p}_t^{\sigma'}\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g}v_A'(\mu_{\sigma'})+\widehat{p}_t^{\sigma''}\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g}v_A'(\mu_{\sigma''})\right] \\
& +\mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}]\left(\frac{\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right) \\
& \cdot\left[-(1-t)[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})]+\widehat{p}_t^{\sigma'}\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b}v_A'(\mu_{\sigma'})+\widehat{p}_t^{\sigma''}\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b}v_A'(\mu_{\sigma''})\right].
\end{aligned}$$

Using $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_b} = -\frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}}\mu_{\sigma'}$, $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_b} = \frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}}\mu_{\sigma''}$, $\frac{\mathrm{d}\mu_{\sigma'}}{\mathrm{d}\epsilon_g} = \frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}}(1-\mu_{\sigma'})$ and $\frac{\mathrm{d}\mu_{\sigma''}}{\mathrm{d}\epsilon_g} = -\frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}}(1-\mu_{\sigma''})$, we get

$$\begin{aligned}
\ldots \quad = \quad & \mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}] \\
& \cdot\left\{\left[-t\left(\frac{\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)-(1-t)\left(\frac{\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\right][v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})]\right. \\
& +\widehat{p}_t^{\sigma'}v_A'(\mu_{\sigma'})\left[\left(\frac{\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}}(1-\mu_{\sigma'})-\left(\frac{\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma'}}\mu_{\sigma'}\right] \\
& \left.+\widehat{p}_t^{\sigma''}v_A'(\mu_{\sigma''})\left[-\left(\frac{\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\frac{\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}}(1-\mu_{\sigma''})+\left(\frac{\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\frac{1-\mu_Y}{\widehat{p}_{\mu_Y}^{\sigma''}}\mu_{\sigma''}\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}] \\
&\quad \cdot\Bigg\{-\left(\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})] \\
&\quad +\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}v_A'(\mu_{\sigma'})\left[\left(\frac{\mu_Y\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\mu_Y\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)(1-\mu_{\sigma'})-\left(\frac{(1-\mu_Y)\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{(1-\mu_Y)\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\mu_{\sigma'}\right] \\
&\quad +\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}v_A'(\mu_{\sigma''})\left[-\left(\frac{\mu_Y\widehat{p}_1^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\mu_Y\widehat{p}_1^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)(1-\mu_{\sigma''})+\left(\frac{(1-\mu_Y)\widehat{p}_0^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{(1-\mu_Y)\widehat{p}_0^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right)\mu_{\sigma''}\right]\Bigg\} \\
&= \mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}]\left(\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right) \\
&\quad \cdot\left\{\left(\frac{\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}}{\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}}v_A'(\mu_{\sigma'})+\frac{\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}}{\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}}v_A'(\mu_{\sigma''})\right)[\mu_{\sigma''}-\mu_{\sigma'}]-[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})]\right\}.
\end{aligned}
$$

Consider $\epsilon_b=\epsilon_g=0$ such that $\mu_{\sigma''}>\mu_{\sigma'}$. By strict concavity of v_A on $(0,1)$, $v_A'(\mu_{\sigma'})>v_A'(\mu_{\sigma''})$. By (e), the coefficient in front of $v_A'(\mu_{\sigma'})$ is larger than the coefficient in front of $v_A'(\mu_{\sigma''})$. It follows

$$
\begin{aligned}
\ldots &\geq \mu_Y(1-\mu_Y)[\mu_{\sigma''}-\mu_{\sigma'}]\left(\frac{\widehat{p}_t^{\sigma'}}{\widehat{p}_{\mu_Y}^{\sigma'}}+\frac{\widehat{p}_t^{\sigma''}}{\widehat{p}_{\mu_Y}^{\sigma''}}\right) \\
&\quad \cdot\left\{\frac{1}{2}\left(v_A'(\mu_{\sigma'})+v_A'(\mu_{\sigma''})\right)[\mu_{\sigma''}-\mu_{\sigma'}]-[v_A(\mu_{\sigma''})-v_A(\mu_{\sigma'})]\right\}.
\end{aligned}
$$

By (d), this expression is non–negative. q.e.d.

We now consider tests $\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b))$ where $\epsilon_g(\epsilon_b)$ is chosen such that the participation behavior is kept constant and show how $\mathbb{M}^2(\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b)),x^B)$ depends on ϵ_b. Note that there exists C_2 which neither depends on ϵ_b nor on ϵ_g such that $\mathbb{M}^2(\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b)),x^B)=P(x^B)\left(M^2(\epsilon_b,\epsilon_g(\epsilon_b))+C_2\right)+(1-P(x^B))\mu_N(x^B)^2$. To see how $\mathbb{M}^2(\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b)),x^B)$ varies with ϵ_b, we only have to investigate how $M^2(\epsilon_b,\epsilon_g(\epsilon_b))$ varies with ϵ_b. By the Implicit Function Theorem, $\epsilon_g'(\epsilon_b)=-\frac{\partial\widehat{V}_t^Y}{\partial\epsilon_b}(\epsilon_b,\epsilon_g)/\frac{\partial\widehat{V}_t^Y}{\partial\epsilon_g}(\epsilon_b,\epsilon_g)$ (with $\widehat{V}_t^Y$ as defined in the proof to Lemma 2). By Lemma A 5 (b) and (c) in the proof to Lemma 2, $\epsilon_g'(0)>0$ such that test $\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b))$ is indeed a test for sufficiently small but strictly positive ϵ_b. In which direction $\mathbb{M}^2(\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b)),x^B)$ changes when ϵ_b increases slightly is given by the sign of

$$
\frac{\partial M^2}{\partial\epsilon_b}(0,0)+\epsilon_g'(0)\frac{\partial M^2}{\partial\epsilon_g}(0,0)=\frac{\frac{\partial M^2}{\partial\epsilon_b}(0,0)\frac{\partial\widehat{V}_t^Y}{\partial\epsilon_g}(0,0)-\frac{\partial M^2}{\partial\epsilon_g}(0,0)\frac{\partial\widehat{V}_t^Y}{\partial\epsilon_b}(0,0)}{\frac{\partial\widehat{V}_t^Y}{\partial\epsilon_g}(0,0)}.
$$

By Lemma A 5 (c) in the proof to Lemma 2, the denominator is positive such that the sign of the expression is the sign of the numerator. By Lemma A 6 (f), the numerator is non–negative when v_A exhibits HARA with $c_1\geq-1$. Hence, $\mathbb{M}^2(T(\epsilon_b,\epsilon_g(\epsilon_b)),x^B)$ can be increased by slightly increasing ϵ_b. It follows that it is possible to increase the second moment when more than one quality belief is strictly smaller than one. Hence, the second moment is maximized by a binary test for which one belief equals one, i.e., by a binary test which is not subject to false positives.

