Economic Theory (2024) 77:849–890
https://doi.org/10.1007/s00199-023-01514-z

RESEARCH ARTICLE

Testing under information manipulation

Silvia Martinez-Gorricho[1] · Carlos Oyarzun[2]

Received: 16 August 2022 / Accepted: 19 July 2023 / Published online: 9 August 2023

Abstract
A principal makes a binary decision based on evidence that can be manipulated by a privately informed agent. The principal's objective is to minimize the expected loss associated to type I and II errors. When the principal can commit to an acceptance standard, the optimal test features ex-post inefficient standards, to internalize the agent's manipulation incentives. We provide conditions for the principal to set soft or harsh standards, that is, lower or higher standards, respectively, than the ex-post optimal standard. When misaligned manipulation (i.e., manipulation by the low type) is dominant, the principal sets soft standards when the prior probability that the candidate is low type is relatively small. In contrast, when aligned manipulation (i.e., manipulation by the high type) is dominant, the principal sets soft standards when the prior probability that the candidate is low type is relatively large. In both scenarios, these soft standards result in that the non-commitment equilibrium outcome is Pareto dominated by the equilibrium outcome under commitment. We also provide conditions for the optimal revelation mechanism to Pareto dominate commitment when the prior probability that the agent is low type is relatively large.

Keywords Information manipulation · Commitment · Moral hazard · Soft standards · Harsh standards · Standard of evidence

JEL Classification C72 · D82 · D86

✉ Carlos Oyarzun
c.oyarzun@uq.edu.au

Silvia Martinez-Gorricho
smartinez@ucsc.cl

[1] Department of Economics, Universidad Católica de la Santísima Concepción, A. de Ribera 2850, Concepción, Chile

[2] School of Economics, University of Queensland, Bldg. 39, Brisbane, QLD, Australia

1 Introduction

There are many real world examples in which agents devote resources to influence an assessment. For instance, in the wake of the replicability crisis, the scientific community has come to realize that methodologies and practices followed for decades, or even centuries, are not immune to malign incentives and research misconduct in particular.[1] A number of alternative solutions have been put forth; for instance, recently, the Ministry of Science and Technology and courts in China have moved towards hard penalties for scientific misconduct—the most extreme even considering the death penalty.[2] However, in many fields, in practice it is nearly impossible to detect research misconduct and even more difficult to prove it, rendering potential punishment virtually ineffective.[3] The logic of the problem is not exclusive to research misconduct. In civil litigation, evidence tampering is pervasive. Sanchirico (2004) points out that *"according to many judges and practitioners[,]...documents that should be produced in response to a discovery request are regularly shredded, altered, or suppressed."*

Our paper analyzes the management of information manipulation.[4] We consider a decision maker facing a binary decision problem—approval or rejection. The decision maker's objective is to minimize the expected loss associated to type I and II errors; i.e., approving a low type agent and rejecting a high type. On the other hand, the agent's preference for approval is type-independent. Before choosing an action, the decision maker observes evidence that is partially informative about the agent's type. However, the agent may exert hidden efforts to alter the evidence, in order to improve the chances of a favorable decision.

We consider two forms of information manipulation. *Test-defensiveness* in psychology is a useful application to illustrate. On one hand, according to Butcher (2002), *"When taking psychological tests at pre-employment, pilots who have personality problems and other mental health symptoms can respond in a way to 'mask' those problems"*.[5] On the other hand, Butcher (1994) suggests that high average performance of pilots in psychological tests can be explained by fit pilots' attempts to display "overly favorable response patterns." Therefore, high type agents may also manipulate evidence; and in some contexts, even to a greater extent than low type agents. For example, consider the problem of admission in graduate school programs requiring specific skills or background. Many schools use qualifying or preliminary exams in their in-program selection procedures (e.g., economics, mathematics, and engineering). Qualified candidates' readiness for the exams can increase substantially with their exerted effort, which is likely to be determined by approval cut-offs. In con-

[1] Di Tillio et al. (2017) provide a historical account of the development of experimental methods.

[2] Source: https://www.statnews.com/2017/06/23/china-death-penalty-research-fraud/ (STATNEWS June 23, 2017).

[3] Although punishment is not completely ineffective in all fields, within several of them, uncovering misconduct may not be practical (see, e.g., Fanelli 2009).

[4] In the absence of concerns about collateral effects on the agent's incentives, such decisions are analyzed as standard statistical decision problems (see, e.g., Neyman and Pearson 1933; Karlin and Rubin 1956, and DeGroot 2005).

[5] Airlines' screening of pilots was subject to intensive scrutiny in 2015, in the aftermath of a Germanwings plane crash in the Alps, believed to be deliberately caused by the pilot.

trast, unqualified candidates' readiness may increase very little due to lack of skills or a weak background.[6] Since the decision maker's and high type agents' incentives are aligned, we refer to the effort of the latter as *aligned manipulation*. In contrast, we refer to the effort exerted by low type agents as *misaligned manipulation*. Both types of manipulation, however, are entirely wasteful: they generate a cost to the agent without altering the probability that her type is high. For instance, in the test-defensiveness application, preparation for a psychological test is unlikely to affect the psychological fitness of a pilot.

First, we analyze a simple model of *commitment* to a standard of evidence. In order to discourage misaligned manipulation and incentivize aligned manipulation, the resulting standards under commitment differ from optimal statistical decision-making. We characterize the direction of the deviations from ex-post optimality in terms of two interacting factors: the manipulation effect that is dominant (misaligned or aligned), and the prior probability that the agent's type is low (Propositions 1-4). Second, we analyze a revelation mechanism for this problem and show that offering a decision rule that only relies on a standard for agents reporting a high type is optimal. Since the decision maker is solely concerned with encouraging manipulation when designing the optimal mechanism, the corresponding standard and the standard under commitment may be biased in opposite directions with respect to the ex-post efficient standard.

Commitment improves the expected payoff of the decision maker. Perhaps surprisingly, it generically involves lowering the standard compared to the case without commitment within a non-trivial set of prior beliefs—therefore leading to outcomes that are Pareto superior (Corollaries 1 and 2). In turn, the optimal mechanism Pareto dominates commitment for relatively high priors that the type is low (Propositions 5 and 6), as for such priors, the revelation mechanism results in lower standards than commitment.

We characterize when optimal standards of evidence are *harsh* or *soft*; i.e., require, respectively, more or less favorable evidence than optimal statistical tests for choosing the agent's preferred action.[7] It is instructive to start the analysis with the case of *pure misaligned manipulation*, in which the high type is non-responsive to changes in the standard. For large prior probabilities that the type is high, the ex-ante optimal standard is soft (Proposition 1). This is because soft standards help decision making if low type agents' effort is a strategic complement of the standard. In turn, under the MLRP assumption, strategic complementarity develops for the low standards that arise in equilibrium, when the agent's type is likely to be high (Lemmata 1 and 2). An analogous reasoning reveals that harsh standards are ex-ante optimal when the agent's type is likely to be low.

In contrast, the results reverse in the *pure aligned manipulation* model. In this model, for large prior probabilities that the type is high, the ex-ante optimal standard is harsh; and, for large prior probabilities that the type is low, the ex-ante optimal standard is soft (Proposition 2). Therefore, a key policy insight is that the agent's

[6] For an empirical analysis of the determinants of success in qualifying exams, thesis completion, and research productivity in economics Ph.D. programs, see Grove and Wu (2007).

[7] Commitment does not always lead to ex-post inefficiencies (see, e.g., Li and Suen 2004; Ben-Porath et al. 2019 and Vohra et al. 2021).

preferred action requires less favorable evidence either when the agent's type is likely to be high under pure misaligned manipulation (due to the strategic complementarity between the standard and effort), or when the agent's type is likely to be low under pure aligned manipulation (due to the strategic substitutability between the standard and effort).[8]

Next, we consider the general case in which the decision maker is concerned with both misaligned manipulation and aligned manipulation. The extent of the responsiveness to the standard of each agent type's effort plays a crucial role in the analysis. We consider a parametrized version of the model that allows us to vary both the high type's *natural* readiness and the manipulation cost of the low type. By changing these parameters we are able to modify the responsiveness of aligned and misaligned manipulation, respectively, to changes in the standard. In particular, we show formally that our results for the pure misaligned manipulation case are robust within a range of large values for the natural readiness of the high type (Proposition 3), provided that the marginal benefits to preparation vanish as readiness increases. Analogously, our results for the pure aligned manipulation case are robust within a range of high effort costs of the low type (Proposition 4). For intermediate values of the parameters, there is a tension between encouraging aligned manipulation and discouraging misaligned manipulation.

Finally, we consider revelation mechanisms without transfers: given a reported type, the decision maker sets probabilities of outright acceptance, outright rejection, or taking a test. In the optimal mechanism, only agents reporting a high type are tested. Setting a strictly positive outright acceptance probability for the low type allows the decision maker to extract this agent's manipulation cost and obtain a higher expected payoff than when he commits to a standard. We provide conditions such that, for large prior probabilities that the agent's type is low, the agent prefers the optimal mechanism over simple commitment to a standard, and the opposite is true when those probabilities are small (Proposition 6). This is because the decision maker is solely concerned about incentivising the high type's effort, since the low type is not tested in the optimal mechanism, and strategic substitutability (complementarity) is induced in equilibrium for large (small) prior probabilities that the agent's type is low.

1.1 Related literature

An early antecedent to our work is Li (2001). Both his and this paper highlight that the quality of information is endogenous to decision-making. While he focuses on mitigating free-riding by committee members, we focus on discouraging effort by low type agents and encouraging effort by high type agents—who, in Li's model, are non-strategic. In his setup, committee members' effort reduces the variance of the signal; in ours, agents' effort shifts probability mass to the right. Li (2001) shows

[8] Soft standards for new drugs with good prior prospects are consistent with recent findings on the approval of new drugs for which the FDA has granted a Breakthrough-Drug Designation. This designation is given based on preliminary evidence to drugs that could provide a substantial improvement to what is available on the market. A number of drugs that received this designation, however, were approved by the FDA, despite subsequent trials showing limited efficacy. Lowering the standard may benefit decision making by discouraging misaligned manipulation (see Sect. 8 for further discussion).

that standards biased against the decision favored by prior beliefs are ex-ante optimal. As a counterpart, in our setup, when the decision maker is mainly concerned with incentivizing aligned manipulation, ex-ante optimal standards are harsh if the prior probability that the type is high is large, and soft otherwise.

The relation between Li's model and ours is parallel to that between Espinosa and Ray (2020) and de Haan et al. (2011). In Espinosa and Ray (2020), the agent can increase the variance of the signal distribution; in contrast, in de Haan et al. (2011), the agent can shift probability mass to the right by paying a cost. Both papers focus on a static game, without commitment by the principal, whereas our analysis focuses on how this commitment distorts otherwise ex-post efficient standards.

Three concurrent papers analyze problems that are closely related to our model: Cunningham and Moreno de Barreda (2019) show that costly signal-jamming improves a sender's probability of persuading the receiver in a model with uniformly distributed types. In their model, signal-jamming makes the receiver worse-off and commitment always leads to harsh standards. In contrast, in our model, information manipulation may be dominated by aligned manipulation, making the decision-maker (the receiver) better-off; and, under pure or dominant misaligned or aligned manipulation, commitment *always* leads to soft standards for a range of priors. Ball (2020) analyzes how the receiver's commitment problem can be mitigated by introducing an intermediary who distorts and coarsens primitive information. Finally, Frankel and Kartik (2022) show that "underutilizing data" is optimal when discouraging an agent to "game" a scoring system. As a counterpart, in our setup, when the decision maker is mainly concerned with desincentivizing misaligned manipulation, ex-ante optimal standards are soft if the prior probability that the type is high is large, and harsh otherwise. The driving force of our result, however, is different from that in Frankel and Kartik (2022). As explained above, strategic complementarity between the standard and the agent's effort leads the decision maker to use soft standards when the probability that the agent is low-type is small. This connection between strategic complementarity and priors relies on the MLRP satisfied by the family of signal distributions. Analogously, the connection between strategic substitutibility and priors beliefs results in harsh standards when the probability that the agent is low-type is large. The antithesis to what the papers above and ours do is considered in Degan and Li (2021), where the sender can commit to a certain level of precision of her signal, in a setup without moral hazard. In equilibrium, the sender commits to make her signal somewhat precise only if the priors about the type are uniform or slightly pessimistic.

Taylor and Yildirim (2011) consider how evidence standards play a dual role: as a selection criterion and as a tool to incentivize an agent whose effort increases the probability that her type is high.[9] Although their analysis focuses on comparing blind versus informed reviews, they also consider a model with commitment to a standard in which the principal observes the agent's ability but not her type, and accordingly, sets different standards. Optimal standards are harsh (soft) for agents with low-cost (high-cost) effort, resembling our findings for the pure aligned manipulation model. The driving forces behind their findings and ours are different, however, as in our

[9] Boleslavsky and Cotton (2015) and Zapechelnyuk (2020) also consider grading and certification systems, respectively, that incetivize producers to provide high quality.

model, the agent has private information and effort is purely wasteful (it does not affect the probability that the agent's type is high).

Our paper contributes to a growing literature on research practices and economic incentives.[10] Di Tillio et al. (2017, 2021) study how scientists' persuasion bias affects the informativeness of experiments, explicitly considering the probabilistic structure of sample selection. In contrast, our analysis abstracts from the specific manner in which information is manipulated. Our framework is very different from standard Bayesian persuasion models (see Kamenica 2019 and Kamenica et al. 2021 for a survey and open issues, respectively): (i) in our model, there is asymmetric information, because the sender (the agent) knows her type; (ii) our model has moral hazard: the receiver (the decision maker) does not observe the signal distribution chosen by the sender; (iii) the sender is restricted to choose within a set of signal distributions, and it is costly to choose more favorable signal distributions; and (iv) the receiver affects the sender's incentives by committing to a standard. Regarding the last point, in a similar spirit to what we do, Tsakas et al. (2021) modify the standard Bayesian persuasion framework by making the receiver strategic. They show that, by committing to incur some cost (burn money) if she picks an action preferred by the sender, the receiver can incentivize the sender to choose a more informative signal.

Perez-Richet and Skreta (2022) analyze optimal testing design under information manipulation when the probability structure of the test is unrestricted. Closer to what we do, Whitmeyer (2019) studies a setup in which the receiver can benefit from committing to garbling the signal from the sender. His analysis imposes very little structure on the process generating the signals and the receiver's garbling. In contrast, in our analysis, a fixed family of distributions satisfying MLRP and the use of standards are assumed as exogenously given; furthermore, our results focus on how ex-ante optimal standards deviate from ex-post optimality and the factors driving such differences.

The distinctive feature of our model, in comparison to classical statistical problems, is the presence of moral hazard. Our mechanism design approach, however, is rather shaped by information asymmetry: in the optimal revelation mechanism, the menu offered by the decision-maker has features resembling a discrete-type version of Mussa and Rosen (1978) price discrimination model.[11] Our setup, however, is different, because the decision maker has aligned (opposite) interests with the high (low) type agent.[12]

[10] See, e.g., Henry and Ottaviani (2019), Di Tillio et al. (2017, 2021), Herresthal (2022), McClellan (2022), and references therein.

[11] For instance, the low type agent is indifferent between reporting her true type or lying, whereas the high type agent strictly prefers reporting her type. The work of Ederer et al. (2018) is related to ours as well. They analyze how "opaque" contracts help a principal incentivize an agent to exert balanced efforts between tasks. In contrast, in our paper, standards' distortions aim to discourage (encourage) the low (high) type agent to exert effort.

[12] Our paper also relates to the literature on optimal evidentiary legal standards to induce adequate behavior (see, e.g., Demougin and Fluet 2008; Ganuza et al. 2015; Gerlach 2013; Kaplow 2011; Sanchirico 2012). The informative role of evidentiary standards has received little attention in this literature. Stephenson (2008) analyzes the effect of standards on the research effort of agencies seeking court approvals, and Mungan and Samuel (2019) show that harsh standards deter crime when guilty agents mimic innocent ones. Their work, however, has no counterpart to our characterizations of harsh and soft standards.

Finally, there is a growing literature in computer science on strategic classification that relates to our work (see, e.g., Hu et al. 2019; Milli et al. 2019). In those papers, the principal publishes a deterministic decision rule (a classifier) which combines a set of features for assessing an agent. The classifier provides incentives to some agents to exert (undesirable) efforts in order to improve their features, just to meet the classification boundary. We contribute to this literature by introducing uncertainty over the result of the test. Random test results are more realistic and substantially change equilibrium behaviour.

2 The model

A manager (the principal or receiver) faces a binary decision: he decides whether to hire or reject a candidate (the agent or sender). The candidate's type is binary: she is either *fit* or *unfit*. The manager prefers to hire the candidate if she is fit and to reject the candidate if she is unfit. The candidate's fitness, however, is not observable to the manager. She is fit with a prior probability strictly between 0 and 1, and unfit otherwise. The prior unfitness odds, i.e., the prior probability that the candidate is unfit divided by the prior probability that the candidate is fit, are denoted by κ.

The manager is risk-neutral and minimizes expected losses. Without loss of generality, the manager's losses due to hiring unfit candidates and rejecting fit candidates are normalized to 1.[13] For $\kappa < (>)1$, if the manager were to make his decision based on prior information only, he would choose hiring (rejection). Throughout the paper we refer to κ simply as the *prior*. A useful interpretation for the reader to keep in mind is that κ corresponds to a measure of the manager's relative expected loss from hiring given the prior information.

2.1 Evidence

The manager runs a test to obtain further evidence on the candidate's fitness. The result of the test is the realization of a signal $z \in [0, 1]$. The distribution of the signal is determined by the candidate's *readiness* for the test, $\theta \in [\underline{\theta}, \overline{\theta}] =: \Theta$. The distribution and density functions of a candidate's signal with readiness θ are denoted by $F(\cdot, \theta)$ and $f(\cdot, \theta)$, respectively.[14] Thus, the domain of F is $D := [0, 1] \times \Theta$ and its interior is denoted by D°; similarly, the interior of Θ is denoted by Θ°. We assume that the distribution F is atomless and thrice continuously differentiable on D, being the third-order partial derivatives of F continuous real functions defined over D. Further, we assume that $f(z, \theta) > 0$ for all $(z, \theta) \in (0, 1) \times \Theta$ and that the density

[13] Our model also applies to situations in which the principal weights more heavily type I errors than type II errors, or vice versa. An increase in the relative weight of hiring the unfit candidate over the weight of rejecting the fit candidate has the same effect as an increase in κ.

[14] In principle, fit and unfit candidates could face different distributions. We impose, however, that the distributions are the same, F. The interpretation is that that both type of candidates face the same "test," but their readiness for it can be different, which is captured completely by θ.

is log-supermodular:

$$\frac{\partial^2 \ln f(z,\theta)}{\partial\theta\partial z} > 0 \tag{1}$$

for all $(z,\theta) \in D^\circ$. The log-supermodularity of the density function implies the strict Monotone Likelihood Ratio Property (MLRP): if $\theta' > \theta$, then $\frac{f(\cdot,\theta')}{f(\cdot,\theta)}$ is strictly increasing, which in turn implies strict first-order stochastic dominance (FOSD), $F(z,\theta) > F(z,\theta')$ for all $z \in (0,1)$.