Part II: $\mathbb{M}^3(T,x^B)$ *is maximized by a binary test which is not subject to false positives.* As in part I, consider tests $\widehat{T}(\epsilon_b,\epsilon_g(\epsilon_b))$ where $\epsilon_g(\epsilon_b)$ is chosen such that the participation behavior is kept constant. We use the following lemma.

Lemma A 7 *Consider* $\mu = \mu(x^B, \widehat{T}(\epsilon_b, \epsilon_g(\epsilon_b)))$. $\mu_{\sigma'}$ *and* $\mu_{\sigma''}$ *both increase in* ϵ_b.

Proof. The proof is analogous to that of Lemma 5 in [31]. q.e.d.

Note that there exists C_3 which neither depends on ϵ_b nor on ϵ_g such that

$$\begin{aligned}\mathbb{M}^3(\widehat{T}(\epsilon_b, \epsilon_g(\epsilon_b)), x^B) &= P(x^B)\left((\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y})\left[\frac{\widehat{p}^{\sigma'}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^3_{\sigma'} + \frac{\widehat{p}^{\sigma''}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^3_{\sigma''}\right] + C_3\right) \\ &\quad +(1 - P(x^B))\mu_N(x^B)^3.\end{aligned}$$

ϵ_b affects only $\frac{\widehat{p}^{\sigma'}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^3_{\sigma'} + \frac{\widehat{p}^{\sigma''}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^3_{\sigma''}$. By Bayes' Law and by the construction of $\widehat{T}(\epsilon_b, \epsilon_g(\epsilon_b))$, ϵ_b does not affect $\frac{\widehat{p}^{\sigma'}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu_{\sigma'} + \frac{\widehat{p}^{\sigma''}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu_{\sigma''}$. By Part I, $\frac{\widehat{p}^{\sigma'}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^2_{\sigma'} + \frac{\widehat{p}^{\sigma''}_{\mu_Y}}{\widehat{p}^{\sigma'}_{\mu_Y} + \widehat{p}^{\sigma''}_{\mu_Y}}\mu^2_{\sigma''}$ is increasing in ϵ_b. By Lemma A 7, $\mu_{\sigma'}$ and $\mu_{\sigma''}$ both increase in ϵ_b. We obtain the result from the following lemma and a reasoning like that at the end of Part I.

Lemma A 8 *When a binary quality belief lottery is changed in a way such that the first moment stays constant, the second moment increases and both quality beliefs increase, then the third moment increases.*

Proof. Consider a parameterized family of binary quality belief distribution where quality belief $m_1(s)$ occurs with probability $p_1(s)$ and quality belief $m_2(s) > m_1(s)$ occurs with probability $p_2(s) = 1 - p_1(s)$. Suppose $m_1(s)$, $m_2(s)$ and $p_1(s)$ are differentiable. Suppose further (i) $p_1(s)m_1(s) + p_2(s)m_2(s) = m_Y$, (ii) $m_1(s)$ and $m_2(s)$ are increasing in s, and (iii) $p_1(s)m_1(s)^2 + p_2(s)m_2(s)^2$ is increasing in s.

(i) implies $p_2(s) = \frac{m_Y - m_1(s)}{m_2(s) - m_1(s)}$ and $p_1(s) = \frac{m_2(s) - m_Y}{m_2(s) - m_1(s)}$. We have $p_1(s)m_1(s)^2 + p_2(s)m_2(s)^2 = m_Y(m_1(s) + m_2(s)) - m_1(s)m_2(s)$. Differentiating this expression, we obtain $m_1'(s)[m_Y - m_2(s)] + m_2'(s)[m_Y - m_1(s)]$. (iii) implies $m_2'(s)[m_Y - m_1(s)] > m_1'(s)[m_2(s) - m_Y]$.

Using (i), we can write $p_1(s)m_1(s)^3 + p_2(s)m_2(s)^3 = -m_1(s)^2 m_2(s) - m_1(s)m_2(s)^2 + m_Y m_2(s)^2 + m_Y m_1(s)m_2(s) + m_Y m_1(s)^2$. Differentiating this expression, we get $-m_1'(s)[m_2(s) - m_Y][2m_1(s) + m_2(s)] + m_2'(s)[m_Y - m_1(s)][m_1(s) + 2m_2(s)]$. By $m_2(s) > m_1(s)$, $[m_1(s) + 2m_2(s)] > [2m_1(s) + m_2(s)]$. By (iii), $m_2'(s)[m_Y - m_1(s)] > m_1'(s)[m_2(s) - m_Y]$. By (i), $m_1'(s)[m_2(s) - m_Y] > 0$. It follows that the derivative of $p_1(s)m_1(s)^3 + p_2(s)m_2(s)^3$ is strictly positive. q.e.d.

Derivation of the results in Subsection 2.4

Statement 1: *For any* $B \in \mathcal{B}$, p_f *is unique.*

Proof: For $B \in \mathcal{B}^a$, $p_f = 0$. It remains to show that for $B \in \mathcal{B}\backslash\mathcal{B}^a$ there exists a unique p_f for which the participation constraint is binding, $V_t^Y(T(p_f), \mu(x^B, T(p_f))) = [t(1-p_f)]v_A(1) + [tp_f + (1-t)]v_A(\mu_1) = v_A(\mu_N)$ with $\mu_1 = \frac{p_f \mu_Y}{p_f \mu_Y + (1-\mu_Y)}$. $B \in \mathcal{B}\backslash\mathcal{B}^a$ implies $V_t^Y(T(p_f), \mu(x^B, T(p_f))) < v_A(\mu_N)$ for $p_f = 0$. The converse is true for $p_f \to 1$. Moreover, $V_t^Y(T(p_f), \mu(x^B, T(p_f)))$ and $v_A(\mu_N)$ are continuous in p_f. It follows that a p_f exists for which the participation constraint is binding. It remains to establish uniqueness. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p_f}V_t^Y(T(p_f), \mu(x^B, T(p_f))) &= -t[v_A(1) - v_A(\mu_1)] + [tp_f + (1-t)]v_A'(\mu_1)\frac{\mu_Y(1-\mu_Y)}{(p_f\mu_Y + (1-\mu_Y))^2} \\ &= -t[v_A(1) - v_A(\mu_1)] + \frac{tp_f\mu_Y + (1-t)\mu_Y}{p_f\mu_Y + (1-\mu_Y)}v_A'(\mu_1)(1-\mu_1) \\ &= -t[v_A(1) - v_A(\mu_1)] + [t + \frac{\mu_Y - t}{p_f\mu_Y + (1-\mu_Y)}]v_A'(\mu_1)(1-\mu_1).\end{aligned}$$

Since $[t + \frac{\mu_Y - t}{p_f \mu_Y + (1-\mu_Y)}] > t$ and since $v'_A(\mu_1)(1-\mu_1) > [v_A(1) - v_A(\mu_1)]$ by strict concavity of v_A on $(0,1)$, $\frac{\mathrm{d}}{\mathrm{d}p_f} V_t^Y(T(p_f), \mu(x^B, T(p_f))) > 0$. Since $v_A(\mu_N)$ is not affected by p_f, we obtain uniqueness.

Statement 2: *Restricting attention to participation behaviors $B = (t, 1)$ with $t \in \Theta$ and $t < \overline{\theta}$ is without loss of generality.*

Proof: Inducing participation with probability zero, i.e. with threshold $t = \overline{\theta}$, cannot be optimal for the principal. As there are no points of mass in the distribution of θ, the mixing probability neither affects the quality beliefs nor the principal's expected utility. That is, the principal is indifferent between any participation behaviors (t, r') and (t, r''). It remains to argue that for any $t < \overline{\theta}$ participation behavior $(t, 1)$ is inducible if participation behavior (t, r) is inducible for some $r \in [0, 1)$. If participation behavior (t, r) is inducible, it is induced with a binding participation constraint, even if $r = 0$. Otherwise, there would exist a signal just above t for which the agent would not want to participate. Since the participation constraint is binding for any $r \in [0, 1)$ and since there are no points of mass in the distribution, the participation constraint is also binding for $r = 1$.

Statement 3: *There exists a unique intersection of $\bar{t}(\delta)$ and $\overline{\theta}(\delta)$, which is at $\delta = \frac{1}{4c_2 - 2}$.*

Proof: We start by proving three properties: *(1): $\bar{t}(\delta)$ is continuous and strictly decreasing.* $\bar{t}$ is continuous as it is a composition of continuous functions. We have $\bar{t}'(\delta) = \frac{\mathrm{d}\underline{\theta}}{\mathrm{d}\delta}(-1 + \frac{1}{2}\frac{4(2c_2-1)}{\sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)}})$. Since $\frac{\mathrm{d}\underline{\theta}}{\mathrm{d}\delta} = -\frac{1}{2}$, we need to show that the bracketed expression is strictly positive. I.e., $4c_2 - 2 > \sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)}$. The inequality is equivalent to $\underline{\theta}(\delta) < c_2 \frac{3c_2 - 2}{2c_2 - 1}$. As the left–hand side is strictly smaller than one and as the right–hand side is larger than one for $c_2 \geq 1$, $\bar{t}'(\delta) < 0$. *(2):* $\lim_{\delta \downarrow 0} \bar{t}(\delta) > \lim_{\delta \downarrow 0} \overline{\theta}(\delta)$. This is $-(2c_2 - 2 + \frac{1}{2}) + \sqrt{(2c_2-2)^2 + 2(2c_2-1)} > \frac{1}{2}$ which is equivalent to $(2c_2-2)^2 + 2(2c_2-1) > (2c_2-1)^2$. It can easily be checked that the inequality holds. *(3):* $\lim_{\delta \uparrow 1} \bar{t}(\delta) = \lim_{\delta \uparrow 1} \underline{\theta}(\delta)$. This is $-(2c_2-2) + \sqrt{(2c_2-2)^2} = 0$, which is obviously true. Existence and uniqueness of an intersection follow by an Intermediate Value Theorem from (1), (2), (3) and from $\overline{\theta}(\delta)$ being continuous and strictly increasing. It remains to prove $\bar{t}(\delta) = \overline{\theta}(\delta)$ for $\delta = \frac{1}{4c_2-2}$. Using $\underline{\theta}(\delta) + \overline{\theta}(\delta) = 1$, $\bar{t}(\delta) = \overline{\theta}(\delta)$ is equivalent to $\sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)} = (2c_2 - 1)$. It follows that $\delta = \frac{1}{4c_2-2}$.

Statement 4: $\bar{t}(\delta) > \underline{\theta}(\delta)$.

Proof: This is $-(2c_2 - 2 + \underline{\theta}(\delta)) + \sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)} > \underline{\theta}(\delta) \Leftrightarrow (2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta) > (2c_2 - 2 + 2\underline{\theta}(\delta))^2 \Leftrightarrow 4\underline{\theta}(\delta) > 4\underline{\theta}(\delta)^2$. This holds for any $\delta \in (0,1)$.

Proof to Result 1: *(1): Inducement with a binding participation constraint.* By Proposition 2 (b), $(t, r) \in \mathcal{B}^b$ if and only if

$$
\begin{aligned}
& t v_A(1) + (1-t) v_A(0) \leq v_A(\mathbf{E}[\theta | \theta \leq t]) \\
\Leftrightarrow \quad & -t\frac{1}{2}(c_2 - 1)^2 - (1-t)\frac{1}{2}(c_2 - 0)^2 \leq -\frac{1}{2}\left(c_2 - \frac{\underline{\theta}(\delta) + t}{2}\right)^2 \\
\Leftrightarrow \quad & t \leq -(2c_2 - 2 + \underline{\theta}(\delta)) + \sqrt{(2c_2-2)^2 + 4(2c_2-1)\underline{\theta}(\delta)} = \bar{t}(\delta). \qquad (5)
\end{aligned}
$$

It follows from Statement 4 that the set of thresholds that is inducible with a binding participation constraint is non–empty. It follows from the proof to Statement 3 that $\bar{t}(\delta) > \overline{\theta}(\delta)$ if $\delta < \frac{1}{4c_2-2}$ and

$\bar{t}(\delta) \leq \overline{\theta}(\delta)$ if $\delta \geq \frac{1}{4c_2-2}$. Hence, if $\delta \in (0, \frac{1}{4c_2-2})$, any $t \in \Theta$ can be induced with a binding participation constraint, if $\delta \in [\frac{1}{4c_2-2}, 1)$, only $t \in [\underline{\theta}(\delta), \bar{t}(\delta)]$ can be induced with a binding participation constraint. *(2): Inducement by an accurate test.* By strict concavity of v_A, $t = \underline{\theta}(\delta)$ can not be induced by an accurate test. By continuity of the distribution, any participation behavior $(t, 1)$ with $t \in (\underline{\theta}(\delta), \overline{\theta}(\delta))$ can only be induced by an accurate if the participation constraint is binding for an accurate test. By (5), $(t, 1)$ is inducible by an accurate test if and only if $t = \bar{t}(\delta)$.