2.2 Information manipulation

The candidate has a baseline preparedness for the test that we call her *natural* readiness, normalized to $\underline{\theta}$ if she is unfit, and given by some $\underline{\theta}_q > \underline{\theta}$ if she is fit. By exerting a costly effort, the candidate can increase her readiness. The "final" readiness of the unfit and fit candidates are denoted by θ_u and θ_q, respectively. Their respective manipulation *efforts* are $\theta_u - \underline{\theta}$ and $\theta_q - \underline{\theta}_q$, and the incurred costs are denoted by $C_u(\theta_u; \underline{\theta})$ and $C_q(\theta_q; \underline{\theta}_q)$, respectively. These cost functions are assumed to be thrice continuously differentiable, with both $C_u(\cdot; \underline{\theta})$ and $C_q(\cdot; \underline{\theta}_q)$ strictly increasing. Both the cost and marginal cost of making no effort are set equal to zero: $C_u(\underline{\theta}; \underline{\theta}) = C'_u(\underline{\theta}; \underline{\theta}) = 0$, and for all $\underline{\theta}_q \in \Theta^\circ$, $C_q(\underline{\theta}_q; \underline{\theta}_q) = C'_q(\underline{\theta}_q; \underline{\theta}_q) = 0$.[15] To avoid notation cluttering, we often omit the dependence of the cost function on the natural readiness, and simply write $C_i(\theta_i)$ (and similarly for its derivatives) for $i = u, q$.

Each candidate minimizes the sum of the expected loss from rejection and the cost of effort. The candidate's loss from rejection is normalized to 1. As discussed below, the manager sets a standard, $s \in [0,1]$, such that the candidate is accepted if and only if her signal realization is greater than the standard. Given any standard $s \in [0,1]$, candidate i's objective function is

$$U_i(s,\theta_i) := F(s,\theta_i) + C_i(\theta_i), \tag{2}$$

for $i = u, q$, respectively, for all $\theta_u \in \Theta$ and $\theta_q \in \Theta_q := [\underline{\theta}_q, \overline{\theta}]$. Given a standard s, the optimal readiness, denoted by $\theta_i^*(s)$, is a minimizer of $U_i(s,\cdot)$ for $i = u, q$.

In order to ensure a *unique* solution of the candidate's problem, guaranteeing the convexity of the candidate's loss function, we assume $C''_i > -\partial^2 F(s,\cdot)/\partial\theta^2$, for all $s \in (0,1)$ and $i = u, q$. A sufficient (but not necessary) condition for this requirement is that the cost functions are strictly convex ($C''_u > 0$ and $C''_q(\cdot; \underline{\theta}_q) > 0$) while F is convex in θ ($\partial^2 F(s,\theta)/\partial\theta^2 \geq 0$ for all $(s,\theta) \in D^\circ$).[16] Additionally, in order to ensure an *interior* solution of the candidate's problem (i.e., $\overline{\theta}$ is not an optimal choice for the candidate), we assume that $C'_i(\overline{\theta}) > -F_\theta(s,\overline{\theta})$ for all $s \in (0,1)$ and $i = u, q$.

15 The derivatives of functions of one variable are denoted by a prime or d/dx, where x is the variable.

16 The convexity of F in θ is not needed. For instance, any distribution defined by $f(z,\theta) = \gamma(\theta) f_1(z) + (1-\gamma(\theta)) f_0(z)$ for all $(z,\theta) \in [0,1]^2$, where $d(f_1/f_0)/dz > 0$ and $\gamma : \Theta \to [0,1]$ is strictly increasing and convex, with $0 < \gamma''(\theta) < C''_i(\theta)$ for all $\theta \in \Theta$, satisfies $C''_i > -\partial^2 F(s,\cdot)/\partial\theta^2$.

Finally, we impose that $C'_q(\theta; \underline{\theta}_q) < C'_u(\theta)$ for all $\theta \in \left(\underline{\theta}_q, \overline{\theta}\right)$, which is equivalent to the standard sorting condition.

2.3 Applications

Many applications discussed in the literature fit well within our model: pharmaceutical companies trying to get their drugs approved by the FDA (Li 2001), borrowers trying to get a loan approval (Ball 2020), or students trying to be admitted to a college (Cunningham and Moreno de Barreda 2019).

We use the college admission application to illustrate the different elements of the model. Here, the principal is the college and the agent is the student trying to get admitted to the college. The student is fit if she has academic aptitudes suitable for college education, and unfit otherwise. College applicants are asked to provide their SAT score. Since the test is administrated by the Educational Testing Service, each individual college can be assumed not to be involved in the design of the test. Our model captures this by the assumption that F is exogenously given and only the admission standard can be decided by each college. Finally, students decide the level of effort that they allocate to get ready for the test. Their score distribution is determined by both their natural or nurtured academic aptitudes, $\underline{\theta}$ for unfit students and $\underline{\theta}_q$ for fit ones, and their test preparation effort, which, altogether, results in their actual readiness, θ_u and θ_q, respectively.

2.4 Evidence standards

Throughout the paper, we only consider readiness pairs $\boldsymbol{\theta} := (\theta_u, \theta_q) \in \boldsymbol{\Theta} := \{\boldsymbol{\theta} \in \Theta^2 : \theta_u < \theta_q\}$, due to the sorting condition. We assume that the manager chooses a "threshold" strategy ($s \in [0, 1]$) such that the candidate is hired if her signal is greater or equal to the standard (i.e., $z \geq s$), and she is rejected otherwise.[17] For any standard s, and readiness profile (θ_u, θ_q), the probabilities of wrongful rejection and wrongful hiring are, respectively, $F(s, \theta_q)$ times the prior probability that the candidate is fit and $(1 - F(s, \theta_u))$ times the prior probability that the candidate is unfit. Thus, the manager's expected loss is an affine transformation of[18]

$$V(s, \boldsymbol{\theta}) = F(s, \theta_q) - \kappa F(s, \theta_u) \tag{3}$$

for all $(s, \boldsymbol{\theta}) \in [0, 1] \times \boldsymbol{\Theta}$. We will often explicitly indicate the dependence of V on the prior κ, writing $V(s, \boldsymbol{\theta}; \kappa)$ instead of $V(s, \boldsymbol{\theta})$.

[17] In the classical statistical setup, given $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the MLRP implies that the manager's optimal policy is to set an acceptance standard. On the other hand, in our model, $\boldsymbol{\theta}$ is not fixed, thus, in principle we could consider other policies, different from setting a standard. A plausible scenario, however, is that agents are capable to tailor downwards the result of the test (e.g., in a multiple choice test in which students can tell whether they know each answer, if there are two scores $s_2 > s_1$, where the former score results in fail and the latter in pass, a student who can score s_2 could deliberately give wrong answers to score s_1). This obligates the manager to define hiring and rejection sets using a standard, and it may explain why standards are widespread in practical decision-making.

[18] The manager's expected loss is $(V + \kappa)$ times the prior probability that the candidate is fit.

Define the *ex-post optimal standard* to be the function s^* mapping any given readiness profile and prior $(\boldsymbol{\theta}, \kappa)$ to the standard minimizing the manager's expected loss; that is, $s^*(\boldsymbol{\theta}, \kappa)$ is the minimizer of $V(\cdot, \boldsymbol{\theta}; \kappa)$. Also define the likelihood ratio function $g(s, \boldsymbol{\theta}) := f(s, \theta_q)/f(s, \theta_u)$ for all $(s, \boldsymbol{\theta}) \in [0, 1] \times \boldsymbol{\Theta}$. By the strict MLRP, $g(\cdot, \boldsymbol{\theta})$ is strictly increasing for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Thus, for the range of priors for which the problem is *test-worthy* (i.e., for which the optimal standard is interior, so that the candidate is not outright hired or outright rejected), given by $\kappa \in (g(0, \boldsymbol{\theta}), g(1, \boldsymbol{\theta}))$, the optimal standard is strictly increasing in the prior, as it satisfies $g(s^*(\boldsymbol{\theta}; \kappa), \boldsymbol{\theta}) \equiv \kappa$.

The problem solved by the manager facing a candidate with her natural readiness is called *classical statistical problem*. The manager's expected loss in this problem is $V(s, \underline{\theta}, \underline{\theta}_q) = F(s, \underline{\theta}_q) - \kappa F(s, \underline{\theta})$, for all $s \in [0, 1]$. The classical statistical problem is then test-worthy if $\kappa \in (\underline{\kappa}, \overline{\kappa})$ where $\underline{\kappa} := g(0, \underline{\theta}, \underline{\theta}_q)$ and $\overline{\kappa} := g(1, \underline{\theta}, \underline{\theta}_q)$. MLRP implies that $\underline{\kappa} < 1 < \overline{\kappa}$. Furthermore, the range of priors for which the classical statistical problem is test-worthy gets larger as $\underline{\theta}_q$ increases since $\underline{\kappa}$ $(\overline{\kappa})$ is weakly decreasing (weakly increasing) in $\underline{\theta}_q$ (this is a direct consequence of Remark 2 in the "Appendix"). In other words, the greater the difference between the natural readiness of the fit and unfit candidates, the more informative the test and the larger the range of priors for which the manager relies on this test.

Finally, before analyzing the role of commitment, we briefly consider an imperfect information *static game*, denoted by Γ_0, between the manager and the candidate, who simultaneously choose the standard and readiness, respectively. Their expected loss are given by (3) and (2), respectively. A *pure strategy Bayesian Nash equilibrium* (BNE) of Γ_0 is a triplet $(s^*_{NE}, \theta_{uNE}, \theta_{qNE}) \in D \times \Theta_q$, with $s^*_{NE} = s^*(\theta_{uNE}, \theta_{qNE}; \kappa)$, $\theta_{uNE} = \theta^*_u(s^*_{NE})$, and $\theta_{qNE} = \theta^*_q(s^*_{NE})$.

For all $\kappa \in \left(\inf_{s\in(0,1)} g(s, \theta^*_u(s), \theta^*_q(s)), \sup_{s\in(0,1)} g(s, \theta^*_u(s), \theta^*_q(s))\right)$, Γ_0 has at least one BNE with $s^*_{NE} \in (0, 1)$, and hence, satisfying that $g(s^*_{NE}, \theta^*_u(s^*_{NE}), \theta^*_q(s^*_{NE})) = \kappa$. These equilibria are not necessarily unique. It is not difficult, however, to provide conditions for uniqueness (for instance, see the proof of Corollary 1 below). On the other hand, $(0, \underline{\theta}, \underline{\theta}_q)$ is the unique BNE if $\kappa < \inf_{s\in(0,1)} g(s, \theta^*_u(s), \theta^*_q(s))$, and $(1, \underline{\theta}, \underline{\theta}_q)$ is the unique BNE if $\kappa > \sup_{s\in(0,1)} g(s, \theta^*_u(s), \theta^*_q(s))$.

3 Analysis of commitment to standards

If the manager could commit in advance to a standard and the candidate could manipulate her readiness (at a type-dependent cost), the equilibrium extent of manipulation and hence, the endogenous informativeness of the test, would depend on the announced standard. With the aim of addressing how the manager can benefit from commitment, we analyse a dynamic game, denoted by Γ, in which, (i) first, Nature chooses the candidate's type (fit or unfit) and reveals it to the candidate; (ii) the manager (without observing the candidate's type) commits to a standard $s \in [0, 1]$. Then, in stage (iii), having observed the standard chosen by the manager, the candidate chooses readiness θ. (iv) Nature chooses a signal realization $z \in [0, 1]$ according to $F(\cdot, \theta)$, and the candidate is hired if and only if $z \geq s$.

The candidate's and manager's expected losses are given by (2) and (3), respectively, and their preferences, including κ, $\underline{\theta}_q$, F, C_u, and C_q, are common knowledge. The unfit and fit candidates' sets of strategies, denoted by $\Theta^{[0,1]}$ and $\Theta_q^{[0,1]}$, respectively, are sets of functions mapping standards to readiness. Our analysis focuses on *Subgame Perfect Nash Equilibria* in pure strategies (SPNE) (strictly speaking, the relevant equilibrium concept is Perfect Bayesian Equilibrium, but we omit specifying beliefs as they are trivial: the manager's beliefs are the same as the prior, and the candidate observes the standard and her type). A SPNE always exists and we let $\kappa \rightrightarrows \mathcal{S}^*(\kappa)$ be the correspondence mapping $\kappa \in (0, \infty)$ to the set of standards in such equilibria (with $\mathcal{S}^{*-1}$ being the inverse). We typically have a unique equilibrium in Γ. Some of these games, however, have multiple equilibria for knife-edge values of the prior κ.

3.1 Strategic complementarity/substituibility of readiness

Since $F(0, \theta) = 0$ and $F(1, \theta) = 1$ for all $\theta \in \Theta$, $\theta_u^*(0) = \theta_u^*(1) = \underline{\theta}$ and $\theta_q^*(0) = \theta_q^*(1) = \underline{\theta}_q$. Furthermore, $\theta_i^*(s)$ is an interior solution and satisfies

$$C_i'(\theta_i^*(s)) = -F_\theta(s, \theta_i^*(s)), \tag{4}$$

for all $s \in (0, 1)$ and $i = q, u$. Thus, at the optimal readiness, the marginal cost is equal to the marginal return to readiness, the rate at which the probability of failing the test decreases with readiness for a given standard. Further, $\theta_q^* > \theta_u^*$ by the sorting condition.

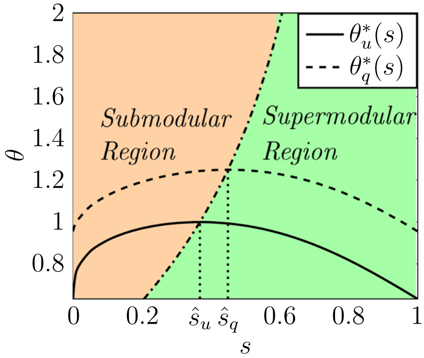
By the assumption $C_i'' > -\partial^2 F(s, \cdot)/\partial\theta^2$, the sign of the effect of the standard on the candidate's optimal readiness is determined by whether F is sub or supermodular at $(s, \theta_i^*(s))$ (i.e., by the sign of $\partial f(s, \theta_i^*(s))/\partial\theta$) since:

$$\frac{d\theta_i^*(s)}{ds} = -\frac{\partial f(s, \theta_i^*(s))}{\partial\theta}\left(C_i''(\theta_i^*(s)) + \frac{\partial^2 F(s, \theta_i^*(s))}{\partial\theta^2}\right)^{-1} \tag{5}$$

for all $s \in (0, 1)$ and $i = q, u$. Furthermore, there exists a continuous function, mapping readiness to the standard, that separates the submodular region of the domain of F from the supermodular region (see Remark 2 in the proof of Lemma 1 in the "Appendix"). As a result, the candidate's optimal readiness $\theta_i^*(s)$ is single-peaked: there exists a cut-off $\hat{s}_i$ for $i = u, q$, which we call the *modularity-switch point*, such that for standards smaller than $\hat{s}_i$, F is submodular at $(s, \theta_i^*(s))$ and hence, θ_i^* is strategic complement of the standard over $[0, \hat{s}_i)$. Analogously, θ_i^* is a strategic substitute of the standard over $(\hat{s}_i, 1]$ by the supermodularity of F at $(s, \theta_i^*(s))$ over $(\hat{s}_i, 1]$. We call this property *single modularity-switch*:

Lemma 1 *For $i = u, q$, there exists $\hat{s}_i \in (0, 1)$ such that*

$$\frac{d\theta_i^*(s)}{ds}\begin{cases} > 0 & \text{if } 0 \le s < \hat{s}_i \\ = 0 & \text{if } s = \hat{s}_i \\ < 0 & \text{if } \hat{s}_i < s \le 1. \end{cases} \tag{6}$$



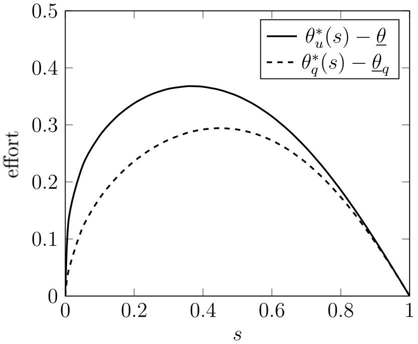


Fig. 1 The left panel shows the candidates' best response to the standard and the submodular and supermodular regions, in the setting of Example 1 with $\underline{\theta}_q = 1$. The right panel shows the corresponding optimal efforts

Example 1 Consider $F(z, \theta) = z^\theta$ for all $(z, \theta) \in [0, 1] \times \Theta$ with $\Theta = [1 - \mathrm{e}^{-1}, 2]$, and cost functions given by $C_u(\theta) = \frac{1}{2}(\theta - \underline{\theta})^2$ for all $\theta \in \Theta$ and $C_q(\theta) = \frac{1}{2}(\theta - \underline{\theta}_q)^2$ for all $\theta \in \Theta_q$. Since $\partial f(z, \theta)/\partial\theta = z^{\theta-1}(\theta \ln z + 1)$, equalizing this expression to 0, we find that the function separating the submodular and supermodular regions is given by $z = e^{-\frac{1}{\theta}}$. The left panel of Fig. 1 illustrates the equilibrium readiness θ_u^* and θ_q^*, the submodular and supermodular regions and the modularity switch-points. The right panel displays the corresponding efforts.

A distribution is said to have a *neutral signal s* (c.f., Milgrom 1981) if $f(s, \theta) = f(s, \theta')$ for all $\theta, \theta' \in \Theta$. If a distribution F has a neutral signal s, then F is submodular at (z, θ) if $z < s$, supermodular if $z > s$, and the modularity-switch point for both candidate types is the neutral signal.

Example 2 Consider $F(z, \theta) = \theta z^2 + (1 - \theta)z$ for all $z \in [0, 1]$, $\theta \in \Theta = [0, 1]$ and $\underline{\theta}_q \in \left(\frac{1}{4}, \frac{3}{4}\right)$. In the classical statistical problem, the likelihood ratio function is given by $g(s, \underline{\theta}, \underline{\theta}_q) = 1 + \underline{\theta}_q(2s - 1)$, which is strictly increasing in s, so the problem is test-worthy if $\kappa \in (1 - \underline{\theta}_q, 1 + \underline{\theta}_q)$. Since $\partial f(z, \theta)/\partial\theta = 2z - 1$, F has a neutral signal at $z = \frac{1}{2}$.

The unfit candidate's cost function is $C_u(\theta) = \frac{1}{2}\theta^2$ for all $\theta \in \Theta$, thus her optimal readiness is $\theta_u^*(s) = s(1 - s)$ for all $s \in [0, 1]$. Similarly, the fit candidate's cost function is $C_q(\theta) = \frac{1}{2}(\theta - \underline{\theta}_q)^2$ for all $\theta \in [\underline{\theta}_q, 1]$; thus her readiness is $\theta_q^*(s) = \underline{\theta}_q + s(1 - s)$ for all $s \in [0, 1]$. Hence, the optimal readiness of both candidate's types are strategic complements of the standard for $s < 1/2$ and strategic substitutes for $s > 1/2$.