Proof to Result 2: If t is inducible, it is by Result 1 inducible with a binding participation constraint, $PC(t, p_f) := V_t^Y(T(p_f), \mu(x^B, T(p_f))) - v_A(\mu_N(x^B)) = 0$. By Statement 1, for any inducible t, there exists a unique p_f for which the participation constraint is binding. By the Implicit Function Theorem, $\frac{\mathrm{d}p_f}{\mathrm{d}t} = -\frac{\partial PC(t,pf)}{\partial t} / \frac{\partial PC(t,pf)}{\partial p_f}$. $\frac{\partial PC(t,pf)}{\partial p_f} > 0$ follows from the proof to Statement 1. To prove $\frac{\mathrm{d}p_f}{\mathrm{d}t} < 0$, it suffices to show $\frac{\partial PC(t,pf)}{\partial t} > 0$. We have

$$\begin{aligned} \frac{\partial PC(t,p_f)}{\partial t} &= (1-p_f)[v_A(1) - v_A(\mu_1(\cdot))] \\ &+ \frac{1}{2}[tp_f + (1-t)]\frac{p_f}{[\mu_Y(\cdot)p_f + (1-\mu_Y(\cdot))]^2} v_A'(\mu_1(\cdot)) - \frac{1}{2}v_A'(\mu_N(\cdot)). \end{aligned}$$

First, $\mu_Y(\cdot) > t$ implies $\frac{tp_f+(1-t)}{\mu(\cdot)p_f+(1-\mu_Y(\cdot))} > 1$. Second, necessary for the participation constraint being binding is $\mu_1(\cdot) < \mu_N(\cdot)$. This and the concavity assumption on v_A imply $v_A'(\mu_1(\cdot)) > v_A'(\mu_N(\cdot))$. Using these two properties, we obtain

$$\begin{aligned} \frac{\partial PC(t,p_f)}{\partial t} &> (1-p_f)[v_A(1) - v_A(\mu_1(\cdot))] \\ &+ \frac{1}{2}\frac{p_f}{\mu_Y(\cdot)p_f + (1-\mu_Y(\cdot))} v_A'(\mu_N(\cdot)) - \frac{1}{2}v_A'(\mu_N(\cdot)) \\ &= (1-p_f)[v_A(1) - v_A(\mu_1(\cdot))] + \frac{1}{2}v_A'(\mu_N(\cdot))\left\{\frac{p_f}{\mu_Y(\cdot)p_f + (1-\mu_Y(\cdot))} - 1\right\}. \end{aligned}$$

Now observe that the expression in curly brackets corresponds to $-(1-p_f)(1-\mu_1(\cdot))$. Hence,

$$\frac{\partial PC(t,p_f)}{\partial t} > (1-p_f)\left\{[v_A(1) - v_A(\mu_1(\cdot))] - \frac{1}{2}v_A'(\mu_N(\cdot))(1-\mu_1(\cdot))\right\}.$$

Using the quadratic structure of v_A, we get $v_A(1) - v_A(\mu_1(\cdot)) = \frac{1}{2}(1-\mu_1(\cdot))[2c_2 - (1+\mu_1(\cdot))]$ and $v_A'(\mu_N(\cdot)) = (c_2 - \mu_N(\cdot))$. With this we obtain

$$\frac{\partial PC(t,p_f)}{\partial t} > \frac{1}{2}(1-p_f)(1-\mu_1(\cdot))\{(c_2-1) + (\mu_N(\cdot) - \mu_1(\cdot))\}.$$

Since the right–hand side is positive, $\frac{\partial PC(t,p_f)}{\partial t} > 0$.

Statement 5: *For any fixed $\delta \in (0,1)$, learning through private information is the better, the closer the participation threshold lies to $\frac{1}{2}$.*

Proof: We prove the statement by showing that for any $\delta \in (0,1)$ the principal's expected utility from learning whether the agent's information is smaller or larger than t is strictly concave in t with maximum at $t = \frac{1}{2}$.

The principal's expected utility from learning whether the agent's information is smaller or larger than t is

$$F(t)v_P(\mathbf{E}[\theta|\theta < t]) + (1-F(t))v_P(\mathbf{E}[\theta|\theta \geq t]).$$

Using $\frac{\mathrm{d}\mathbf{E}[\theta|\theta<t]}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{E}[\theta|\theta\geq t]}{\mathrm{d}t} = \frac{1}{2}$ and $f(t) = \frac{1}{\delta}$, the first derivative is

$$\frac{1}{\delta}\left\{v_P(\mathbf{E}[\theta|\theta < t]) - v_P(\mathbf{E}[\theta|\theta \geq t])\right\} + F(t)\frac{1}{2}v'_P(\mathbf{E}[\theta|\theta < t]) + (1-F(t))\frac{1}{2}v'_P(\mathbf{E}[\theta|\theta \geq t])$$

and the second derivative is

$$\frac{1}{\delta}\left\{v'_P(\mathbf{E}[\theta|\theta < t]) - v'_P(\mathbf{E}[\theta|\theta \geq t])\right\} + F(t)\frac{1}{4}v''_P(\mathbf{E}[\theta|\theta < t]) + (1-F(t))\frac{1}{4}v''_P(\mathbf{E}[\theta|\theta \geq t]).$$

Consider first the second derivative. Since $v_P(m)$ is quadratic, $v''_P(m)$ is constant, say $v''_P(m) := v''_P$, and we have $v'_P(m_1) - v'_P(m_2) = v''_P \cdot (m_1 - m_2)$. Using this, the second derivative becomes

$$\frac{1}{\delta}v''_P \cdot \{\mathbf{E}[\theta|\theta < t] - \mathbf{E}[\theta|\theta \geq t]\} + \frac{1}{4}v''_P = \frac{1}{\delta}v''_P \cdot \left\{-\frac{1}{2}\delta\right\} + \frac{1}{4}v''_P = -\frac{1}{4}v''_P.$$

Since $v_P(m)$ is strictly convex, the principal's expected utility from learning whether θ is above or below t is strictly concave in t. Sufficient for the second part of the result is that the first derivative is zero at $t = \frac{1}{2}$. That is,

$$\frac{1}{\delta}\left\{v_P(\mathbf{E}[\theta|\theta < \frac{1}{2}]) - v_P(\mathbf{E}[\theta|\theta \geq \frac{1}{2}])\right\} + \frac{1}{4}v'_P(\mathbf{E}[\theta|\theta < \frac{1}{2}]) + \frac{1}{4}v'_P(\mathbf{E}[\theta|\theta \geq \frac{1}{2}]) = 0.$$

Since $v_P(m)$ is quadratic, we can write $v_P(m) = \frac{1}{2}v''_P \cdot m^2 + v'_P(0)m + v_P(0)$. Using this, we obtain

$$\frac{1}{\delta}(\mathbf{E}[\theta|\theta < \frac{1}{2}] - \mathbf{E}[\theta|\theta \geq \frac{1}{2}])\left\{\frac{1}{2}v''_P \cdot (\mathbf{E}[\theta|\theta < \frac{1}{2}] + \mathbf{E}[\theta|\theta \geq \frac{1}{2}]) + v'_P(0)\right\}$$
$$+\frac{1}{4}\left\{v''_P \cdot (\mathbf{E}[\theta|\theta < \frac{1}{2}] + \mathbf{E}[\theta|\theta \geq \frac{1}{2}]) + 2v'_P(0)\right\} = 0.$$

Using $\mathbf{E}[\theta|\theta < \frac{1}{2}] - \mathbf{E}[\theta|\theta \geq \frac{1}{2}] = -\frac{1}{2}\delta$ and $\mathbf{E}[\theta|\theta < \frac{1}{2}] + \mathbf{E}[\theta|\theta \geq \frac{1}{2}] = 1$ yields the result.

Proof to Result 3: Follows from the structure of the set of inducible participation behavior (see Result 1).

Proof to Proposition 4

Since inducing participation with probability zero cannot be optimal, the proposition follows directly from the following lemma.

Lemma A 9 *Fix any $T \in \mathcal{T}$. (a) If (x, μ) specifies an equilibrium of the reduced–form game, then there exists $\{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z} \in \mathcal{F}$ such that $(x, y, \{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z})$ with $y(\widehat{\sigma}) = \frac{\mu_{\widehat{\sigma}}}{\mu_{\widehat{\sigma}}+(1-\mu_{\widehat{\sigma}})\gamma_R}$ specifies an equilibrium of the non–reduced game. Moreover, $\mathcal{V}_P(T, x, \mu) = \widetilde{\mathcal{V}}_P(T, x, y)$. (b) If $(x, y, \{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z})$ specifies an equilibrium of the non–reduced game and $P(x) > 0$, then $(\widetilde{x}, \mu)$ with $\widetilde{x}(\theta) = x(\theta)$ if $P(x) < 1$ and $\widetilde{x}(\theta) = 1$ if $P(x) = 1$, and with $\mu_N := \mu_N(\widetilde{x})$ and $\mu_\sigma := \mu_\sigma(\mu_Y(\widetilde{x}), p^\sigma_b, p^\sigma_g)$ specifies an equilibrium of the reduced–form game. Moreover, $\mathcal{V}_P(T, \widetilde{x}, \mu) = \widetilde{\mathcal{V}}_P(T, x, y)$.*

Proof. (a) Suppose (I) and (II) hold for (x, μ) and $P(x) > 0$. Define $F^\omega_{\widehat{\sigma}}(\theta)$ by (2) whenever Bayes' Law is applicable. I.e., for $\widehat{\sigma} = N$ if $P(x) < 1$ and for $\widehat{\sigma} \in \Sigma^Z$ if $P(x) > 0$. If $P(x) = 0$, define for any $\sigma \in \Sigma^Z$ and for $\omega \in \{g, b\}$, $F^\omega_\sigma(\theta) = \mathbf{1}_{\theta\geq\overline{\theta}}\frac{q^\omega_{\overline{\theta}}p^\sigma_\omega}{q^g_{\overline{\theta}}p^\sigma_g+q^b_{\overline{\theta}}p^\sigma_b}$. If $P(x) = 1$, define for $\omega \in \{g, b\}$, $F^\omega_N(\theta) = \mathbf{1}_{\theta\geq\underline{\theta}}q^\omega_{\underline{\theta}}$.

It is easily confirmed that $\{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z} \in \mathcal{F}$. By construction of $\{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z}$, (III') is satisfied. By (I), $x_\theta = x(\theta)$ maximizes

$$\begin{aligned}
& x_\theta \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega v_A(\mu_\sigma) \right) + (1 - x_\theta) v_A(\mu_N) \\
\overset{(3),(4)}{=} \; & x_\theta \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega \widetilde{v}_A\left(\frac{\mu_\sigma}{\mu_\sigma + (1-\mu_\sigma)\gamma_R}\right) \right) + (1 - x_\theta) \widetilde{v}_A\left(\frac{\mu_N}{\mu_N + (1-\mu_N)\gamma_R}\right) \\
= \; & x_\theta \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega \widetilde{v}_A(y(\sigma)) \right) + (1 - x_\theta) \widetilde{v}_A(y(N)).
\end{aligned}$$

As this is the objective function in (I'), it follows that $x_\theta = x(\theta)$ maximizes also (I'). Next, we show that for any $\widehat{\sigma} \in \widehat{\Sigma}^Z$, $a = y(\widehat{\sigma})$ maximizes

$$\sum_\omega \int_\Theta \widetilde{v}_P(a,\omega) \mathrm{d}F^\omega_{\widehat{\sigma}}(\theta) \;=\; -\int_\Theta \gamma_P (a-0)^2 \mathrm{d}F^b_{\widehat{\sigma}}(\theta) - \int_\Theta (a-1)^2 \mathrm{d}F^g_{\widehat{\sigma}}(\theta).$$

The first order condition is given by

$$-2\gamma_P a \int_\Theta \mathrm{d}F^b_{\widehat{\sigma}}(\theta) - 2(a-1) \int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta) = 0.$$