3.2 Deviations from ex-post efficiency

The standard set by the manager in the BNE of the static game Γ_0, in which the candidate does not observe the standard before choosing her readiness, is *ex-post efficient*. Our focus, however, is on the dynamic game Γ in which the manager also takes into account the candidate's incentives to exert effort when setting the optimal standard. Addressing the effects of the standard on manipulation effort leads the manager to set *ex-post inefficient* standards.

In equilibrium, the manager's expected loss as a function of the standard is given by:

$$V(s, \theta_u^*(s), \theta_q^*(s)) = F(s, \theta_q^*(s)) - \kappa F(s, \theta_u^*(s)).$$

Since

$$\frac{dV(s, \theta_u^*(s), \theta_q^*(s))}{ds} = f(s, \theta_q^*(s)) + F_\theta(s, \theta_q^*(s)) \frac{d\theta_q^*(s)}{ds} \\ -\kappa \left(f(s, \theta_u^*(s)) + F_\theta(s, \theta_u^*(s)) \frac{d\theta_u^*(s)}{ds} \right), \quad (7)$$

for all $s \in (0, 1)$, the F.O.C. of the manager's problem simplifies to $v(s_P^*) = \kappa$ for all SPNE with $s_P^* \in (0, 1)$, where v is the *pseudo likelihood ratio* function, defined by

$$v(s) := \frac{f(s, \theta_q^*(s)) + F_\theta(s, \theta_q^*(s)) \frac{d\theta_q^*(s)}{ds}}{f(s, \theta_u^*(s)) + F_\theta(s, \theta_u^*(s)) \frac{d\theta_u^*(s)}{ds}}$$

or its limit, whenever this limit is well defined, for all $s \in [0, 1]$.

Since the manager wants to encourage manipulation by the fit type and discourage manipulation by the unfit type, addressing these indirect effects of the standard results in that, in general, equilibrium standards are not ex-post efficient. In terms of the F.O.C.'s of the manager problem, with and without commitment, $v(s) \neq g(s, \theta_u^*(s), \theta_q^*(s))$ for almost all $s \in (0, 1)$. We are interested in characterizing when the optimal standard under commitment is higher or lower than the ex-post optimal standard.

Definition 1 Let $(s_P^*, \theta_u^*, \theta_q^*)$ be a SPNE of Γ. The equilibrium standard is *soft (harsh)* if $s_P^* < (>)\, s^*\left(\theta_u^*(s_P^*), \theta_q^*(s_P^*); \kappa\right)$.

Upon observing the signal, at the margin, a harsh manager rejects a candidate even if the expected loss from hiring is strictly less than the expected loss from rejection. Similarly, at the margin, a soft manager hires a candidate even if the expected loss from rejection is strictly less than the expected loss from hiring.

Next, we show that the direction of the deviations from ex-post optimality critically depends on whether the candidate's effort is a strategic substitute or complement of the standard at the equilibrium, and on whether it is more relevant to deter effort from the unfit candidate or to encourage effort from the fit candidate.

4 Benchmark models

It is instructive to start the analysis considering *pure misaligned manipulation* and *pure aligned manipulation* models; i.e., games where only the unfit and fit candidate, respectively, are responsive to changes in the standard. An important feature of these cases is that $\mathcal{S}^*$ is *increasing* in the prior, κ. Specifically, we say that $\mathcal{S}^*$ is *weakly* (*strictly*) *increasing* within a given interval $I \subseteq (0, \infty)$ if $\kappa' > \kappa$, $s \in \mathcal{S}^*(\kappa)$ and $s' \in \mathcal{S}^*(\kappa')$ imply that $s' \geq s$ ($s' > s$), for all $\kappa, \kappa' \in I$. The monotonicity of $\mathcal{S}^*$ and the single modularity-switch property imply the existence of a unique threshold for the prior such that, the candidate's effort is a strategic complement of the standard in equilibrium for sufficiently low priors and a strategic substitute of the standard in equilibrium for sufficiently high priors.

As a consequence, as the prior increases, a soft-harsh pattern of equilibrium standards arises in pure misaligned manipulation games, while a harsh-soft pattern arises in pure aligned manipulation games. The intuition behind the different patterns is simple. On one hand, when the manager is mainly concerned with deterring effort from the unfit candidate, he optimally induces less effort from this candidate by setting a soft standard if the prior is sufficiently low (as the candidate's effort is a strategic complement of the standard for low priors), or alternatively, a harsh standard if the prior is sufficiently high (as the candidate's effort is a strategic substitute of the standard for high priors).

On the other hand, when the manager is mainly concerned with encouraging effort from the fit candidate, he optimally induces more effort from this candidate by setting a harsh standard for low priors (by the strategic complementarity role of the standard) and a soft standard for high priors (by its strategic substitutability).

4.1 Pure misaligned manipulation

A game in which the fit candidate's readiness is exogenously given is called a *pure misaligned manipulation game*. In this game, denoted by Γ_u, $\Theta_q = \{\underline{\theta}_q\}$ for some $\underline{\theta}_q \in \left(\theta_u^*(\hat{s}_u), \overline{\theta}\right]$, and the unfit candidate exerts effort to make her signal distribution more similar to the fit candidate's signal distribution. All other assumptions on Γ, laid out in Sects. 2–3, remain valid. Just as in the solution of the classical statistical problem, the equilibrium standard is increasing in the prior. Intuitively, the higher the prior probability that the candidate is unfit, the higher is the required realization of the signal to induce posterior beliefs that make the manager's expected payoff from hiring greater than the expected payoff from rejection.

Lemma 2 *In every game* Γ_u*,* $\mathcal{S}^*$ *is weakly increasing over* $(0, \infty)$ *and strictly increasing over* $\mathcal{S}^{*-1}(0, 1)$*. Furthermore,* $\mathcal{S}^*(\kappa) = \{0\}$ *if* $\kappa \in (0, \underline{\kappa}]$ *and* $\mathcal{S}^*(\kappa) = \{1\}$ *if* $\kappa \in [\overline{\kappa}, \infty)$.

Consider an intermediate value of the prior leading to an interior SPNE standard, $s_P^* \in (0, 1)$. From (5) and the single modularity-switch property, v and $g(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$ cross only once at the unfit candidate's modularity switch point $\hat{s}_u$. The left and right panel of Fig. 2 show v and $g(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$ from Example 2 for different values of $\underline{\theta}_q$.

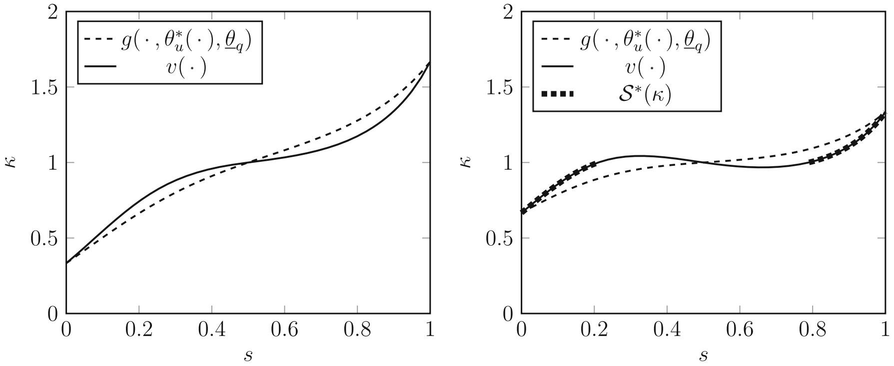


Fig. 2 The left and right panel show the pseudo likelihood ratio function v (solid line) and equilibrium standards for each κ in the static game Γ_0 (dashed line). The cost functions and distribution are those in Example 2, with $\underline{\theta}_q = \frac{2}{3}$ in the left panel and $\underline{\theta}_q = \frac{1}{3}$ in the right panel. In the left panel, the equilibrium standards of Γ_u coincide with v and, in the right panel, they correspond to the dash-dotted line

Since $g(\cdot, \theta_u^*(s_P^*), \underline{\theta}_q)$ is increasing, we have that $s_P^* < \hat{s}_u$ if and only if $v(s_P^*) > g(s_P^*, \theta_u^*(s_P^*), \underline{\theta}_q)$ and $s_P^* > \hat{s}_u$ if and only if $v(s_P^*) < g(s_P^*, \theta_u^*(s_P^*), \underline{\theta}_q)$.[19] Therefore, by the monotonicity of the equilibrium standard in the prior, the equilibrium standard is soft if the prior probability that the candidate is unfit is relatively low and harsh if it is relatively high.[20]

Proposition 1 *For any game Γ_u, there exists $\widetilde{\kappa}_u \in (\underline{\kappa}, \overline{\kappa})$ such that*

$$\text{the optimal standard is} \begin{cases} \textit{ex-post efficient} & \text{if } \kappa \in \left(0, \underline{\kappa}\right] \\ \textit{soft} & \text{if } \kappa \in (\underline{\kappa}, \widetilde{\kappa}_u) \\ \textit{harsh} & \text{if } \kappa \in (\widetilde{\kappa}_u, \overline{\kappa}) \\ \textit{ex-post efficient} & \text{if } \kappa \in [\overline{\kappa}, \infty). \end{cases} \tag{8}$$

Proposition 1 states that the manager optimally applies soft standards for relatively low priors and harsh standards for relatively high priors. This result follows from Lemmata 1-2: by the increasingness of the standard in the prior (Lemma 2), relatively low priors lead to relatively low standards in equilibrium. But, if the equilibrium standard is relatively low, then the equilibrium standard and unfit readiness pair $(s_P^*, \theta_u^*(s_P^*))$ is located in the submodular region of F, being the unfit type's effort a strategic complement of the standard (Lemma 1). Therefore, for low priors, the manager benefits

---

[19] If $dF(s, \theta_u^*(s))/ds \leq 0$ for some $s \in (0, 1)$, then from (7), we have $dV(s, \theta_u^*(s), \underline{\theta}_q)/ds > 0$, and, hence, s cannot be an equilibrium standard. All examples in the paper satisfy that $v(s) > 0$ for all $s \in (0, 1)$. There are, however, games such that $v(s) < 0$ within some intervals. For instance, in the game Γ_u, defined by $F(s, \theta) = \theta s^{10} + (1 - \theta)s$, $\underline{\theta}_q = 1$, and $C_u(\theta) = \frac{1}{2}\theta^2$ for all $s \in [0, 1]$ and $\theta \in \Theta = [0, 1]$, we have that $v(s) < 0$ for all $s \in (0.52, 0.65)$.

[20] On the other hand, the equilibrium standard associated with extreme values of the prior (i.e., if $\kappa \in (0, \underline{\kappa}] \cup [\overline{\kappa}, \infty)$) is ex-post efficient since the manager does not rely on the test outcome (the SPNE standard is a corner solution), as it cannot overturn extreme prior beliefs. The equilibrium standard is also ex-post efficient if the evidence from the test is valuable but a commitment to a standard is not; i.e., if $\kappa = \widetilde{\kappa}_u$ and $\mathcal{S}^*(\widetilde{\kappa}_u) = \{\hat{s}_u\}$. Example 2 illustrates that the manager may not choose ex-post efficient standards at $\widetilde{\kappa}_u$ when $\mathcal{S}^*(\widetilde{\kappa}_u)$ is multivalued.

from lowering the standard below the ex-post efficient level, because this discourages misaligned manipulation. An analogous argument reveals that the manager optimally induces less effort by the unfit candidate if he sets harsh standards for relatively high priors, due to the strategic substitutability role of the standard. If v is strictly increasing, then $\mathcal{S}^*$ is single-valued at all κ, and the cut-off prior for soft and harsh standards is equal to the likelihood ratio function evaluated at the modularity-switch point, $\widetilde{\kappa}_u = g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)$, as it is illustrated in the left-panel of Fig. 2. Since the unfit candidate becomes less responsive to changes in the standard as her cost of improving readiness increases, it is possible to provide sufficient conditions for relatively high costs C_u to generate a positive and monotone pseudo likelihood ratio function v (see Remark 3 in the “Appendix”).[21] In contrast, if the candidate’s effort is sufficiently responsive to changes in the standard (which is allowed by low manipulation costs), then v might be non-monotone and the equilibrium standard varies *discontinuously* with changes in the prior. This implies that candidates with arbitrarily similar priors may be subject to very different standards as it is illustrated in the right-panel of Fig. 2. These two observations are addressed formally by Remark 4 in the “Appendix”. On the other hand, $g(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$ is strictly increasing, as (5) and direct computations reveal.

Example 3 (Example 2 revisited) Consider $F(z, \theta) = \theta z^2 + (1-\theta)z$ for all $z \in [0, 1]$, $\theta \in \Theta = [0, 1]$ and $\underline{\theta}_q \in (1/4, 3/4)$. Thus, $\underline{\kappa} = 1 - \underline{\theta}_q$ and $\overline{\kappa} = 1 + \underline{\theta}_q$. Recall that $\hat{s}_u = \frac{1}{2}$. By Proposition 1, in the game Γ_u, the optimal standard is soft for $\kappa \in (1 - \underline{\theta}_q, \widetilde{\kappa}_u)$, harsh for $\kappa \in (\widetilde{\kappa}_u, 1 + \underline{\theta}_q)$, and ex-post efficient for $\kappa \leq 1 - \underline{\theta}_q$ and $\kappa \geq 1 + \underline{\theta}_q$.

For $\underline{\theta}_q \in (1/2, 3/4)$, direct computations reveal that $v'(s) > 0$ for all $s \in (0, 1)$. Therefore, by Remark 4, $\widetilde{\kappa}_u = g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q) = 1$ and the manager is ex-post efficient at $\widetilde{\kappa}_u$. For $\underline{\theta}_q \in (1/4, 1/2)$, v is non-monotone. In this case, $\widetilde{\kappa}_u = 1$ and $\mathcal{S}^*(\widetilde{\kappa}_u) = \{s_1, s_2\}$, where s_1 and s_2 are, respectively, the smallest and greatest root of $s(1-s) = \underline{\theta}_q/2$. The right-panel of Fig. 2 shows how the standard varies discontinuously with κ for $\underline{\theta}_q = 1/3$, with $s_1 = 0.21$ and $s_2 = 0.79$. The manager is not ex-post efficient at $\widetilde{\kappa}_u$—he is soft at $(0.21, \theta_u^*, 1/3)$ and harsh at $(0.79, \theta_u^*, 1/3)$.

4.2 Pure aligned manipulation

In a *pure aligned manipulation game*, denoted by Γ_q, only the fit candidate’s readiness is responsive to changes in the standard. Here, the unfit candidate is non-strategic: her readiness is exogenously given by her natural readiness; i.e., $\theta_u^*(s) = \underline{\theta}$ for all $s \in [0, 1]$. All other assumptions on Γ, laid out in Sects. 2–3, remain valid. Loosely speaking, Γ_q is the game obtained from any game Γ, as $C_u(\theta)$ goes to infinity for all $\theta > \underline{\theta}$. Let $\underline{\kappa}_q$ and $\overline{\kappa}_q$ be the smallest and largest priors, respectively, at which testing

[21] Similarly, it is easy to show that if v is monotone for a sufficiently high natural readiness of the fit candidate ($\underline{\theta}'_q$), then it is also monotone for all natural readiness that exceed such value (i.e., for all $\underline{\theta}_q > \underline{\theta}'_q$), by the log-supermodularity of F.

is worthy in the pure aligned manipulation setting.[22] Similarly to the pure misaligned manipulation scenario (Lemma 2), the equilibrium standard is monotone in the prior.

Lemma 3 *In game* Γ_q, $\mathcal{S}^*$ *is weakly increasing over* $(0,\infty)$ *and strictly increasing over* $(\underline{\kappa}_q, \overline{\kappa}_q)$. *Further,* $\mathcal{S}^*(\kappa) = \{0\}$ *(corresp.,* $\subset (0,1), = \{1\}$*) for all* $\kappa \in (0, \underline{\kappa}_q)$ *(corresp.,* $(\underline{\kappa}_q, \overline{\kappa}_q)$, $(\overline{\kappa}_q, \infty)$*).*

In turn, the monotonicity result and the single modularity-switch property allow us to connect the level of the prior with the strategic role of the standard in equilibrium.

Proposition 2 *For any game* Γ_q, $\underline{\kappa}_q < \widetilde{\kappa}_q < \overline{\kappa}_q$ *and*

$$\text{the optimal standard is} \begin{cases} \textit{ex-post efficient} & \text{if } \kappa \in \left(0, \underline{\kappa}_q\right) \\ \textit{harsh} & \text{if } \kappa \in (\underline{\kappa}_q, \widetilde{\kappa}_q) \\ \textit{ex-post efficient} & \text{if } \kappa = \widetilde{\kappa}_q \\ \textit{soft} & \text{if } \kappa \in (\widetilde{\kappa}_q, \overline{\kappa}_q) \\ \textit{ex-post efficient} & \text{if } \kappa \in (\overline{\kappa}_q, \infty), \end{cases} \tag{9}$$

where $\widetilde{\kappa}_q := g(\hat{s}_q, \underline{\theta}, \theta_q^*(\hat{s}_q))$.

According to Proposition 2, the manager biases the standard in order to encourage manipulation by the fit candidate: he chooses harsh standards for relatively low priors and soft standards for relatively high priors. The intuition behind the reversed equilibrium pattern (when compared to Proposition 1) is that the manager wants to incentivize effort by the fit candidate, instead of desincentivize it. Higher effort by the fit candidate increases the probability that this candidate is hired, which is the right decision for the manager.

In contrast to the pure misaligned manipulation setting, aligned manipulation enlarges the set of priors for which testing is worthy (since $\underline{\kappa}_q \leq \underline{\kappa}$ and $\overline{\kappa} \leq \overline{\kappa}_q$) and the optimal standard associated to the cut-off prior is always the modularity-switch point; that is, $\mathcal{S}^*(\widetilde{\kappa}_q) = \{\hat{s}_q\}$ (see the proof of Proposition 2). This allows us to provide a closed-form expression for $\widetilde{\kappa}_q$, corresponding to the likelihood ratio at the fit candidate's modularity-switch point, and to guarantee ex-post efficiency at the cut-off.

5 The general model

In the general set-up, both candidate types are strategic, but the intuition from the pure manipulation models is robust: the manager is still willing to give up ex-post efficiency to take advantage of the standard's incentive role.

In any SPNE $(s_P^*, \theta_u^*, \theta_q^*)$ such that $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the supermodular and submodular regions of F respectively, the unfit and fit candidates'

[22] Formally, $\underline{\kappa}_q := \inf_{s\in(0,1)} \left\{ \frac{F(s,\theta_q^*(s))}{F(s,\underline{\theta})} \right\}$ and $\overline{\kappa}_q := \sup_{s\in(0,1)} \left\{ \frac{1-F(s,\theta_q^*(s))}{1-F(s,\underline{\theta})} \right\}$ (see proof of Lemma 3 for details).

readiness are, respectively, strategic substitute and strategic complement of the standard. In such equilibria, there is no strategic tension, and the manager applies harsh standards in order to discourage effort of the unfit candidate and encourage effort of the fit candidate. This case arises when the function separating the submodular and supermodular regions of F is strictly increasing and $s_P^* \in (\hat{s}_u, \hat{s}_q)$ (see Remark 2 in the "Appendix" and Fig. 1). This function is increasing if and only if $f(s, \cdot)$ is strictly log-concave. For instance, the distribution in Example 1 satisfies this condition. Analogously, the manager applies soft standards in equilibria such that $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the submodular and supermodular regions of F, respectively. This case arises when $f(s, \cdot)$ is strictly log-convex.