Since the problem is strictly concave in a, a global maximum is assumed at

$$a = \frac{\int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta)}{\int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta) + \gamma_R \int_\Theta \mathrm{d}F^b_{\widehat{\sigma}}(\theta)} = \frac{\int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta)}{\int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta) + \gamma_R (1 - \int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta))}.$$

It remains to show $\int_\Theta \mathrm{d}F^g_{\widehat{\sigma}}(\theta) = \mu_{\widehat{\sigma}}$. Consider first $\widehat{\sigma} = N$. If $P(x) < 1$, (2) implies $\int_\Theta \mathrm{d}F^g_N(\theta) = F^g_N(\overline{\theta}) = \frac{\int_\Theta q^g_\theta (1-x(\theta))\mathrm{d}F(\theta)}{\int_\Theta (1-x(\theta))\mathrm{d}F(\theta)}$. If $P(x) = 1$, $\int_\Theta \mathrm{d}F^g_N(\theta) = \underline{\theta}$. In both cases, $\int_\Theta \mathrm{d}F^g_N(\theta)$ coincides with μ_N as introduced in (II). Consider now $\widehat{\sigma} = \sigma \in \Sigma^Z$. If $P(x) > 0$, (2) implies $\int_\Theta \mathrm{d}F^g_\sigma(\theta) = F^g_\sigma(\overline{\theta}) = \frac{\int_\Theta q^g_\theta p^\sigma_g x(\theta) \mathrm{d}F(\theta)}{\int_\Theta [q^g_\theta p^\sigma_g + q^b_\theta p^\sigma_b] x(\theta) \mathrm{d}F(\theta)}$ which can be rewritten as

$$\frac{\int_\Theta q^g_\theta x(\theta) \mathrm{d}F(\theta) p^\sigma_g}{\int_\Theta q^g_\theta x(\theta) \mathrm{d}F(\theta) p^\sigma_g + \int_\Theta q^b_\theta x(\theta) \mathrm{d}F(\theta) p^\sigma_b} \;=\; \frac{\frac{\int_\Theta q^g_\theta x(\theta)\mathrm{d}F(\theta)}{\int_\Theta x(\theta)\mathrm{d}F(\theta)} p^\sigma_g}{\frac{\int_\Theta q^g_\theta x(\theta)\mathrm{d}F(\theta)}{\int_\Theta x(\theta)\mathrm{d}F(\theta)} p^\sigma_g + (1 - \frac{\int_\Theta q^g_\theta x(\theta)\mathrm{d}F(\theta)}{\int_\Theta x(\theta)\mathrm{d}F(\theta)}) p^\sigma_b}.$$

If $P(x) = 0$, $\int_\Theta \mathrm{d}F^g_\sigma(\theta) = \frac{\overline{\theta} p^\sigma_g}{\overline{\theta} p^\sigma_g + (1-\overline{\theta}) p^\sigma_b}$. In both cases, $\int_\Theta \mathrm{d}F^g_\sigma(\theta)$ coincides with μ_σ as introduced in (II). (II') follows. We obtain that $(x, y, \{(F^g_{\widehat{\sigma}}, F^b_{\widehat{\sigma}})\}_{\widehat{\sigma}\in\widehat{\Sigma}^Z})$ is an equilibrium of the non–reduced game.

We now show that the expected utility of the principal is equal in both cases. Suppose first $P(x) \in (0, 1)$.

$$\begin{aligned}
\mathcal{V}_P(T, x, \mu) \;&=\; \int_\Theta \left[x(\theta) \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega v_P(\mu_\sigma) \right) + (1 - x(\theta)) v_P(\mu_N) \right] \mathrm{d}F(\theta) \\
&=\; \int_\Theta \Bigg[x(\theta) \left(\sum_{\omega,\sigma} q^\omega_\theta p^\sigma_\omega (\mu_\sigma \widetilde{v}_P(y(\sigma), g) + (1-\mu_\sigma) \widetilde{v}_P(y(\sigma), b)) \right) \\
&\quad + (1 - x(\theta))(\mu_N \widetilde{v}_P(y(N), g) + (1-\mu_N) \widetilde{v}_P(y(N), b)) \Bigg] \mathrm{d}F(\theta) \\
&=\; \sum_\sigma \left[\int_\Theta x(\theta) [q^g_\theta p^\sigma_g + q^b_\theta p^\sigma_b] \mathrm{d}F(\theta) \right] (\mu_\sigma \widetilde{v}_P(y(\sigma), g) + (1-\mu_\sigma) \widetilde{v}_P(y(\sigma), b)) \\
&\quad + \left[\int_\Theta (1 - x(\theta)) \mathrm{d}F(\theta) \right] (\mu_N \widetilde{v}_P(y(N), g) + (1-\mu_N) \widetilde{v}_P(y(N), b))
\end{aligned}$$

Using $\mu_N = \frac{\int_\Theta (1-x(\theta))q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta (1-x(\theta))\mathrm{d}F(\theta)}$ and $\mu_\sigma = \frac{\int_\Theta x(\theta) q_\theta^g p_g^\sigma \mathrm{d}F(\theta)}{\int_\Theta x(\theta)[q_\theta^g p_g^\sigma + q_\theta^b p_b^\sigma]\mathrm{d}F(\theta)}$ for $P(x) \in (0,1)$, we get

$$
\begin{aligned}
\ldots &= \sum_\sigma \left(\int_\Theta x(\theta) q_\theta^g p_g^\sigma \mathrm{d}F(\theta) \widetilde{v}_P(y(\sigma), g) + \int_\Theta x(\theta) q_\theta^b p_b^\sigma \mathrm{d}F(\theta) \widetilde{v}_P(y(\sigma), b) \right) \\
&\quad + \left(\int_\Theta (1-x(\theta)) q_\theta^g \mathrm{d}F(\theta) \widetilde{v}_P(y(N), g) + \int_\Theta (1-x(\theta)) q_\theta^b \mathrm{d}F(\theta) \widetilde{v}_P(y(N), b) \right) \\
&= \int_\Theta \left[x(\theta) \left(\sum_{\omega,\sigma} q_\theta^\omega p_\omega^\sigma \widetilde{v}_P(y(\sigma), \omega) \right) + (1-x(\theta)) \left(\sum_\omega q_\theta^\omega \widetilde{v}_P(y(N), \omega) \right) \right] \mathrm{d}F(\theta) \\
&= \widetilde{\mathcal{V}}_P(T, x, y).
\end{aligned}
$$

The proofs for $P(x) = 0$ and for $P(x) = 1$ are similar. The only difference is that some of the integrals disappear.

(b) Suppose (I'), (II') and (III') hold for $(x, y, \{(F_{\widehat{\sigma}}^g, F_{\widehat{\sigma}}^b)\}_{\widehat{\sigma} \in \widehat{\Sigma}^Z})$. (II) holds by construction of μ. (II') implies the structure of y as function of $\{(F_{\widehat{\sigma}}^g, F_{\widehat{\sigma}}^b)\}_{\widehat{\sigma} \in \widehat{\Sigma}^Z}$. We obtain $y(\widehat{\sigma}) = \frac{F_{\widehat{\sigma}}^g(\overline{\theta})}{F_{\widehat{\sigma}}^g(\overline{\theta}) + \gamma_R(1 - F_{\widehat{\sigma}}^g(\overline{\theta}))}$. By the weak consistency requirement implicit in the definition of $\mathcal{F}$, there exists a sequence $x_n^Y \in \mathcal{X}$ with $P(x_n^Y) > 0$ and a sequence $x_n^N \in \mathcal{X}$ with $P^N(x_n^Y) < 1$ such that

$$
\begin{aligned}
F_{\widehat{\sigma}}^g(\overline{\theta}) &= \begin{cases} \lim\limits_{n\to\infty} \frac{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta) p_g^\sigma}{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta) p_g^\sigma + \int_\Theta x_n^Y(\theta) q_\theta^b \mathrm{d}F(\theta) p_b^\sigma} & \text{if } \widehat{\sigma} = \sigma \in \Sigma^Z \\ \lim\limits_{n\to\infty} \frac{\int_\Theta (1-x_n^N(\theta)) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta (1-x_n^N(\theta)) \mathrm{d}F(\theta)} & \text{if } \widehat{\sigma} = N \end{cases} \\
&= \begin{cases} \lim\limits_{n\to\infty} \frac{\frac{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta x_n^Y(\theta) \mathrm{d}F(\theta)} p_g^\sigma}{\frac{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta x_n^Y(\theta) \mathrm{d}F(\theta)} p_g^\sigma + (1 - \frac{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta x_n^Y(\theta) \mathrm{d}F(\theta)}) p_b^\sigma} & \text{if } \widehat{\sigma} = \sigma \in \Sigma^Z \\ \lim\limits_{n\to\infty} \frac{\int_\Theta (1-x_n^N(\theta)) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta (1-x_n^N(\theta)) \mathrm{d}F(\theta)} & \text{if } \widehat{\sigma} = N \end{cases}.
\end{aligned}
$$

$m_n^Y := \frac{\int_\Theta x_n^Y(\theta) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta x_n^Y(\theta) \mathrm{d}F(\theta)}$ and $m_n^N := \frac{\int_\Theta (1-x_n^N(\theta)) q_\theta^g \mathrm{d}F(\theta)}{\int_\Theta (1-x_n^N(\theta)) \mathrm{d}F(\theta)}$ define convergent sequences in $[\underline{\theta}, \overline{\theta}]$. It follows

$$
F_{\widehat{\sigma}}^g(\overline{\theta}) = \begin{cases} \lim\limits_{n\to\infty} \mu_\sigma(m_n^Y, p_b^\sigma, p_g^\sigma) & \text{if } \widehat{\sigma} = \sigma \in \Sigma^Z \\ \lim\limits_{n\to\infty} m_n^N & \text{if } \widehat{\sigma} = N \end{cases}.
$$

We distinguish two cases. Case (i): $P(x) \in (0,1)$. Here, $\widetilde{x}(\theta) = x(\theta)$. Since $P(x) > 0$, it follows from (III') $\lim\limits_{n\to\infty} \mu_\sigma(m_n^Y, p_b^\sigma, p_g^\sigma) = \mu_\sigma$ for any σ and thus $v_A(\mu_\sigma) = \widetilde{v}_A(y(\sigma))$ for any σ. Since $P(x) < 1$, it follows from (III') $\lim_{n\to\infty} m_n^N = \mu_N$ and thus $v_A(\mu_N) = \widetilde{v}_A(y(N))$. Since the objective functions in (I') and (I) coincide, (I) follows from (I'). Case (ii): $P(x) = 1$. Here, $\widetilde{x}(\theta) = 1$. By Proposition 1 (b), only threshold participation behavior is inducible. I.e., $x = x^B$ for some $B \in \mathcal{B}$. Case (ii.1): $B = (\underline{\theta}, 1)$. That is, $\widetilde{x}(\theta) = x(\theta)$. (I') implies $\sum_{\omega,\sigma} q_{\underline{\theta}}^\omega p_\omega^\sigma \widetilde{v}_A(y(\sigma)) \geq \widetilde{v}_A(y(N))$. It follows from $m_n^N \in [\underline{\theta}, \overline{\theta}]$, $\lim\limits_{n\to\infty} m_n^N \geq \mu_N$ and thus $v_A(\mu_N) \leq \widetilde{v}_A(y(N))$. Moreover, (III') implies $v_A(\mu_N) = \widetilde{v}_A(y(N))$. It follows $\sum_{\omega,\sigma} q_{\underline{\theta}}^\omega p_\omega^\sigma v_A(\mu_\sigma) \geq v_A(\mu_N)$. Hence, (I) holds for $\widetilde{x} = x$. Case (ii.2): $B = (\underline{\theta}, r)$ with $r \in [0,1)$. $P(x) = 1$ together with $r < 1$ requires $\text{Prob}\{\theta = \underline{\theta}\} = 0$. This has two consequences. First, it requires that $\Theta \cap (\underline{\theta}, \underline{\theta} + \epsilon) \neq \emptyset$ for any $\epsilon > 0$. It follows that (I') implies $\sum_{\omega,\sigma} q_{\underline{\theta}}^\omega p_\omega^\sigma \widetilde{v}_A(y(\sigma)) = \widetilde{v}_A(y(N))$ even for $r = 0$. Second, it implies $\mu_\sigma(\mu_Y(x), p_b^\sigma, p_g^\sigma) = \mu_\sigma$ for any σ and thus $\widetilde{v}_A(y(\sigma)) = v_A(\mu_\sigma)$ for any σ. It follows from $m_n^N \in [\underline{\theta}, \overline{\theta}]$, $\lim\limits_{n\to\infty} m_n^N \geq \mu_N$ and thus $v_A(\mu_N) \leq \widetilde{v}_A(y(N))$. Thus, $\sum_{\omega,\sigma} q_{\underline{\theta}}^\omega p_\omega^\sigma v_A(y(\sigma)) \geq v_A(\mu_N)$. Hence, (I) holds for $\widetilde{x} = 1$. We obtain that $(\widetilde{x}, \mu)$ is an equilibrium of the reduced–form game.