However, if both $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in either the submodular region of F, or the supermodular region of F, then the manager faces a trade-off between discouraging effort by the unfit candidate and encouraging effort by the fit candidate (for instance, if F has a neutral signal). In this case, the relative magnitudes of the strategic effects of the standard on the optimal efforts exerted by the candidates will play a critical role. We define the *strategic ratio* as the ratio of the strategic (i.e., indirect) effects of the standard on the signal distributions of the fit candidate over that of the unfit candidate:

$$r(s) := F_\theta(s, \theta_q^*(s)) \frac{d\theta_q^*(s)}{ds} \left(F_\theta(s, \theta_u^*(s)) \frac{d\theta_u^*(s)}{ds} \right)^{-1}$$

for all $s \in (0, 1) \setminus \{\hat{s}_u\}$, $r(0) := \lim_{s \to 0} r(s)$, $r(\hat{s}_u) := \lim_{s \to \hat{s}_u} r(s)$ and $r(1) := \lim_{s \to 1} r(s)$, respectively, whenever these limits exist. Next result follows directly from (7):

Lemma 4 *For any SPNE $(s_P^*, \theta_u^*, \theta_q^*)$ of Γ with $s_P^* \in (0, 1)$:*

(i) if $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the submodular and supermodular regions of F, respectively, i.e., $s_P^* \in (\hat{s}_q, \hat{s}_u)$, then the optimal standard is soft;

(ii) if $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the supermodular and submodular regions of F, respectively, i.e., $s_P^* \in (\hat{s}_u, \hat{s}_q)$, then the optimal standard is harsh;

(iii) if both $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the submodular region, i.e., $s_P^* \in (0, \min\{\hat{s}_q, \hat{s}_u\})$, then the optimal standard is soft (harsh) if and only if $r(s_P^*) < (>) \kappa$; and

(iv) if both $(s_P^*, \theta_u^*(s_P^*))$ and $(s_P^*, \theta_q^*(s_P^*))$ are located in the supermodular region, i.e., $s_P^* \in (\max\{\hat{s}_q, \hat{s}_u\}, 1)$, then the optimal standard is soft (harsh) if and only if $r(s_P^*) > (<) \kappa$.

We discuss the intuition of parts (iii) and (iv) for the case in which $r(s_P^*) < \kappa$. If the strategic ratio is less than the prior, then the manager is relatively more concerned with the effect of his commitment on the effort exerted by the unfit candidate and we say that *misaligned manipulation is dominant at this prior*. Since the effort exerted by the unfit candidate is a strategic complement of the standard over the submodular region and a strategic substitute over the supermodular region, the manager optimally commits to soft and harsh standards, respectively. The intuition of the case $r(s_P^*) > \kappa$ is analogous, and in this case we say that *aligned manipulation is dominant at this prior.*

As an application, consider a distribution F, such that $F(z, \cdot)$ is an affine transformation of $-\theta$ for all $z \in [0, 1]$. The corresponding marginal return to readiness is independent of readiness itself and thus, F has a neutral signal. Furthermore, assume that candidates have quadratic costs with $C''_q = C''_u$, as in Example 2. From (5), the candidates' manipulation responds in the same manner to changes in the standard, which yields that the strategic ratio is equal to 1 for every standard. Since v is equal to 1 only at the neutral signal, by (iii) and (iv) in Lemma 4, the manager applies either harsh standards or ex-post efficient standards for all priors. On the other hand, multiple soft-harsh or harsh-soft cut-offs may arise whenever neither misaligned manipulation nor aligned manipulation is dominant at all priors.

In contrast to the pure manipulation models, the monotonicity of standard in the prior cannot be guaranteed without imposing further assumptions in the general model. However, intuition suggests that most qualitative aspects of the results obtained in the pure manipulation games are robust when both candidate types are allowed to manipulate readiness, but the manipulation of one of the types is much more responsive to changes in the standard than the other's. In the next subsections, we show this formally.

5.1 Dominating misaligned manipulation

The manager is dominantly concerned about the indirect effect of the standard on the effort exerted by the unfit candidate if, for instance, the fit candidate may benefit little from exerting effort. This is the case if returns to readiness are decreasing and the fit candidate has a high natural readiness.

Let $\Gamma(\underline{\theta}_q)$ be the game Γ defined by a triplet (F, C_q, C_u), with $\underline{\theta}_q \in \Theta^\circ$. In the sequel, we denote explicitly the dependence on $\underline{\theta}_q$ of the smallest and largest priors at which the problem is test-worthy, $\underline{\kappa}$ and $\overline{\kappa}$, by writing $\underline{\kappa}(\underline{\theta}_q)$ and $\overline{\kappa}(\underline{\theta}_q)$, respectively. The following result shows that the arguments leading to Proposition 1 still apply when the fit candidate's effort changes little in response to changes in the standard.

Proposition 3 *Assume $F_\theta(s, \overline{\theta}) = 0$ for all $s \in (0, 1)$. Then, there exists $\underline{\underline{\theta}}_q \in \Theta^\circ$ such that for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_q, \overline{\theta}\right)$, there exists $\widetilde{\kappa}_u(\underline{\theta}_q) \in (\underline{\kappa}(\underline{\theta}_q), \overline{\kappa}(\underline{\theta}_q))$ such that*

$$\textit{the optimal standard is} \begin{cases} \textit{ex-post efficient} & \text{if } \kappa \in \left(0, \underline{\kappa}(\underline{\theta}_q)\right] \\ \textit{soft} & \text{if } \kappa \in (\underline{\kappa}(\underline{\theta}_q), \widetilde{\kappa}_u(\underline{\theta}_q)) \\ \textit{harsh} & \text{if } \kappa \in (\widetilde{\kappa}_u(\underline{\theta}_q), \overline{\kappa}(\underline{\theta}_q)) \\ \textit{ex-post efficient} & \text{if } \kappa \in [\overline{\kappa}(\underline{\theta}_q), \infty). \end{cases} \tag{10}$$

Proposition 3 states that if the marginal benefit from exerting effort vanishes as the fit candidate's natural readiness approaches its largest value and the fit candidate's natural readiness is large, then, in the limit, the strategic ratio goes to 0, and the manager prioritises deterring effort from the unfit candidate by applying soft standards for relatively low priors and harsh standards for relatively high priors (by Lemmata 1-4). Example 4 illustrates Proposition 3.

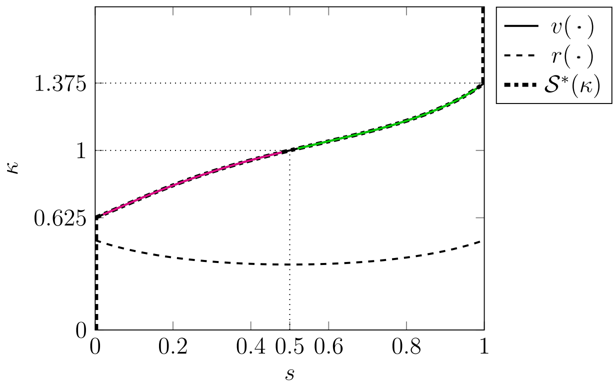


Fig. 3 Example 4. Strategic ratio $r(s)$ versus pseudo likelihood ratio function $v(s)$ and equilibrium standards $\mathcal{S}^*(\kappa)$ of Γ for $\underline{\theta}_q = 0.5$. The equilibrium standard is soft for $\kappa \in (0.625, 1)$ (highlighted in magenta), harsh for $\kappa \in (1, 1.375)$ (highlighted in green) and ex-post efficient otherwise

Example 4 Consider the distribution

$$F(z,\theta) = \theta\left(1-\frac{\theta}{2}\right)z^2 + \left(1-\theta\left(1-\frac{\theta}{2}\right)\right)z,$$

for all $(z,\theta) \in [0,1]^2$. The unfit candidate's cost function is $C_u(\theta) = \theta^2$ for all $\theta \in [0,1]$. The fit candidate's natural readiness is $\underline{\theta}_q \in (0,1)$, and her cost function is $C_q(\theta;\underline{\theta}_q) = \frac{1}{2}(\theta - \underline{\theta}_q)^2$ for all $\theta \in [\underline{\theta}_q, 1]$. Since F has a neutral signal at 0.5, we have that $\hat{s}_u = \hat{s}_q = 0.5$.

The best response functions of unfit and fit candidates are $\theta_u^*(s) = s(1-s)(s(1-s)+2)^{-1}$ and $\theta_q^*(s;\underline{\theta}_q) = (s(1-s)+\underline{\theta}_q)(s(1-s)+1)^{-1}$, respectively, for all $s \in [0,1]$. This yields an strategic ratio $r(s) = \left((1-\underline{\theta}_q)/2\right)^2 ((s(1-s)+2)/(s(1-s)+1))^3$. The strategic ratio $r(s)$ is decreasing (increasing) over (0, 0.5) (over (0.5, 1)), attaining its minimum at the neutral signal. Furthermore, as $\underline{\theta}_q$ increases, $r(s)$ shifts down, and $v(s)$ "rotates" around $(0.5, v(0.5)) = (0.5, 1)$, becoming "steeper" ($v(s)$ is decreasing (increasing) in $\underline{\theta}_q$ for all $s < (>)0.5$). For sufficiently high fit candidates' natural readiness—namely, for $\underline{\theta}_q \in (0.42, 1]$—misaligned manipulation is dominant (at $\underline{\theta}_q = 0.42$, $v(0) = r(0)$, whereas $r(s) < v(s)$ for all $s \in [0,1]$ for larger $\underline{\theta}_q$), and hence, a soft-harsh pattern arises as κ increases (see Fig. 3).

5.2 Dominating aligned manipulation

In this section, we consider the case in which the manager's strategic concerns are dominated by the effect of the standard on the fit candidate's effort. The main qualitative features of the pure aligned manipulation game Γ_q arise in game Γ as well, provided that the unfit candidate's effort is sufficiently costly, since this candidate's type becomes less responsive to changes in the standard as her cost of improving readiness increases.

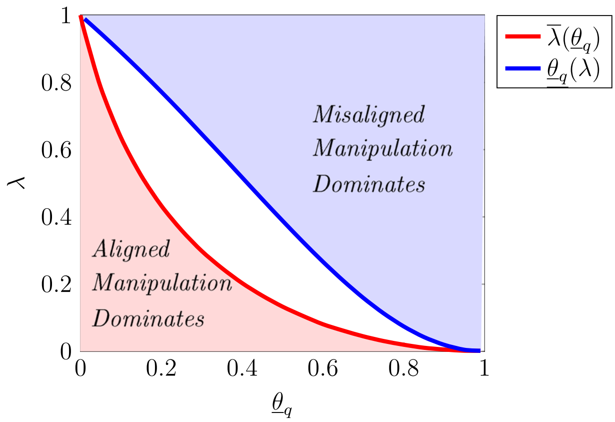


Fig. 4 Example 5. Aligned manipulation is dominant in the red shaded region while misaligned manipulation is dominant in the blue shaded region. The parameter $\overline{\lambda}(\underline{\theta}_q)$ equalizes the minimum of the difference between strategic ratio and the pseudo likelihood ratio $(r(s) - v(s))$ to 0, for each $\underline{\theta}_q \in (0, 1)$. The parameter $\underline{\theta}_q(\lambda)$ equalizes the maximum of the difference between strategic ratio and the pseudo likelihood ratio $(r(s) - v(s))$ to 0, for each $\lambda \in (0, 1)$

For any game Γ, defined by a triplet $(F, C_q(\cdot; \underline{\theta}_q), C_u)$, we define a set of games indexed by $\lambda \in [0, 1]$, and denoted by $\Gamma(\lambda)$, where the only difference between Γ and $\Gamma(\lambda)$ is that the cost function of the unfit candidate in the latter is $\lambda^{-1}C_u$, for all $\lambda \in (0, 1]$, whereas $\Gamma(0)$ corresponds to Γ_q. For an interpretation, λ may measure the unfit candidate's degree of access to a manipulation technology. As in the analysis of pure aligned manipulation games, we let $\underline{\kappa}_q(\lambda)$ and $\overline{\kappa}_q(\lambda)$ be the be the smallest and largest priors, respectively, at which testing is worthy in the game $\Gamma(\lambda)$.[23]

Proposition 4 *There exists* $\overline{\lambda} > 0$ *such that for all* $\lambda \in (0, \overline{\lambda})$*, there exists* $\widetilde{\kappa}_q(\lambda) \in (\underline{\kappa}_q(\lambda), \overline{\kappa}_q(\lambda))$ *such that*

$$\text{the optimal standard is} \begin{cases} \textit{ex-post efficient} & \text{if } \kappa \in \big(0, \underline{\kappa}_q(\lambda)\big) \\ \textit{harsh} & \text{if } \kappa \in (\underline{\kappa}_q(\lambda), \widetilde{\kappa}_q(\lambda)) \\ \textit{ex-post efficient} & \text{if } \kappa = \widetilde{\kappa}_q(\lambda) \\ \textit{soft} & \text{if } \kappa \in (\widetilde{\kappa}_q(\lambda), \overline{\kappa}_q(\lambda)) \\ \textit{ex-post efficient} & \text{if } \kappa \in (\overline{\kappa}_q(\lambda), \infty). \end{cases} \tag{11}$$

This result reveals that, if increasing readiness is sufficiently costly for the unfit candidate, the manager encourages effort of the fit candidates by applying harsh standards for relatively low priors and soft standards for relatively high priors (by Lemmata 1-4). For instance, the development of plagiarism checkers has helped decision makers (e.g., editors) ensure the originality of submitted manuscripts, by detecting misconduct. These innovations have transformed editors' decision problems, making them

[23] Formally, $\underline{\kappa}_q(\lambda) := \inf_{s\in(0,1)} \left\{ \frac{F(s,\theta_q^*(s))}{F(s,\theta_u^*(s;\lambda))} \right\}$ and $\overline{\kappa}_q(\lambda) := \sup_{s\in(0,1)} \left\{ \frac{1-F(s,\theta_q^*(s))}{1-F(s,\theta_u^*(s;\lambda))} \right\}$, where $\theta_u^*(\cdot; \lambda)$ is the best response of the unfit candidate with cost function $\lambda^{-1}C_u$ for all $\lambda \in (0, 1]$, and $\theta_u^*(\cdot; 0) = \underline{\theta}$.

to resemble more closely the aligned manipulation setup. Proposition 4 is illustrated in Example 5.

Example 5 We revisit Example 4. The fit candidate's cost function remains the same but the unfit candidate's cost function is now given by $C_u(\theta) = \frac{1}{2\lambda}\theta^2$ for all $\theta \in [0, 1]$ (e.g., in Example 4, $\lambda = 0.5$).

The best response functions of unfit and fit candidates are given by, respectively, $\theta_u^*(s; \lambda) = s(1-s)(s(1-s) + 1/\lambda)^{-1}$ and $\theta_q^*(s; \underline{\theta}_q) = (s(1-s) + \underline{\theta}_q)(s(1-s)+1)^{-1}$, for all $s \in [0, 1]$. The strategic ratio is $r(s) = \lambda^2(1 - \underline{\theta}_q)^2 \,((s(1-s) + 1/\lambda)/(s(1-s)+1))^3$.

For each possible fit candidate's natural readiness $\underline{\theta}_q$, there exist large cost functions for the unfit candidate—namely, for $\lambda \in (0, \overline{\lambda}(\underline{\theta}_q))$—such that $r(s) > v(s)$ for all $s \in [0, 1]$, and, thus, the manager's concerns about the fit candidates' effort dominate. Hence, a harsh-soft pattern arises as κ increases. For instance, for $\lambda = 1/2$ and $\underline{\theta}_q = 0.10$, the manager is harsh for $\kappa \in (0.89, 1)$, soft for $\kappa \in (1, 1.10)$, and ex-post efficient otherwise.

Figure 4 displays the threshold for $\overline{\lambda}$, as a function of $\underline{\theta}_q$, such that for all $\lambda < \overline{\lambda}$, aligned manipulation is dominant. Intuitively, the marginal benefit from fit candidates' effort is lower for larger values of $\underline{\theta}_q$; thus, in order to have dominant aligned manipulation, the marginal cost of the unfit candidate must be larger, or equivalently, λ must be lower.

Example 5 also illustrates Proposition 3. For each λ, as $\underline{\theta}_q$ increases, we move from aligned manipulation dominance to misaligned manipulation dominance. Larger gaps between candidates' readinesses, associated with larger $\underline{\theta}_q$, result in a less responsive fit candidate, making the effects of the standard on the readiness of the unfit candidate dominant.

6 Welfare analysis

In this section, we analyze the impact on welfare of the manager's ability to commit. We compare equilibrium standards and payoffs in the dynamic game Γ and the static game Γ_0—in which the standard and readiness are chosen simultaneously. In the previous sections, we found that when the candidate can manipulate the signal distribution, the ability to commit to a standard increases the expected payoff of the manager. This is achieved by either inducing a lower readiness by the unfit candidate or a greater readiness by the fit candidate, or both.

The candidate is strictly better-off with lower standards. Intuitively, a lower standard gives the candidate a higher probability of acceptance if she keeps her readiness unchanged; additionally she can adjust her readiness, which can only increase her payoffs even further.

Therefore, the equilibrium outcome of the dynamic game Pareto dominates the equilibrium outcome of the static game if and only if the standard in the former is lower than in the latter.

Remark 1 A SPNE of Γ Pareto dominates a BNE of Γ_0 if $s_P^* < s_{NE}^*$. If $s_P^* > s_{NE}^*$ the manager is weakly better off and the candidates are worse off in the SPNE.

Aligned manipulation often leads to non-unique BNE; however, as the return to effort vanishes, as $\underline{\theta}_q$ increases, uniqueness is recovered.[24] Our discussion in the rest of this section focuses on the case in which we have a unique BNE. We find conditions on prior beliefs and the type of manipulation that is dominant for which commitment leads to outcomes that Pareto dominate the outcome without commitment.

We assume that the fit candidate's natural readiness is relatively large. This guarantees that $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$, the likelihood ratio with adapted readiness, is increasing, which ensures that (i) the BNE of Γ_0 is unique, and (ii) the equilibrium standard under commitment is soft if and only if it is lower than the standard in the BNE.[25] To see why (ii) holds when $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$ is increasing, notice that when the standard under commitment, s_P^*, is soft, $g(s, \theta_u^*(s_P^*), \theta_q^*(s_P^*)) = \kappa$ for a standard $s > s_P^*$. Since $g(\cdot, \theta_u^*(s_P^*), \theta_q^*(s_P^*))$ is increasing, $g(s_P^*, \theta_u^*(s_P^*), \theta_q^*(s_P^*)) < \kappa$. On the other hand, the standard in the equilibrium of the static game, s_{NE}^* is ex-post efficient; i.e., it satisfies $g(s_{NE}^*, \theta_u^*(s_{NE}^*), \theta_q^*(s_{NE}^*)) = \kappa$. Thus, if $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$ is increasing, $s_{NE}^* > s_P^*$ as well. An analogous argument proves the converse.

Our next result (Corollary 1) considers the case in which the effect of standards on the unfit candidate's effort dominates, whereas the second one (Corollary 2) considers the case in which the effect of the standard on the fit candidate's effort dominates.

Corollary 1 *Assume that* $F_\theta(z, \overline{\theta}) = 0$ *for all* $z \in (0, 1)$ *and (1) also holds at* $(0, \theta)$ *and* $(1, \theta)$ *for all* $\theta \in \Theta^\circ$. *Then, for all large enough fit candidate's natural readiness* $\underline{\theta}_q$ *and relatively low priors for which the problem is test-worthy, the (unique) BNE of* $\Gamma_0(\underline{\theta}_q)$ *is Pareto dominated by every SPNE of* $\Gamma(\underline{\theta}_q)$.

This result is a consequence of Proposition 3. By Propositions 1 and 3, when misaligned manipulation is dominant, the manager applies soft standards in the dynamic game when there is a relatively low prior probability that the candidate is unfit.

In contrast, when aligned manipulation is dominant, the manager sets soft standards when there is a relatively high prior probability that the candidate is unfit (by Propositions 2 and 4). Let $\Gamma_0(\underline{\theta}_q, \lambda)$ and $\Gamma(\underline{\theta}_q, \lambda)$ be the static and dynamic games, respectively, with fit candidate's natural readiness $\underline{\theta}_q \in (\underline{\theta}, \overline{\theta}]$ and unfit candidate's cost function $\lambda^{-1} C_u$, with $\lambda \in [0, 1]$ (where $\lambda = 0$ corresponds to the pure aligned manipulation case). Next result is essentially a consequence of Proposition 4.

Corollary 2 *Assume that* $F_\theta(z, \overline{\theta}) = 0$ *for all* $z \in (0, 1)$ *and (1) also holds at* $(0, \theta)$ *and* $(1, \theta)$ *for all* $\theta \in \Theta^\circ$. *Then, for all combination of large enough fit candidate's*

[24] The source and nature of the non-uniqueness of the SPNE of Γ and the BNE of Γ_0 are very different. The SPNE of Γ is generically unique. Non-uniqueness of the SPNE only can arise because, for specific values of κ, the objective function of the manager, $V(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$, has multiple minimisers. This can only occur when v is non-monotone. The non-uniqueness of the BNE, on the other hand, arises when $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$ is non-monotone. In this case, we can have intervals of values of κ for which we have multiple BNEs.

[25] Our focus on relatively large natural readiness of the fit candidate does not seem to be too restrictive. For instance, in Example 4, the possible range for $\underline{\theta}_q$ is (0, 1] and $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot))$ is increasing for all $\underline{\theta}_q > 0.16$.

natural readiness $\underline{\theta}_q$ with large enough unfit candidate's manipulation costs (i.e., small enough λ), and relatively high priors for which the problem is test-worthy, the (unique) BNE of $\Gamma_0(\underline{\theta}_q)$ is Pareto dominated by every SPNE of $\Gamma(\underline{\theta}_q)$.

7 A mechanism design approach

Announcing a standard (either in hiring or internal promotion processes) is nowadays a ubiquitous response to the pre-contractual informational asymmetries. In this section, however, we show that the manager could do better by offering the candidate a menu of testing options to choose from; that is, by designing a revelation mechanism. Indeed, applying such a mechanism can yield a Pareto improving outcome when there is a high likelihood that the type of the candidate is unfit.

By the Revelation Principle, we can focus on direct mechanisms that are truthful (i.e., that, in equilibrium, induce the candidate to reveal her true type). We restrict attention to mechanisms without monetary transfers. Thus, the mechanisms that we consider are described by a decision rule mapping each report (*unfit* or *fit*) to probabilities of outright rejection, outright hiring, and using a test with approval standard s to make the decision.

In the "Appendix" we show that, without loss of generality, the analysis can be restricted to the class of mechanisms in which: (i) any candidate who claims to be unfit is outright rejected with probability $p \in [0, 1]$ and hired otherwise, and (ii) any candidate who claims to be fit is asked to take a test. Thus, we only consider truthful revelation mechanisms characterized by a duplet $(s, p) \in [0, 1]^2$, where s is the standard applied to a candidate reporting to be *fit*, and p is the probability of outright rejection for a candidate reporting to be *unfit*.

The individual rationality constraint for the unfit candidate is redundant: rejecting the contract yields a loss of $1 \geq p$ for all $p \in [0, 1]$. The same applies to the fit candidate: $1 \geq F(s, \underline{\theta}_q) \geq F(s, \theta_q^*(s)) + C_q(\theta_q^*(s))$ for all $s \in [0, 1]$.

Incentive-compatibility requires $p \leq F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))$ and $F(s, \theta_q^*(s)) + C_q(\theta_q^*(s)) \leq p$ for the unfit and fit candidate, respectively. The first restriction is binding, as the expected loss to the manager is decreasing in p, whereas the second is not. Therefore, incentive compatibility for the unfit candidate implies that in any mechanism (s, p), we have $p = F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))$ and the payoff to the manager is $V_M(s) := F(s, \theta_q^*(s)) - \kappa \left(F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))\right)$. Thus, if (s_M, p_M) is an optimal mechanism for the manager, then s_M solves $\min_{s\in[0,1]} V_M(s)$ and $p_M = F(s_M, \theta_u^*(s_M)) + C_u(\theta_u^*(s_M))$.

Let $\underline{\kappa}_M$ be the minimum of the ratio of the probability of rejecting the fit candidate over the probability of rejecting the unfit candidate across different standards; and let $\overline{\kappa}_M$ be the maximum of the ratio of the probability of accepting the fit candidate over the probability of accepting the unfit candidate across different standards.[26] We also

[26] Formally, $\underline{\kappa}_M := \inf_{s\in(0,1)} \left\{ \frac{F(s,\theta_q^*(s))}{F(s,\theta_u^*(s))+C_u(\theta_u^*(s))} \right\}$ and $\overline{\kappa}_M := \sup_{s\in(0,1)} \left\{ \frac{1-F(s,\theta_q^*(s))}{1-F(s,\theta_u^*(s))-C_u(\theta_u^*(s))} \right\}$.

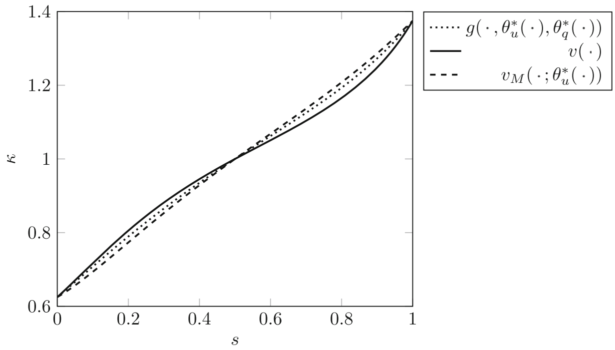


Fig. 5 Example from Sect. 5 with $\lambda = \underline{\theta}_q = 0.5$: equilibrium standard in Γ (solid line) and the optimal mechanism standard (dashed line), for each prior κ

define the pseudo likelihood ratio function relevant to this problem,

$$v_M(s) := \frac{f(s, \theta_q^*(s)) + F_\theta(s, \theta_q^*(s)) \frac{d\theta_q^*(s)}{ds}}{f(s, \theta_u^*(s))} \tag{12}$$

for all $s \in [0, 1]$. Finally, we let $\mathcal{S}_M^*$ be the correspondence mapping $\kappa \in (0, \infty)$ to the set of standards applied to candidates reporting to be fit in an optimal mechanism and adopt the definition of weakly increasing correspondence introduced in Sect. 4.[27]

Proposition 5 *The correspondence* $\mathcal{S}_M^*$ *is weakly increasing over* $(0, \infty)$. *If* (s_M, p_M) *is an optimal mechanism, then, (i)* $(s_M, p_M) = (0, 0)$ *for all* $\kappa < \underline{\kappa}_M$, $(s_M, p_M) = (s, F(s, \theta_u^*(s)) + C_u(\theta_u^*(s)))$, *for some* $s \in (0, 1)$ *satisfying* $v_M(s) = \kappa$, *for all* $\kappa \in (\underline{\kappa}_M, \overline{\kappa}_M)$, *and* $(s_M, p_M) = (1, 1)$ *for all* $\kappa > \overline{\kappa}_M$; *and (ii) the manager strictly prefers the optimal mechanism to the equilibria of* Γ *for all* $\kappa \in (\underline{\kappa}_M, \overline{\kappa}_M)$.

The manager is better-off using the optimal mechanism than simply committing to a standard, for all priors leading to an interior equilibrium standard, due to the higher probability of rejecting the unfit candidate. The proof of Proposition 5 reveals that, as in the pure aligned manipulation scenario of the game in which the manager commits to a standard, the optimal mechanism enlarges the range of priors for which screening is worthy.

Candidates are better-off with the revelation mechanism than under simple commitment to a standard if $s_M < s_P^*$ and worse-off if $s_M > s_P^*$. Provided that the pseudo likelihood ratio function of Γ, v, and v_M are both strictly increasing, $s_M < s_P^*$ if $v_M(s_P^*) > v(s_P^*)$, for all $\kappa \in \mathcal{S}_M^{*-1}(0, 1)$. Similarly, $s_M > s_P^*$ if $v_M(s_P^*) < v(s_P^*)$.

Functions v and v_M differ in the same manner that v and $g(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$ differ in the pure misaligned manipulation setup: by the presence of the indirect effect of the standard, $F_\theta(\cdot, \theta_u^*(\cdot)) \frac{d\theta_u^*(\cdot)}{ds}$, that appears in the denominator of v. Thus, if the problem is sufficiently well-behaved, so that both functions are well-defined in $[0, 1]$,[28] for all

[27] As in the game in which the manager simply commits to a hiring standard (see Sect. 4.1), the possibility of multiple equilibria for some knife-edge values of κ, in general, cannot be ruled out.

[28] See the discussion in footnote 19.

$s \in (0, 1)$ we have $v(s) > (=, <) v_M(s)$ if $s < (=, >) \hat{s}_u$. In this scenario, an argument analogous to the proof of Proposition 1 shows that standards in the commitment setup are lower than in the optimal mechanism for low priors and higher for high priors. This is illustrated in Fig. 5 for the example described in Sect. 5. Under the assumptions of Proposition 6, if F has a neutral signal and misaligned (aligned) manipulation dominates in Γ, then, the manager's deviations from the standard in the static game are in opposite (the same) direction under commitment and the optimal mechanism. When they are in the same direction, the deviation under the optimal mechanism is larger than under commitment, because of the buffering effect of the unfit candidate's effort in the latter.

The economics behind this, however, is very different: with the optimal mechanism, at the margin, the manager ignores the effect of the standard on the unfit candidate's effort because, since $p_M = F(s_M, \theta_u^*(s_M)) + C_u(\theta_u^*(s_M))$, the optimal menu offsets changes in $F(\cdot, \theta_u^*(\cdot))$ with changes in $C_u(\theta_u^*(\cdot))$ (which, at the margin, are the same). Since no test is needed when the unfit type is revealed, the manager is only concerned with incentivising the fit candidate's effort when setting up the standard in the optimal mechanism. Thus, the standard of the optimal mechanism and under commitment with pure misaligned manipulation deviate in opposite directions with respect to the standard without commitment.

Our last result shows that a sufficient condition for both v and v_M to be well-defined and strictly increasing is that the fit candidate's natural readiness and the unfit candidate's marginal cost are large. Under these assumptions, the candidates' ranking over the two settings (commitment and the revelation mechanism) depends on the prior. In particular, the revelation mechanism is Pareto superior for sufficiently high priors that the candidate's type is low.

Recall that, given a triplet (F, C_u, C_q), $\Gamma(\underline{\theta}_q, \lambda)$ is the game in which the manager commits to a standard, the fit candidate has a natural readiness $\underline{\theta}_q$, and the unfit candidate has a cost function $\lambda^{-1} C_u$ if $\lambda \in (0, 1]$, or the game Γ_q if $\lambda = 0$. Also, let $\hat{s}_u(\lambda)$ be the modularity-switch point of the unfit candidate with cost function $\lambda^{-1} C_u$ for all $\lambda \in (0, 1]$, and $\underline{\kappa}_M(\underline{\theta}_q, \lambda)$ and $\overline{\kappa}_M(\underline{\theta}_q, \lambda)$, defined as $\underline{\kappa}_M$ and $\overline{\kappa}_M$, respectively, but with $\theta_q^*(\cdot; \underline{\theta}_q)$ and $\theta_u^*(\cdot; \lambda)$ determined by $\underline{\theta}_q \in \Theta^\circ$ and $\lambda \in [0, 1]$, respectively.

Proposition 6 *Assume that* $F_\theta(z, \overline{\theta}) = 0$ *for all* $z \in (0, 1)$ *and that condition (1) also holds at* $(0, \theta)$ *and* $(1, \theta)$ *for all* $\theta \in \Theta^\circ$*. Then, for all combination of large enough fit candidate's natural readiness* $\underline{\theta}_q$ *with large enough unfit candidate's manipulation costs (i.e., small enough* λ*),*

(i) both candidate's types prefer commitment to a standard over the optimal mechanism, for the relatively low priors for which the mechanism is worthy ($\kappa \in (\underline{\kappa}_M(\underline{\theta}_q, \lambda), v(\hat{s}_u(\lambda)))$*), and*

(ii) both candidate's types prefer the optimal mechanism to commitment, and hence the optimal mechanism Pareto Dominates commitment, for the relatively high priors for which the mechanism is worthy ($\kappa \in (v(\hat{s}_u(\lambda)), \overline{\kappa}_M(\underline{\theta}_q, \lambda))$*).*

A large enough fit candidate's natural readiness $\underline{\theta}_q$ and unfit candidate's manipulation costs guarantee the monotonicity of the standard in the prior in both settings. Both candidate's types prefer the scheme with the lowest standard. Compared with commitment, under the optimal revelation mechanism the manager sets higher standards for sufficiently small priors that the candidate's type is low and lower standards for sufficiently high priors that the candidate's type is low.

The logic behind Proposition 6 is independent of which information manipulation effect dominates in the commitment setting. Furthermore, if the distribution has a neutral signal, the prior at which the candidate preferences for the revelation mechanism and commitment reverse is $\kappa = 1$.

8 Discussion

Information manipulation by interested parties is ubiquitous. Motivated by the widespread use of approval standards in applications, we analyse the desirability and implications of the use of commitment to ex-post inefficient standards as a tool to manage information manipulation. Optimal standards trade off classic statistical decision-making for management of information manipulation. Strategic complementarity between readiness and the standard develops in the submodular region of the domain of the signal distribution—i.e., for low standards that arise in equilibrium when agents have good prior prospects. Analogously strategic substitutability arises for agents with bad prior prospects.

An application often discussed in the literature is the drug approval process by regulatory agencies such as the FDA or the ABPI (see, e.g., Li 2001; Henry and Ottaviani 2019). As Li (2001) observes, "most of the evidence concerning effectiveness of a new drug is provided by its producer, not by the panelists." Pharmaceutical companies engage in a range of information manipulation practices, including hiding data, cherry-picking variables, manipulating experimental conditions, etc. (see, e.g., Goldacre 2014). In the light of our results, a question that arises is whether regulatory agencies' approval standards are tilted in the right direction to manage information manipulation incentives.

For instance, our model predicts that, when misaligned manipulation is dominant, drugs with good prospects (low κ) should be subject to soft approval standards; namely, drugs with ex-post evidence marginally-negative expected values should be approved. A natural choice for drugs having good prospects are those in the Breakthrough-Drug Designation (BDD) program of the FDA. As Darrow et al. (2018) show, trials following the nomination of many of the drugs in the BDD program have confirmed their good prospects, producing good results, and have been approved by the FDA. Nevertheless, trial results for some of the drugs in the program have shown little efficacy, even failing to meet customary standards (see Darrow et al. 2018, p. 1449). Yet, a number of such drugs have been approved by the FDA. Propositions 1 and 3 suggest that softening the standard has a positive side-effect of discouraging misaligned manipulation.[29] An empirical study of these issues is a subject of interest for future research.

[29] The BDD scheme was conceived to provide a "fast-track" approval process. The softening of standards that we refer to, however, is not related to the "fast-track" aspect of the program, but exclusively to the evidence documented in Darrow et al. (2018) on the approval of drugs that showed little efficacy in trials run after the drug was granted the designation.

As psychological screening has become widespread,[30] practitioners have emphasized the importance of assessing it properly (see, e.g., Dattner 2013; Caska 2020). The assessment of screening tests not only needs to take into account their performance in terms of wrong hiring/rejection, but also, their effect on test-preparation incentives. While the performance of selection procedures is determined by many factors, practitioners should bear in mind the insights of our analysis: managing information manipulation can further benefit from other incentive schemes in the economics toolkit.

Our analysis shows that a revelation mechanism allows the decision maker to obtain a higher expected payoff than simple commitment to a standard and it may be Pareto improving. Thus, protocols more involved than plain tests, as those described in our mechanism design approach, may be advantageous for managers' hiring procedures. Other possible approaches include mechanisms with transfers, manager's randomizations (probabilities of outright hiring/rejection), and hiring/rejection sets that are not monotone (i.e., not determined by a single standard) for both the commitment setup and revelation mechanisms. Additionally, the decision maker could introduce randomizations where the probability of hiring/rejection is conditional on the signal realization. We leave the analysis of these variations of the problem for future research.

Finally, we make a number of assumptions that aid the tractability of the analysis. For instance, the agent's type distribution and the principal's choice set are binary. It is not difficult to imagine situations where agents' heterogeneity may play an important role. Future research extending the analysis to continuous types is encouraged.

Acknowledgements We thank participants at ACE 2015 (QUT), AETW 2018 (ANU), ESWM 2018 (University of Naples Federico II), ESAM 2019 (Curtin University), QuETE 2019 (QUT), IWET 2022 (UCSC), LATAM 2022 (U. Chile), University of New South Wales, The University of Queensland, Universidad de Chile, the joint Series of Universidad del Bío-bío and Universidad de Concepción, and Universidad Católica de la Santísima Concepción for their comments. In particular, we thank Juan Carlos Carbajal, Eduardo Engel, Heiko Gerlach, Gabriele Gratton, Jonas Hedlund, Richard Holden, Keiichi Kawai, Takashi Kunimoto, Li Hao, Hongyi Li, Thomas Mariotti, Andrew McLennan, Friederike Mengel, Claudio Mezzetti, Marco Ottaviani, Lisette Pregelj, Francisco Silva, Miguel Sánchez Villalba, Joel Sobel, Roland Strausz, Alejandro Vidal, and Rodrigo Wagner for genuinely fruitful discussions. James Butcher provided us references on test-defensiveness. Kate Huang provided great research assistance. Financial support from the Spanish Ministry of Economics and Competitiveness (ECO 2012-34928) is gratefully acknowledged.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

[30] See SHL 2018 Global Assessment Trends Report https://www.shl.com/en/assessments/trends/.