As x and $\widetilde{x}$ differ at most on a set of measure zero, $\mathcal{V}_P(T,\widetilde{x},\mu) = \widetilde{\mathcal{V}}_P(T,x,y)$ is established by a reasoning like in part (a). q.e.d.

References

[1] Carlos Alós-Ferrer and Julien Prat. Job market signaling and employer learning. Journal of Economic Theory, forthcoming.

[2] Jean-Pierre Benoît and Juan Dubra. Why do good cops defend bad cops? *International Economic Review*, 45:787–809, 2004.

[3] Dirk Bergemann and Martin Pesendorfer. Information structures in optimal auctions. *Journal of Economic Theory*, 137:580–609, 2007.

[4] David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, 1951.

[5] K. Burdett and Dale T. Mortensen. Testing for ability in a competitive labor market. *Journal of Economic Theory*, 25:42–66, 1981.

[6] Andrew Caplin and Kfir Eliaz. Aids policy and psychology: A mechanism–design approach. *The RAND Journal of Economics*, 34:631–646, 2003.

[7] Brendan Daley and Brett Green. Market signaling with grades. Mimeo, October 27, 2011.

[8] Sankar De and Prafulla Nabar. Economic implications of imperfect quality certification. *Economics Letters*, 37:333–337, 1991.

[9] David Dranove and Ginger Z. Jin. Quality disclosure and certification. *Journal of Economic Literature*, 48:935–963, 2010.

[10] Emmanuel Farhi, Josh Lerner, and Jean Tirole. Fear of rejection? Tiered certification and transparency. IDEI Working Paper, n. 534, July 2011.

[11] Antoine Faure-Grimaud, Eloïc Peyrach, and Lucía Quesada. The ownership of ratings. *The RAND Journal of Economics*, 40:234–257, 2009.

[12] Juan-José Ganuza and José Penalva. Signal orderings based on dispersion and the supply of private information in auctions. *Econometrica*, 78:1007–1030, 2010.

[13] David Gill and Daniel Sgroi. The optimal choice of pre–launch reviewer. Journal of Economic Theory, forthcoming.

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

[15] Rick Harbaugh and Eric Rasmusen. Coarse grades: Informing the public when certification is voluntary. Mimeo, May 13, 2011.

[16] Jack Hirshleifer. The private and social value of information and the reward to inventive activity. *The American Economic Review*, 61:561–574, 1971.

[17] Michael Hoel, Tor Iversen, Tore Nilssen, and Jon Vislie. Genetic testing in competitive insurance markets with repulsion from chance: A welfare analysis. *Journal of Health Economics*, 25:847–860, 2006.

[18] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101:2590–2615, 2011.

[19] Alessandro Lizzeri. Information revelation and certification intermediaries. *RAND Journal of Economics*, 30:214–231, 1999.

[20] Glenn M. MacDonald. Person–specific information in the labor market. *The Journal of Political Economy*, 88(3):578–597, June 1980.

[21] Charles F. Mason and Frederic P. Sterbenz. Imperfect quality testing and market size. *International Economic Review*, 35:61–86, 1994.

[22] Steven Matthews and Andrew Postlewaite. Quality testing and disclosure. *The RAND Journal of Economics*, 16:328–340, 1985.

[23] Thomas J. Miceli. Optimal prosecution of defendants whose guilt is uncertain. *Journal of Law, Economics, & Organization*, 6:189–201, 1990.

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

[25] Paul R. Milgrom and Robert J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, September 1982.

[26] Masahiro Okuno-Fujiwara, Andrew Postlewaite, and Kotaro Suzumura. Strategic information revelation. *The Review of Ecnomic Studies*, 57(1):25–47, January 1990.

[27] Michael Ostrovsky and Michael Schwarz. Information disclosure and unraveling in matching markets. *American Economic Journal: Microeconomics*, 2:34–63, 2010.

[28] Marco Ottaviani and Andrea Prat. The value of public information in monopoly. *Econometrica*, 69:1673–1683, 2001.

[29] Jeffrey S. Parker and Bruce H. Kobayashi. *Encyclopedia of Law and Economics*, chapter Evidence. Aldershot, Edward Elgar, 1999.

[30] Thomas Philippon and Vasiliki Skreta. Optimal interventions in markets with adverse selection. *American Economic Review*, 102:1–30, 2012.

[31] Frank Rosar and Elisabeth Schulte. Imperfect private information and the design of information–generating mechanisms. Mimeo, March 24, 2010.

[32] Daniel L. Rubinfeld and David E. M. Sappington. Efficient awards and standards of proof in judical proceedings. *The RAND Journal of Economics*, 18:308–315, 1987.

[33] Michael Spence. Job market signaling. *The Quaterly Journal of Economics*, 87(3):355–374, August 1973.

[34] Joseph E. Stiglitz. The theory of “screening,” education and the distribution of income. *The American Economic Review*, 65(3):283–300, June 1975.

[35] Jean Tirole. Overcoming adverse selection: How public intervention can restore market functioning. *American Economic Review*, 102:29–59, 2012.

[36] Matthijs van den Berg, Danielle R. M. Timmermans, Johanna H. Kleinveld, Elisa Garcia, John M. G. van Vugt, and Gerrit van der Wal. Accepting or declining the offer of prenatal screening for congenital defects: test uptake and women’s reasons. *Parental Diagnosis*, 25:84–90, 2005.

[37] Michael Waldman. Information on worker ability: An analysis of investment within the firm. *Information Economics and Policy*, 4:57–80, 1989.

[38] Andrew Weiss. A sorting-cum-learning model of education. *Journal of Political Economy*, 91:420–442, 1983.

[39] Okan Yilankaya. A model of evidence production and optimal standard of proof and penalty in criminal trials. *The Canadian Journal of Economics*, 35:385–409, 2002.