Appendix: Proofs and Ancillary material

Proofs and Ancillary material of Sect. 2

Formal analysis of ex-post optimal standard

Since sign $\{\partial V(s,\boldsymbol{\theta})/\partial s\} = \text{sign}\{g(s,\boldsymbol{\theta}) - \kappa\}$ for all $(s,\boldsymbol{\theta}) \in (0,1) \times \boldsymbol{\Theta}$, we have that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the optimal standard is:

$$s^*(\boldsymbol{\theta};\kappa) = \begin{cases} 0 & \text{if } 0 < \kappa \leq g(0,\boldsymbol{\theta}) \\ s^*_{\boldsymbol{\theta},\kappa} & \text{if } g(0,\boldsymbol{\theta}) < \kappa < g(1,\boldsymbol{\theta}) \\ 1 & \text{if } g(1,\boldsymbol{\theta}) \leq \kappa, \end{cases} \tag{13}$$

where $s^*_{\boldsymbol{\theta},\kappa}$ is defined by $g(s^*_{\boldsymbol{\theta},\kappa},\boldsymbol{\theta}) \equiv \kappa$ for all $\kappa \in (g(0,\boldsymbol{\theta}), g(1,\boldsymbol{\theta}))$. Since $g(\cdot,\boldsymbol{\theta})$ is strictly increasing by the MLRP, $s^*(\boldsymbol{\theta};\cdot)$ is weakly increasing for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Proofs and Ancillary material for Sect. 3

Let $m(z,\theta) := \frac{1}{f(z,\theta)}\frac{\partial f(z,\theta)}{\partial\theta}$ for all $(z,\theta) \in D$. Note that $m(\cdot,\theta)$ is strictly increasing for all $\theta < \overline{\theta}$ since for all $z \in (0,1)$ and $\theta < \overline{\theta}$, we have that $\partial m(z,\theta)/\partial z = \partial^2 \ln f(z,\theta)/\partial\theta\partial z > 0$.

Claim 1 $F_\theta(z,\theta) < 0$ for all $z \in (0,1)$ and $\theta < \overline{\theta}$.

Proof For all $z \in (0,1)$ and $\theta < \overline{\theta}$,

$$F_\theta(z,\theta) = \int_0^z \frac{\partial f(z',\theta)}{\partial\theta}dz' = \int_0^z f(z',\theta)m(z',\theta)dz' \tag{14}$$

By the MLRP, we know that $F_\theta(z,\theta) \leq 0$ for all $z \in (0,1)$ and $\theta < \overline{\theta}$. We now show that this inequality is indeed strict. By the positiveness of the density function and strict monotonicity of $m(\cdot,\theta)$, if (14) is equal to zero for some $z \in (0,1)$ and $\theta < \overline{\theta}$, then $f(z',\theta)m(z',\theta) > 0$ and $F_\theta(z',\theta) > 0$ for all $z' \in (z,1)$. But $F_\theta(z',\theta) > 0$ contradicts FOSD (and hence the MLRP). □

Proof of Lemma 1. The proof of Lemma 1 relies on the fact that we can separate the submodular regions of the domain of F from the supermodular regions:

Remark 2 For all $\theta < \overline{\theta}$ there exists $\widetilde{s}(\theta) \in (0,1)$ such that

$$\frac{\partial f(z,\theta)}{\partial\theta} \begin{cases} < 0 & \text{if } z < \widetilde{s}(\theta) \\ = 0 & \text{if } z = \widetilde{s}(\theta) \\ > 0 & \text{if } z > \widetilde{s}(\theta) \end{cases} \tag{15}$$

for all $z \in [0,1]$.[31]

[31] We allow for $F_\theta(z,\overline{\theta}) = 0$ for all $z \in [0,1]$, thus, it is possible that $\frac{\partial f(z,\overline{\theta})}{\partial\theta} = 0$ for all $z \in [0,1]$.

Proof By Claim 1, for all $z \in (0,1)$ and $\theta < \overline{\theta}$, we have that $\int_0^z (\partial F_\theta(z', \theta)/\partial z) dz' = F_\theta(z, \theta) < 0$, where the equality follows from the fact that $F_\theta(0, \theta) = 0$ for all $\theta < \overline{\theta}$. Thus there is $z' \in (0, z)$ such that $\partial F_\theta(z', \theta)/\partial z < 0$; and similarly, there is $z'' \in (z, 1)$ such that $\partial F_\theta(z'', \theta)/\partial z > 0$ since $F_\theta(1, \theta) = 0$ for all $\theta < \overline{\theta}$. Therefore, $m(z', \theta) < 0$ and $m(z'', \theta) > 0$, and by continuity of m, there is $z''' \in (z', z'')$ such that $m(z''', \theta) = 0$. Indeed, z''' is the unique root of $m(\cdot, \theta) = 0$ because of the strict monotonicity of $m(\cdot, \theta)$ for all $\theta < \overline{\theta}$. Thus, $\theta \mapsto \widetilde{s}(\theta)$ maps θ to the unique root of $m(\cdot, \theta) = 0$, for all $\theta < \overline{\theta}$. □

Now, by the Implicit Function Theorem, $\widetilde{s}$ is continuous, with

$$\frac{d\widetilde{s}(\theta)}{d\theta} = -\frac{\partial m(\widetilde{s}(\theta), \theta)}{\partial \theta} \left(\frac{\partial m(\widetilde{s}(\theta), \theta)}{\partial s} \right)^{-1}$$

for all $\theta < \overline{\theta}$. In particular, $\partial m(\widetilde{s}(\theta), \theta)/\partial s > 0$ and hence, $d\widetilde{s}(\theta)/d\theta$ is finite for all $\theta < \overline{\theta}$.

Let $\hat{s}_u$ be a global maximizer θ_u^*. By the properties of the cost function, $\hat{s}_u \in (0, 1)$, and $\theta_u^*(\hat{s}_u) > \underline{\theta}$. Further, $d\theta_u^*(\hat{s}_u)/ds = 0$ and hence, by (5), $\widetilde{s}(\theta_u^*(\hat{s}_u)) = \hat{s}_u$. Indeed, $\hat{s}_u$ is the unique maximizer of θ_u^*: if $\hat{s}'_u \neq \hat{s}_u$ is another maximizer of θ_u^*, then $\widetilde{s}(\theta_u^*(\hat{s}_u)) = \hat{s}'_u$, contradicting that $\widetilde{s}$ is a function.

Suppose there exists $s' \neq \hat{s}_u$ such that $\widetilde{s}(\theta_u^*(s')) = s'$. If s' is not a local extreme of θ_u^*, then s' is a tangency point between θ_u^* and the inverse of $\widetilde{s}$.[32] But this would imply $0 = d\theta_u^*(s')/ds = (d\widetilde{s}(\theta_u^*(s'))/d\theta)^{-1}$, which leads to a contradiction because $d\widetilde{s}(\theta)/d\theta$ is finite for all $\theta < \overline{\theta}$. Furthermore, s' cannot be a local minimum of θ_u^* as this would imply that for some $\theta > \theta_u^*(s')$, there are $s'' < s' < s'''$ with $\theta_u^*(s'') = \theta_u^*(s''') = \theta$ and such that $d\theta_u^*(s'')/ds < 0$ and $d\theta_u^*(s''')/ds > 0$, implying $\widetilde{s}(\theta) < s'' < s''' < \widetilde{s}(\theta)$, a contradiction. Therefore s' can only be a local maximum of θ_u^*. But this would imply that there is a local minimum of θ_u^* in the interval $(\min\{s', \hat{s}_u\}, \max\{s', \hat{s}_u\})$, contradicting that θ_u^* does not have local minima in $(0, 1)$. We conclude that θ_u^* intersects $\widetilde{s}$ only once at $(\hat{s}_u, \theta_u^*(\hat{s}_u))$. The argument for θ_q^* is analogous. The thesis of the Lemma 1 follows immediately. □

Proofs and Ancillary material of Sect. 4

Proof of Lemma 2. We prove that $\mathcal{S}^*$ is weakly increasing using an indirect argument. Consider $\kappa' > \kappa$, $s \in \mathcal{S}^*(\kappa)$, and $s' \in \mathcal{S}^*(\kappa')$. Then,

$$F(s, \underline{\theta}_q) - \kappa F(s, \theta_u^*(s)) \leq F(s', \underline{\theta}_q) - \kappa F(s', \theta_u^*(s'))$$
$$F(s', \underline{\theta}_q) - \kappa' F(s', \theta_u^*(s')) \leq F(s, \underline{\theta}_q) - \kappa' F(s, \theta_u^*(s)).$$

Adding these inequalities yields $(\kappa' - \kappa)\left(F(s, \theta_u^*(s)) - F(s', \theta_u^*(s'))\right) \leq 0$, which implies $F(s, \theta_u^*(s)) \leq F(s', \theta_u^*(s'))$. Now suppose $s' < s$; since densities are

[32] The inverse of $\widetilde{s}$ then could be defined over an open interval containing s' because the tangency occurring under the working hypothesis would imply that $d\widetilde{s}(\theta_u^*(s'))/d\theta \neq 0$.

strictly positive, we have $F(s', \underline{\theta}_q) < F(s, \underline{\theta}_q)$. Thus, $F(s', \underline{\theta}_q) - \kappa F(s', \theta_u^*(s')) < F(s, \underline{\theta}_q) - \kappa F(s, \theta_u^*(s))$, contradicting that s is the equilibrium standard for κ.

Now we show that $\mathcal{S}^*(\underline{\kappa}) = \{0\}$. Notice that $V(0, \underline{\theta}, \underline{\theta}_q; \underline{\kappa}) < V(s, \underline{\theta}, \underline{\theta}_q; \underline{\kappa}) \leq V(s, \theta_u^*(s), \underline{\theta}_q; \underline{\kappa})$, for all $s > 0$, where the strict inequality follows from the fact that $s^*(\underline{\theta}, \underline{\theta}_q; \underline{\kappa}) = 0$, and the weak inequality follows from FOSD.

Since $\mathcal{S}^*$ is weakly increasing, we conclude $\mathcal{S}^*(\kappa) = \{0\}$ for all $\kappa \in (0, \underline{\kappa}]$. An analogous argument proves that $\mathcal{S}^*(\kappa) = \{1\}$ for all $\kappa \in [\overline{\kappa}, \infty)$.

Finally, we show that $\mathcal{S}^*$ is strictly increasing over $\mathcal{S}^{*-1}(0, 1)$. Notice that for all $\kappa \in \mathcal{S}^{*-1}(0, 1)$, $s \in \mathcal{S}^*(\kappa)$ only if the right hand side of (7) is equal to 0. Thus, s can only be an element of $\mathcal{S}^*(\kappa)$ for only one $\kappa \in \mathcal{S}^{*-1}(0, 1)$. □

Proof of Proposition 1. We start with the following lemma:

Lemma 5 *Let $(s_P^*, \theta_u^*, \underline{\theta}_q)$ be a SPNE of Γ_u such that $s_P^* \in (0, 1)$. The equilibrium standard is soft (harsh) if and only if F is strictly submodular (strictly supermodular) at $(s_P^*, \theta_u^*(s_P^*))$.*

Proof Since s_P^* is an interior minimizer of $V(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$, from (7), it satisfies

$$g(s_P^*, \theta_u^*(s_P^*), \underline{\theta}_q) = \kappa \left(1 + \frac{F_\theta(s_P^*, \theta_u^*(s_P^*))}{f(s_P^*, \theta_u^*(s_P^*))} \frac{d\theta_u^*(s_P^*)}{ds}\right). \tag{16}$$

By Claim 1, $F_\theta(z, \theta) < 0$ for all $z \in (0, 1)$ and $\theta < \overline{\theta}$. Thus, the sign of the second term on the right-hand side of (16) is the same as the sign of $-d\theta_u^*(s_P^*)/ds$, which in turn, from condition (5), is the same as the sign of $\partial f(s_P^*, \theta_u^*(s_P^*))/\partial\theta$. Thus, $g(s_P^*, \theta_u^*(s_P^*), \underline{\theta}_q) < \kappa$ if and only if F is strictly submodular at $(s_P^*, \theta_u^*(s_P^*))$.

Furthermore, since $g(\cdot, \theta_u^*(s_P^*), \underline{\theta}_q)$ is strictly increasing, we have that either $g(s, \theta_u^*(s_P^*), \underline{\theta}_q) = \kappa$ for some $s > s_P^*$, or the minimizer of $V(\cdot, \theta_u^*(s_P^*), \underline{\theta}_q)$ is $s = 1$. Thus, $s_P^* < s^*(\theta_u^*(s_P^*), \underline{\theta}_q; \kappa)$ if and only if F is strictly submodular at $(s_P^*, \theta_u^*(s_P^*))$. The argument for the case in which F is strictly supermodular at $(s_P^*, \theta_u^*(s_P^*))$ is analogous. □

Now, for any game Γ_u, let $\widetilde{\kappa}_u := \sup\{\kappa \in (0, \infty) : \sup \mathcal{S}^*(\kappa) \leq \hat{s}_u\}$, where $\hat{s}_u$ is the modularity-switch point.

<u>Part 1</u>. We first prove that $\widetilde{\kappa}_u < \overline{\kappa}$:

Case 1. $\overline{\kappa} = \infty$: For any $\kappa \in (0, \infty)$, we have that

$$\begin{aligned} &\min_{s\in[0,\hat{s}_u]} \left\{V(1, \underline{\theta}, \underline{\theta}_q; \kappa) - V(s, \theta_u^*(s), \underline{\theta}_q; \kappa)\right\} \\ &= \min_{s\in[0,\hat{s}_u]} \left\{1 - F(s, \underline{\theta}_q) - \kappa(1 - F(s, \theta_u^*(s)))\right\}. \end{aligned}$$

This expression is negative for a large enough κ°, thus $s \notin \mathcal{S}^*(\kappa^\circ)$ for all $s \in [0, \hat{s}_u]$. Since $\mathcal{S}^*$ is weakly increasing (Lemma 2), $s \notin \mathcal{S}^*(\kappa')$ for all $s \in [0, \hat{s}_u]$ and $\kappa' > \kappa^\circ$. Thus, $\widetilde{\kappa}_u \leq \kappa^\circ < \infty$.

Case 2. $\overline{\kappa} < \infty$: From Lemma 2, $\mathcal{S}^*(\overline{\kappa}) = \{1\}$. Indeed, $V(1, \underline{\theta}, \underline{\theta}_q; \overline{\kappa}) < \min_{s\in[0,\hat{s}_u]} V(s, \theta_u^*(s), \underline{\theta}_q; \overline{\kappa})$. Notice that $V(s, \theta, \underline{\theta}_q; \cdot)$ is continuous over $(0, \infty)$, for all $(s, \theta) \in [0, 1] \times \Theta$, and by the Maximum Theorem, so it is $\min_{s \in [0, \hat{s}_u]} V(s, \theta_u^*(s), \underline{\theta}_q; \cdot)$. Thus, for small enough $\delta > 0$, we have $V(1, \underline{\theta}, \underline{\theta}_q; \overline{\kappa} - \delta) < \min_{s \in [0, \hat{s}_u]} V(s, \theta_u^*(s), \underline{\theta}_q; \overline{\kappa} - \delta)$. Therefore, $\sup \mathcal{S}^*(\overline{\kappa} - \delta) > \hat{s}_u$ and since $\mathcal{S}^*$ is weakly increasing (Lemma 2), we conclude that $\widetilde{\kappa}_u \leq \overline{\kappa} - \delta < \overline{\kappa}$.

Part 2. Now we show that the optimal standard is ex-post efficient for all $\kappa \in [\overline{\kappa}, \infty)$: From Lemma 2, $\mathcal{S}^*(\kappa) = \{1\}$ for all $\kappa \geq \overline{\kappa}$. Recall that $\theta_u^*(1) = \underline{\theta}$. From (13), $s^*(\underline{\theta}, \underline{\theta}_q; \kappa) = 1$ for all $\kappa \geq \overline{\kappa}$. Thus, the standard is ex-post efficient for all $\kappa \geq \overline{\kappa}$.

Part 3. Now we show that the optimal standard is harsh for all $\kappa \in (\widetilde{\kappa}_u, \overline{\kappa})$: consider any $\kappa \in (\widetilde{\kappa}_u, \overline{\kappa})$ and $s \in \mathcal{S}^*(\kappa)$; since $\mathcal{S}^*$ is weakly increasing over $(0, \infty)$ and strictly increasing over $\mathcal{S}^{*-1}(0, 1)$ (by Lemma 2), we have that $s \in (\hat{s}_u, 1]$. By Lemmata 1 and 5, the optimal standard is harsh if $s \in (\hat{s}_u, 1)$. And if $s = 1$ the optimal standard is harsh because $s^*(\underline{\theta}, \underline{\theta}_q; \kappa) < 1$ for $\kappa < \overline{\kappa}$.

Noting that $\widetilde{\kappa}_u = \inf\{\kappa \in (0, \infty) : \inf \mathcal{S}^*(\kappa) \geq \hat{s}_u\}$, an argument analogous to that of Part 1 shows that $\widetilde{\kappa}_u > \underline{\kappa}$. Similarly, arguments analogous to those in Parts 2 and 3 yield that the optimal standard is ex-post efficient for all $\kappa \in (0, \underline{\kappa}]$ and soft for all $\kappa \in \left(\underline{\kappa}, \widetilde{\kappa}_u\right)$, respectively. □

Remark 3 Assume that (1) also holds at $(0, \theta)$ and $(1, \theta)$ for all $\theta \in \Theta^\circ$. Then, there exists $\overline{\lambda} > 0$ such that the game Γ_u defined by F and the cost function $\lambda^{-1} C_u$ has a strictly increasing pseudo likelihood ratio function v, for all $\lambda \in (0, \overline{\lambda})$.

Proof Let $\theta_u^*(\cdot; \lambda)$ be the unfit candidate's best response for the cost function $\lambda^{-1} C_u$ if $\lambda > 0$ and $\theta_u^*(\cdot; \lambda) = \underline{\theta}$ if $\lambda = 0$. If $\lambda > 0$, the derivative of $F(\cdot, \theta_u^*(\cdot; \lambda))$ is

$$d_u(s; \lambda) := f(s, \theta_u^*(s; \lambda)) - F_\theta(s, \theta_u^*(s; \lambda)) \frac{\partial f(s, \theta_u^*(s; \lambda))}{\partial \theta} \left(\frac{C_u''(\theta_u^*(s; \lambda))}{\lambda} + \frac{\partial^2 F(s, \theta_u^*(s; \lambda))}{\partial \theta^2} \right)^{-1}$$

for all $s \in [0, 1]$. Since $f(\cdot, \underline{\theta}) > 0$, by the Maximum Theorem, there exists $\overline{\lambda}_1 > 0$ such that for all $\lambda \in (0, \overline{\lambda}_1)$, we have that $\min_{s \in [0,1]} \{d_u(s; \lambda)\} > 0$. Thus, for all $\lambda \in (0, \overline{\lambda}_1)$, $v' > 0$ is equivalent to

$$\min_{s \in [0,1]} \left\{ \frac{1}{f(s, \underline{\theta}_q)} \frac{\partial f(s, \underline{\theta}_q)}{\partial s} - \frac{1}{d_u(s; \lambda)} \frac{d\,(d_u(s; \lambda))}{ds} \right\} > 0.$$

Indeed,

$$\begin{aligned} &\lim_{\lambda \to 0} \min_{s \in [0,1]} \left\{ \frac{1}{f(s, \underline{\theta}_q)} \frac{\partial f(s, \underline{\theta}_q)}{\partial s} - \frac{1}{d_u(s; \lambda)} \frac{d(d_u(s; \lambda))}{ds} \right\} \\ &\quad = \min_{s \in [0,1]} \left\{ \frac{1}{f(s, \underline{\theta}_q)} \frac{\partial f(s, \underline{\theta}_q)}{\partial s} - \frac{1}{f(s, \underline{\theta})} \frac{\partial f(s, \underline{\theta})}{\partial s} \right\} > 0, \end{aligned}$$

where the inequality is guaranteed by (1). Thus, there exists $\overline{\lambda} > 0$ such that for all $\lambda \in (0, \overline{\lambda})$, we have $v' > 0$. □

Remark 4 Consider any game Γ_u.

(i) If $v(s) < (>) g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)$ for all $s \in (0, \hat{s}_u)$ ($s \in (\hat{s}_u, 1)$), then $\widetilde{\kappa}_u = g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)$. Therefore, Proposition 1 holds, mutatis mutandis, replacing $\widetilde{\kappa}_u$ by $g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)$. Further, $\mathcal{S}^*(g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)) = \{\hat{s}_u\}$ and thus, the optimal standard is ex-post efficient at $\kappa = g(\hat{s}_u, \theta_u^*(\hat{s}_u), \underline{\theta}_q)$.

(ii) Suppose that v is strictly decreasing over some interval $(\underline{s}, \overline{s})$, with $0 < \underline{s} < \overline{s} < 1$. Then, there exists a prior $\kappa \in (0, \infty)$ and $\delta > 0$ such that $\kappa' < \kappa < \kappa''$, $s' \in \mathcal{S}^*(\kappa')$, and $s'' \in \mathcal{S}^*(\kappa'')$ imply that $s'' - s' > \delta$.

Proof Part (i) is direct, so we proceed directly to prove part (ii). Let $d_u(s) := dF(s, \theta_u^*(s))/ds$.

Case 1. Suppose $d_u(s) > 0$ over $(0, 1)$. For all $s \in (\underline{s}, \overline{s})$, if s is a critical point of $V(\cdot, \theta_u^*(\cdot), \underline{\theta}_q)$, then s is a local maximum and hence, $s \notin \mathcal{S}^*(\kappa)$ for any $\kappa \in (0, \infty)$. Let $\kappa^* := \sup\{\kappa \in (0, \infty) : \sup \mathcal{S}^*(\kappa) \leq \underline{s}\}$ and $\kappa_* := \inf\{\kappa \in (0, \infty) : \inf \mathcal{S}^*(\kappa) \geq \overline{s}\}$.

We observe that $\kappa^* = \kappa_*$: if $\kappa^* < \kappa_*$, then for all $\kappa \in (\kappa^*, \kappa_*)$ we have that $\mathcal{S}^*(\kappa) \cap (\underline{s}, \overline{s}) \neq \emptyset$, contradicting that $s \in (\underline{s}, \overline{s})$ implies that $s \notin \mathcal{S}^*(\kappa)$ for any $\kappa \in (0, \infty)$. On the other hand, if $\kappa^* > \kappa_*$, then for any $\kappa \in (\kappa_*, \kappa^*)$, $\sup \mathcal{S}^*(\kappa) \leq \underline{s}$ and $\inf \mathcal{S}^*(\kappa) \geq \overline{s}$, a contradiction.

Thus, for all $\kappa' < \kappa^*$ we have $\sup \mathcal{S}^*(\kappa') \leq \underline{s}$, and for all $\kappa'' > \kappa^*$ we have $\inf \mathcal{S}^*(\kappa'') \geq \overline{s}$. Hence the thesis holds for $\kappa = \kappa^*$ and all $\delta \in (0, \overline{s} - \underline{s})$.

Case 2. Suppose $d_u(s) \leq 0$ for some $s \in (0, 1)$. Then, for all $\kappa \in (0, \infty)$, $dV(\cdot, \theta_u^*(\cdot), \underline{\theta}_q; \kappa)/ds > 0$ over $(s - \varepsilon, s + \varepsilon)$ for some $\varepsilon \in (0, \min\{s, 1 - s\})$. Thus, $s \notin \mathcal{S}^*(\kappa)$ for any $\kappa \in (0, \infty)$. Analogously to the argument in Case 1, we can define $\kappa'^* := \sup\{\kappa \in (0, \infty) : \sup \mathcal{S}^*(\kappa) \leq s - \varepsilon\}$ and $\kappa'_* := \inf\{\kappa \in (0, \infty) : \inf \mathcal{S}^*(\kappa) \geq s + \varepsilon\}$. The rest of the argument is analogous to Case 1, leading to the statement that thesis holds for $\kappa = \kappa'^*$ and all $\delta \in (0, 2\varepsilon)$. □

Proof of Lemma 3 The proof that $\mathcal{S}^*$ is weakly increasing is indirect and analogous to the one in the proof of Lemma 2, so we omit it.

By definition (see footnote 22), $\underline{\kappa}_q \leq F(s, \theta_q^*(s))/F(s, \underline{\theta})$, which is equivalent to $0 \leq F(s, \theta_q^*(s)) - \underline{\kappa}_q F(s, \underline{\theta})$, for all $s \in (0, 1)$. Thus $0 \in \mathcal{S}^*(\underline{\kappa}_q)$ and since $\mathcal{S}^*$ is weakly increasing, $\mathcal{S}^*(\kappa) = \{0\}$ for all $\kappa < \underline{\kappa}_q$. Also by definition, for all $\kappa > \underline{\kappa}_q$, there exists $s \in (0, 1)$ such that $\kappa > F(s, \theta_q^*(s))/F(s, \underline{\theta})$, which is equivalent to $0 > F(s, \theta_q^*(s)) - \kappa F(s, \underline{\theta})$, and therefore $0 \notin \mathcal{S}^*(\kappa)$.

An analogous argument shows that $1 \in \mathcal{S}^*(\overline{\kappa}_q)$, $\mathcal{S}^*(\kappa) = \{1\}$ for all $\kappa > \overline{\kappa}_q$ and that $1 \notin \mathcal{S}^*(\kappa)$ for all $\kappa < \overline{\kappa}_q$.

Finally, the argument to prove that $\mathcal{S}^*$ is strictly increasing over $(\underline{\kappa}_q, \overline{\kappa}_q)$ is analogous to the corresponding argument in the proof of Lemma 2. □

Proof of Proposition 2. We start with the following lemma:

Lemma 6 *Let $(s_P^*, \underline{\theta}, \theta_q^*)$ be a SPNE of Γ_q such that $s_P^* \in (0, 1)$. The equilibrium standard is harsh (soft) at if and only if F is strictly submodular (strictly supermodular) at $(s_P^*, \theta_q^*(s_P^*))$.*

Proof Since s_P^* is an interior minimizer of $V(\cdot, \underline{\theta}, \theta_q^*(\cdot); \kappa)$, it satisfies $v(s_P^*) = \kappa$. Recall that $F_\theta(s, \theta)$ is strictly negative for all $s \in (0, 1)$ and $\theta \in \Theta^\circ$, thus $g(s_P^*, \underline{\theta}, \theta_q^*(s_P^*)) > v(s_P^*) = \kappa$ if and only if F is strictly submodular at $(s_P^*, \theta_q^*(s_P^*))$.

Since $g(\cdot, \underline{\theta}, \theta_q^*(s_P^*))$ is strictly increasing, either $g(s, \underline{\theta}, \theta_q^*(s_P^*)) = \kappa$ for some $s < s_P^*$, or the minimizer of $V(\cdot, \underline{\theta}, \theta_q^*(s_P^*); \kappa)$ is 0. Thus, $s_P^* > s^*(\underline{\theta}, \theta_q^*(s_P^*); \kappa)$ if and only if F is strictly submodular at $(s_P^*, \theta_q^*(s_P^*))$. The argument for the case in which F is strictly supermodular at $(s_P^*, \theta_q^*(s_P^*))$ is analogous. □

Now we establish that $\widetilde{\kappa}_q \in (\underline{\kappa}_q, \overline{\kappa}_q)$. Observe that $V(\hat{s}_q, \underline{\theta}, \theta_q^*(\hat{s}_q); \widetilde{\kappa}_q) < V(s, \underline{\theta}, \theta_q^*(\hat{s}_q); \widetilde{\kappa}_q) < V(s, \underline{\theta}, \theta_q^*(s); \widetilde{\kappa}_q)$, for all $s \in [0, 1] \setminus \{\hat{s}_q\}$, where the first inequality follows from the fact that $\hat{s}_q$ is the unique minimizer of $V(\cdot, \underline{\theta}, \theta_q^*(\hat{s}_q); \widetilde{\kappa}_q)$,[33] and the second inequality follows from observing that $V(s, \underline{\theta}, \cdot; \widetilde{\kappa}_q)$ is decreasing for all $s \in (0, 1)$ and $\hat{s}_q$ is the unique maximizer of θ_q^*. Thus, $\mathcal{S}^*(\widetilde{\kappa}_q) = \{\hat{s}_q\}$, and since in the proof of Lemma 3 it is shown that $\mathcal{S}^*$ is strictly increasing over $(\underline{\kappa}_q, \overline{\kappa}_q)$, $0 \in \mathcal{S}^*(\underline{\kappa}_q)$ and $1 \in \mathcal{S}^*(\overline{\kappa}_q)$, we conclude $\widetilde{\kappa}_q \in (\underline{\kappa}_q, \overline{\kappa}_q)$.

Now we prove that the optimal standard is ex-post efficient for all $\kappa \in \left(0, \underline{\kappa}_q\right)$. Let $d_q(s) := dF(s, \theta_q^*(s))/ds$. Notice that

$$\underline{\kappa}_q \leq \lim_{s\to 0} F(s, \theta_q^*(s))/F(s, \underline{\theta}) = \lim_{s\to 0} d_q(s)/f(s, \underline{\theta}) \leq \lim_{s\to 0} g(s, \underline{\theta}, \theta_q^*(s)) = g(0, \underline{\theta}, \underline{\theta}_q).$$

Thus, $\underline{\kappa}_q \leq g(0, \underline{\theta}, \underline{\theta}_q)$. The ex-post optimal standard is weakly increasing in κ and $s = 0$ is the only ex-post optimal standard for $\kappa = g(0, \underline{\theta}, \underline{\theta}_q)$ (see Sect. 2.4). Thus, since $\underline{\kappa}_q \leq g(0, \underline{\theta}, \underline{\theta}_q)$ and $\mathcal{S}^*(\kappa) = 0$ for all $\kappa < \underline{\kappa}_q$ (Lemma 3), we have that the optimal standard is ex-post efficient for all $\kappa < \underline{\kappa}_q$.

The argument proving that the optimal standard is harsh for all $\kappa \in (\underline{\kappa}_q, \widetilde{\kappa}_q)$ is analogous to the argument showing that the optimal standard is harsh for all $\kappa \in (\widetilde{\kappa}_u, \overline{\kappa})$ in the proof of Proposition 1. Instead of Lemmata 2 and 5, we use Lemmata 3 and 6.

The arguments showing that the optimal standard is ex-post efficient for all $\kappa \in \left(\overline{\kappa}_q, \infty\right)$ and soft for all $\kappa \in (\widetilde{\kappa}_q, \overline{\kappa}_q)$ are analogous to the arguments showing that the optimal standard is ex-post efficient for all $\kappa \in \left(0, \underline{\kappa}_q\right)$ and harsh for all $\kappa \in (\underline{\kappa}_q, \widetilde{\kappa}_q)$, respectively. Finally, by (7), if $\kappa = \widetilde{\kappa}_q$, the optimal standard is ex-post efficient. □

Proofs and Ancillary material of Sect. 5

Proof of Lemma 4. Let $\mathcal{V}(s) := V(s, \theta_u^*(s), \theta_q^*(s))$ and observe that

$$\begin{aligned}\frac{d\mathcal{V}(s)}{ds} &= f(s, \theta_u^*(s))\left(g(s, \theta_u^*(s), \theta_q^*(s)) - \kappa\right) + F_\theta(s, \theta_q^*(s))\frac{d\theta_q^*(s)}{ds} - \kappa F_\theta(s, \theta_u^*(s))\frac{d\theta_u^*(s)}{ds}\\ &= f(s, \theta_u^*(s))\left(g(s, \theta_u^*(s), \theta_q^*(s)) - \kappa\right) + F_\theta(s, \theta_u^*(s))\frac{d\theta_u^*(s)}{ds}(r(s) - \kappa)\end{aligned}$$

for all $s \in (0, 1) \setminus \{\hat{s}_u\}$. The manager is soft (harsh) in an equilibrium with standard $s_P^* \in (0, 1)$ if $g(s_P^*, \theta_u^*(s_P^*), \theta_q^*(s_P^*)) < (>)\kappa$. Thus, parts (i)-(ii) and (iii)-(iv) follow from the first and second equalities, respectively, using Claim 1 and the fact that readiness is a strategic complement (substitute) of the standard at $s \in [0, 1]$ if and only if $(s, \theta_i^*(s))$ is located in the submodular (supermodular) region of F, for $i = q, u$. □

[33] Notice that by definition of $\widetilde{\kappa}_q$, $\hat{s}_q$ is the ex-post optimal standard if fit candidates' readiness is $\theta_q^*(\hat{s}_q)$, unfit candidates' readiness is $\underline{\theta}$, and $\kappa = \widetilde{\kappa}_q$.

Proof of Proposition 3: In the sequel, when convenient, we make explicit the dependence of d_q, v, r, and θ_q^* on the natural readiness of fit candidates, so, instead of writing $d_q(s)$, $v(s)$, $r(s)$, and $\theta_q^*(s)$, we write $d_q(s;\underline{\theta}_q)$, $v(s;\underline{\theta}_q)$, $r(s;\underline{\theta}_q)$ and $\theta_q^*(s;\underline{\theta}_q)$, respectively, for all $s \in [0,1]$ and $\underline{\theta}_q \in \Theta^\circ$.[34] The proof hinges on the following lemmata:

Lemma 7 *Assume $F_\theta(s,\overline{\theta}) = 0$ for all $s \in (0,1)$. Then, there exists $\underline{\underline{\theta}}_q \in \Theta^\circ$ such that for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_q, \overline{\theta}\right)$, $d_q(s;\underline{\theta}_q) > 0$ for all $s \in [0,1]$.*

Proof The Maximum Theorem and the assumptions of the model on F and the cost functions guarantee that $\min_{s\in[0,1]} d_q(s;\underline{\theta}_q)$ varies continuously with $\underline{\theta}_q$. Further, $F_\theta(s,\overline{\theta}) = 0$ for all $s \in [0,1]$ and, from the hypothesis, $\min_{s\in[0,1]} f(s,\overline{\theta}) > 0$. Thus, there exists $\underline{\underline{\theta}}_q \in \Theta^\circ$ such that $\min_{s\in[0,1]} d_q(s;\underline{\theta}_q) > 0$ for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_q, \overline{\theta}\right)$. □

Lemma 8 *Assume that $F_\theta(s,\overline{\theta}) = 0$ for all $s \in (0,1)$. Then, there exists $\underline{\underline{\theta}}_q \in \Theta^\circ$ such that for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_q, \overline{\theta}\right)$, there exists $\hat{s}(\underline{\theta}_q) \in (0,1)$ such that, if $(s_P^*, \theta_u^*, \theta_q^*)$ is a SPNE of Γ, then the optimal standard is soft if $s_P^* \in (0,\hat{s}(\underline{\theta}_q))$ and harsh if $s_P^* \in (\hat{s}(\underline{\theta}_q), 1)$.*

Proof First, direct computations yield

$$\lim_{(s,\underline{\theta}_q)\to(s_0,\overline{\theta})} \frac{dv(s;\underline{\theta}_q)}{ds} > \lim_{(s,\underline{\theta}_q)\to(s_0,\overline{\theta})} \frac{dg(s,\theta_u^*(s),\theta_q^*(s;\underline{\theta}_q))}{ds} \tag{17}$$

for $s_0 = 0, 1$. Since $v(0;\underline{\theta}_q) = \underline{\kappa}(\underline{\theta}_q)$ and $v(1;\underline{\theta}_q) = \overline{\kappa}(\underline{\theta}_q)$, there exist $0 < \delta_1 < \delta_2 < 1$ and $\underline{\underline{\theta}}_{q_1} \in \Theta^\circ$ such that

$$v(s;\underline{\theta}_q) \begin{cases} > g(s,\theta_u^*(s),\theta_q^*(s;\underline{\theta}_q)) \text{ if } s \in (0,\delta_1) \\ < g(s,\theta_u^*(s),\theta_q^*(s;\underline{\theta}_q)) \text{ if } s \in (\delta_2,1) \end{cases}$$

for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_{q_1}, \overline{\theta}\right)$.

Second, by the log-supermodularity of F, Lemma 1, and direct computations, we have

$$\lim_{(s,\underline{\theta}_q)\to(\hat{s}_u,\overline{\theta})} \frac{dv(s;\underline{\theta}_q)}{ds} < \lim_{(s,\underline{\theta}_q)\to(\hat{s}_u,\overline{\theta})} \frac{dg(s,\theta_u^*(s),\theta_q^*(s;\underline{\theta}_q))}{ds}. \tag{18}$$

[34] Recall that $d_q(s)$ was defined in the proof of Proposition 2: $d_q(s) := dF(s,\theta_q^*(s))/ds$.

Since $v(\hat{s}_u; \overline{\theta}) = g(\hat{s}_u, \theta_u^*(\hat{s}_u), \overline{\theta})$, there exist $\underline{\theta}_{q_2} \in \Theta^\circ$, $\delta_3 \in (\delta_1, \hat{s}_u)$, and $\delta_4 \in (\hat{s}_u, \delta_2)$ such that, for all $\underline{\theta}_q \in \left(\underline{\theta}_{q_2}, \overline{\theta}\right)$, there exists $\hat{s}(\underline{\theta}_q) \in (\delta_3, \delta_4)$ satisfying

$$v(s; \underline{\theta}_q) \begin{cases} > g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) \text{ if } \ s \in (\delta_3, \hat{s}(\underline{\theta}_q)) \\ = g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) \text{ if } \ s = \hat{s}(\underline{\theta}_q) \\ < g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) \text{ if } \ s \in (\hat{s}(\underline{\theta}_q), \delta_4). \end{cases}$$

Third, in game Γ_u, for all $s \in [\delta_1, \delta_3]$, either $v(s; \overline{\theta}) > g(s, \theta_u^*(s), \overline{\theta})$ or $d_u(s) \leq 0$, and $v(s; \overline{\theta}) < g(s, \theta_u^*(s), \overline{\theta})$, for all $s \in [\delta_4, \delta_2]$. Thus, if $d_u > 0$ on $[\delta_1, \delta_3]$, then

$$v(s; \underline{\theta}_q) \begin{cases} > g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) \text{ if } \ s \in [\delta_1, \delta_3] \\ < g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) \text{ if } \ s \in [\delta_4, \delta_2] \end{cases}$$

for all large enough $\underline{\theta}_q \in \Theta^\circ$. We used Weierstrass' Theorem to establish that the difference between $v(s; \overline{\theta})$ and $g(s, \theta_u^*(s), \overline{\theta})$ is strictly greater than zero for all $s \in [\delta_1, \delta_3]$, and the Maximum Theorem to establish that the difference between $v(s; \underline{\theta}_q)$ and $g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q))$ is strictly greater than zero for all $s \in [\delta_1, \delta_3]$, for $\underline{\theta}_q$ close enough to $\overline{\theta}$. An analogous argument applies for the interval $[\delta_4, \delta_2]$. On the other hand, if $d_u(s) < (=) 0$ for some $s \in [\delta_1, \delta_3]$, then by Lemma 7, $v(s; \underline{\theta}_q)$ is negative (not defined) for large enough $\underline{\theta}_q$.

Hence $v(\cdot; \underline{\theta}_q)$ and $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot; \underline{\theta}_q))$ cannot cross over $[\delta_1, \delta_3]$ or $[\delta_4, \delta_2]$ for all $\underline{\theta}_q \in \left(\underline{\theta}_{q_3}, \overline{\theta}\right)$, for a large enough $\underline{\theta}_{q_3} \in \Theta^\circ$.

Therefore, for all $\underline{\theta}_q > \underline{\underline{\theta}_q} := \max\left\{\underline{\theta}_{q_1}, \underline{\theta}_{q_2}, \underline{\theta}_{q_3}\right\}$, $\hat{s}(\underline{\theta}_q)$ is the only root of $v(\cdot; \underline{\theta}_q) - g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot; \underline{\theta}_q))$ over $(0, 1)$, and for any equilibrium $(s_P^*, \theta_q^*, \theta_u^*)$, we have $v(s_P^*; \underline{\theta}_q) > (<) g(s_P^*, \theta_u^*(s_P^*), \theta_q^*(s_P^*; \underline{\theta}_q))$ if $s_P^* < (>) \hat{s}(\underline{\theta}_q)$. Thus, the thesis of the lemma holds. □

Lemma 9 *Assume $F_\theta(s, \overline{\theta}) = 0$ for all $s \in (0, 1)$. Then, there exists $\underline{\underline{\theta}_q} \in \Theta^\circ$ such that for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}_q}, \overline{\theta}\right)$, the thesis of Lemma 2 holds in Γ.*

Proof From Lemma 7, $F(\cdot, \theta_q^*(\cdot; \underline{\theta}_q))$ is strictly increasing for high enough $\underline{\theta}_q$. Thus, the argument showing that $\mathcal{S}^*$ is increasing in the proof of Lemma 2 applies here too.

We prove that $\mathcal{S}^*(\underline{\kappa}(\underline{\theta}_q)) = \{0\}$ (the proof of the statement $\mathcal{S}^*(\overline{\kappa}(\underline{\theta}_q)) = \{1\}$ is analogous). First notice that

$$\lim_{(s, \underline{\theta}_q) \to (0, \overline{\theta})} \frac{dv(s; \underline{\theta}_q)}{ds} = \lim_{s \to 0} \frac{dv(s; \overline{\theta})}{ds} > \lim_{s \to 0} \frac{dg(s, \theta_u^*(s), \overline{\theta})}{ds} > 0. \tag{19}$$

Thus, there exists $\underline{\theta}_{q_1} \in \Theta^\circ$ and $\delta > 0$ such that for all $\underline{\theta}_q \in \left(\underline{\theta}_{q_1}, \overline{\theta}\right)$, $\mathcal{V}(0; \underline{\kappa}(\underline{\theta}_q), \underline{\theta}_q) < \mathcal{V}(s; \underline{\kappa}(\underline{\theta}_q), \underline{\theta}_q)$ for all $s \in (0, \delta)$, where $\mathcal{V}(\cdot; \kappa, \underline{\theta}_q)$ is the expected loss to the manager $\mathcal{V}(s)$ for prior κ when the natural readiness of fit candidates is $\underline{\theta}_q$.

Second, from Lemma 2, we know that $\lim_{\underline{\theta}_q \to \overline{\theta}} \mathcal{V}(0; \underline{\kappa}(\underline{\theta}_q), \underline{\theta}_q) < \lim_{\underline{\theta}_q \to \overline{\theta}} \mathcal{V}(s; \underline{\kappa}(\underline{\theta}_q), \underline{\theta}_q)$ for all $s > 0$. Let $\widetilde{\mathcal{V}}(\cdot, \underline{\theta}_q) := \mathcal{V}(\cdot; \underline{\kappa}(\underline{\theta}_q), \underline{\theta}_q)$ and notice that $\widetilde{\mathcal{V}}(0, \cdot)$ and $\min_{s\in[\delta,1]} \widetilde{\mathcal{V}}(s, \cdot)$ are continuous functions. Thus, there exists $\underline{\theta}_{q_2}$ such that $\widetilde{\mathcal{V}}(0, \underline{\theta}_q) < \min_{s\in[\delta,1]} \widetilde{\mathcal{V}}(s, \underline{\theta}_q)$ for all $\underline{\theta}_q > \underline{\theta}_{q_2}$. Thus, $\widetilde{\mathcal{V}}(0, \underline{\theta}_q) < \widetilde{\mathcal{V}}(s, \underline{\theta}_q)$ for all $s > 0$ and $\underline{\theta}_q > \max\left\{\underline{\theta}_{q_1}, \underline{\theta}_{q_2}\right\}$; that is, $\mathcal{S}^*(\underline{\kappa}(\underline{\theta}_q)) = \{0\}$ for all $\underline{\theta}_q > \max\left\{\underline{\theta}_{q_1}, \underline{\theta}_{q_2}\right\}$.

From (7), for $\kappa > \underline{\kappa}(\underline{\theta}_q)$, we have $d\mathcal{V}(0; \kappa, \underline{\theta}_q)/ds < 0$ and hence $0 \notin \mathcal{S}^*(\kappa)$. Analogously, for $\kappa < \overline{\kappa}(\underline{\theta}_q)$, we have $d\mathcal{V}(1; \kappa, \underline{\theta}_q)/ds > 0$ and, hence, $1 \notin \mathcal{S}^*(\kappa)$.

Finally, an argument analogous to that in the corresponding part of the proof of Lemma 2 proves that $\mathcal{S}^*$ is strictly increasing over $\mathcal{S}^{*-1}(0, 1) = (\underline{\kappa}(\underline{\theta}_q), \overline{\kappa}(\underline{\theta}_q))$. □

The proof of of Proposition 3 is analogous to the proof of Proposition 1, with $\hat{s}(\underline{\theta}_q)$, defined in Lemma 8, playing the role of $\hat{s}_u$[35]: Lemma 8 plays the role of Lemmas 1 and 4, and Lemma 9 plays the role of Lemma 2. □

Proof of Proposition 4: In the sequel, when convenient, we make explicit the dependence of d_u, r, and v on the unfit candidates' cost parameter λ, so, instead of writing $d_u(s)$, $r(s)$, and $v(s)$, we write $d_u(s; \lambda)$, $r(s; \lambda)$, and $v(s; \lambda)$, respectively, for all $s \in [0, 1]$ and $\lambda \in [0, 1]$.[36] The proof hinges on the following lemmata:

Lemma 10 *There exists $\overline{\lambda}_1 > 0$ such that for all $\lambda \in [0, \overline{\lambda}_1)$, $d_u(s; \lambda) > 0$ for all $s \in [0, 1]$.*

Proof The Maximun Theorem guarantees that $\min_{s\in[0,1]} d_u(s; \lambda)$ varies continuously with λ. Further, $\min_{s\in[0,1]} d_u(s; \lambda) = \min_{s\in[0,1]} f(s, \underline{\theta}) > 0$ for $\lambda = 0$. Thus, there exists $\overline{\lambda}_1 > 0$ such that $\min_{s\in[0,1]} d_u(s; \lambda) > 0$ for all $\lambda \in [0, \overline{\lambda}_1)$. □

Lemma 11 *There exists $\overline{\lambda}_2 > 0$ such that for all $\lambda \in \left[0, \overline{\lambda}_2\right)$, there exists $\hat{s}(\lambda) \in (0, 1)$ such that if $(s_P^*, \theta_u^*(\cdot; \lambda), \theta_q^*)$ is a SPNE of $\Gamma(\lambda)$, then the optimal standard is harsh if $s_P^* \in (0, \hat{s}(\lambda))$ and soft if $s_P^* \in (\hat{s}(\lambda), 1)$.*

Proof The proof is analogous to the proof of Lemma 8, using the fact that $v(\cdot; \lambda)$ approaches to $v(\cdot; 0)$ (instead of $v(\cdot; \underline{\theta}_q)$ approaches to $v(\cdot; \overline{\theta})$) and $g(\cdot, \theta_u^*(\cdot; \lambda), \theta_q^*(\cdot))$ approaches to $g(\cdot, \underline{\theta}, \theta_q^*(\cdot))$ (instead of $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot; \underline{\theta}_q))$ approaches to $g(\cdot, \theta_u^*(\cdot), \overline{\theta})$), as $\lambda \to 0$ (instead of as $\underline{\theta}_q \to \overline{\theta}$). □

Lemma 12 *There exists $\overline{\lambda}_3 > 0$ such that for all $\lambda \in \left[0, \overline{\lambda}_3\right)$, the thesis of Lemma 3 holds in $\Gamma(\lambda)$, mutatis mutandis, replacing $\underline{\kappa}_q$ and $\overline{\kappa}_q$ with $\underline{\kappa}_q(\lambda)$ and $\overline{\kappa}_q(\lambda)$, respectively.*

Proof From Lemma 10, $F(\cdot, \theta_u^*(\cdot; \lambda))$ is strictly increasing for all $\lambda < \overline{\lambda}_1$. Then, an argument analogous to the one in the proof of Lemma 3 applies with $\underline{\kappa}_q(\lambda)$ and $\overline{\kappa}_q(\lambda)$ playing the role of $\underline{\kappa}_q$ and $\overline{\kappa}_q$, respectively. □

[35] Case 1 in Part 1 in the the proof of Proposition 1 is not necessary here as the assumption $f > 0$ rules out the possibility that $\overline{\kappa}(\underline{\theta}_q) = \infty$ for all $\underline{\theta}_q \in \Theta^\circ$.

[36] Recall that $d_u(s; \lambda)$ was defined in the proof of Remark 3.

Let $\widetilde{\kappa}_q(\lambda) := g(\hat{s}(\lambda), \theta_u^*(\hat{s}(\lambda);\lambda), \theta_q^*(\hat{s}(\lambda)))$. Since $\lim_{\lambda\to 0}\hat{s}(\lambda) = \hat{s}_q$, we have $\lim_{\lambda\to 0}\widetilde{\kappa}_q(\lambda) = \widetilde{\kappa}_q$. Similarly, $\lim_{\lambda\to 0}\underline{\kappa}_q(\lambda) = \underline{\kappa}_q$ and $\lim_{\lambda\to 0}\overline{\kappa}_q(\lambda) = \overline{\kappa}_q$. Thus, there exists $\overline{\lambda}_4 > 0$ such that $\underline{\kappa}_q(\lambda) < \widetilde{\kappa}_q(\lambda) < \overline{\kappa}_q(\lambda)$ for all $\lambda \in (0, \overline{\lambda}_4)$.

The above lemmata allow us to prove Proposition 4 with an argument analogous to the one in the proof of Proposition 2. Lemma 11 plays the role of Lemma 1 and 4, and Lemma 12 plays the role of Lemma 3. We omit the details. □

Proofs and Ancillary material of Sect. 6

Proof of Remark 1. We provide the argument for the case $s_P^* < s_{NE}^*$ (the case $s_P^* > s_{NE}^*$ is analogous). The manager is better off in the SPNE, because $s_P^* \in \arg\min_{s\in[0,1]}\{V(s, \theta_u^*(s), \theta_q^*(s))\}$ and the BNE $(s_{NE}^*, \theta_{uNE}, \theta_{qNE})$ satisfies $\theta_{uNE} = \theta_u^*(s_{NE}^*)$ and $\theta_{qNE} = \theta_q^*(s_{NE}^*)$. The fact that the candidate is strictly better-off with lower standards follows from the Envelope Theorem, because $dU_i(s, \theta_i^*(s))/ds = f(s, \theta_i^*(s)) > 0$ over $s \in (0, 1)$, for $i = u, q$. □

Proof of Corollary 1. First, we prove the uniqueness of the BNE of the static game, when $\underline{\theta}_q$ is large. It suffices to show that $g(\cdot, \theta_u^*(\cdot), \theta_q^*(\cdot; \underline{\theta}_q))$ is strictly increasing for large $\underline{\theta}_q$.

Notice that $dg(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q))/ds = g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q))\varphi(s; \underline{\theta}_q)$, where

$$\varphi(s;\underline{\theta}_q) := \frac{1}{f(s,\theta_q^*(s;\underline{\theta}_q))}\frac{\partial f(s,\theta_q^*(s;\underline{\theta}_q))}{\partial s} - \frac{1}{f(s,\theta_u^*(s))}\frac{\partial f(s,\theta_u^*(s))}{\partial s} + m(s,\theta_q^*(s;\underline{\theta}_q))\frac{d\theta_q^*(s;\underline{\theta}_q)}{ds} - m(s,\theta_u^*(s))\frac{d\theta_u^*(s)}{ds}, \tag{20}$$

for all $s \in [0, 1]$. Thus, $dg(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q))/ds > 0$ if and only if $\varphi(s; \underline{\theta}_q) > 0$. The first line on the right hand side of (20) is strictly positive for all $s \in [0, 1]$, due to the MLRP, the hypothesis, and the fact that $\theta_q^* > \theta_u^*$. In addition, $F_\theta(\cdot, \bar{\theta}) = 0$ implies $\lim_{\underline{\theta}_q\to\bar{\theta}} m(s, \theta_q^*(s; \underline{\theta}_q))\frac{d\theta_q^*(s;\underline{\theta}_q)}{ds} = 0$. By (5), we have that $-m(s, \theta_u^*(s))\frac{d\theta_u^*(s)}{ds} \geq 0$ for all $s \in [0, 1]$. Since $\varphi(s; \cdot)$ is continuous, by the Maximum Theorem, we have $\lim_{\underline{\theta}_q\to\bar{\theta}} \min_{s\in[0,1]} \varphi(s; \underline{\theta}_q) > 0$.

Now we show that there exists $\underline{\underline{\theta}}_q \in \Theta^\circ$ satisfying that, for all $\underline{\theta}_q \in \left(\underline{\underline{\theta}}_q, \bar{\theta}\right]$, there exists $\widetilde{\kappa}(\underline{\theta}_q) \in (\underline{\kappa}(\underline{\theta}_q), \overline{\kappa}(\underline{\theta}_q))$ such that, for all $\kappa \in (\underline{\kappa}(\underline{\theta}_q), \widetilde{\kappa}(\underline{\theta}_q))$ and SPNE $(s_P^*, \theta_u^*, \theta_q^*(\cdot; \underline{\theta}_q))$ of $\Gamma(\underline{\theta}_q)$, the BNE of $\Gamma_0(\underline{\theta}_q)$ is Pareto dominated by $(s_P^*, \theta_u^*, \theta_q^*(\cdot; \underline{\theta}_q))$.

Take $\underline{\underline{\theta}}_q$ to be the same as in Proposition 3. Consider an arbitrary initial advantage $\underline{\theta}_q > \underline{\underline{\theta}}_q$ and prior $\kappa \in (\underline{\kappa}(\underline{\theta}_q), \widetilde{\kappa}(\underline{\theta}_q))$. From Proposition 3, if $(s_P^*, \theta_u^*, \theta_q^*(\cdot; \underline{\theta}_q))$ is a SPNE of $\Gamma(\underline{\theta}_q)$, then the manager is soft at that equilibrium, and thus, $\kappa = v(s_P^*; \underline{\theta}_q) > g(s_P^*, \theta_u^*(s_P^*), \theta_q^*(s_P^*; \underline{\theta}_q))$. Since $g(1, \theta_u^*(1), \theta_q^*(1; \underline{\theta}_q)) = \overline{\kappa}(\underline{\theta}_q) > \widetilde{\kappa}(\underline{\theta}_q) > \kappa$, there exists $s \in (s_P^*, 1)$ such that $g(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q)) = \kappa$, by the Intermediate Value Theorem. Hence, $(s, \theta_u^*(s), \theta_q^*(s; \underline{\theta}_q))$ is the BNE of $\Gamma_0(\underline{\theta}_q)$. This completes the proof. □

Proof of Corollary 2. First we show that under the hypothesis, there exists $\underline{\theta}_q{}' \in \Theta^\circ$ and $\overline{\lambda}' \in (0,1)$ satisfying that, for all $\left(\underline{\theta}_q, \lambda\right) \in \left(\underline{\theta}_q{}', \overline{\theta}\right] \times \left[0, \overline{\lambda}'\right)$, the BNE of $\Gamma_0(\underline{\theta}_q, \lambda)$ is unique. As in the proof of Corollary 1, uniqueness follows from the fact that for high enough $\underline{\theta}_q$ and λ^{-1}, the minimum with respect to s of the sum of the first three terms on the right-hand side of (20) is positive, and so it is the last term.[37]

Now we fix $\underline{\theta}_q > \underline{\theta}_q{}'$ and prove that there exists $\overline{\lambda} > 0$ satisfying that for all $\lambda \in [0, \overline{\lambda})$, there exists $\widetilde{\kappa}_q(\lambda) \in (\underline{\kappa}_q(\lambda), \overline{\kappa}_q(\lambda))$ such that for all $\kappa \in (\widetilde{\kappa}_q(\lambda), \overline{\kappa}_q(\lambda))$ and SPNE $(s_P^*, \theta_u^*(\cdot;\lambda), \theta_q^*)$ of $\Gamma(\underline{\theta}_q, \lambda)$, the BNE of $\Gamma_0(\underline{\theta}_q, \lambda)$ is Pareto dominated by $(s_P^*, \theta_u^*(\cdot;\lambda), \theta_q^*)$.

Consider $\overline{\lambda}$ as in Proposition 4 and arbitrary $\lambda < \min\left\{\overline{\lambda}, \overline{\lambda}'\right\}$ and prior $\kappa \in (\widetilde{\kappa}_q(\lambda), \overline{\kappa}_q(\lambda))$. From Proposition 4, if $(s_P^*, \theta_u^*(\cdot;\lambda), \theta_q^*)$ is a SPNE of $\Gamma(\underline{\theta}_q, \lambda)$, then the manager is soft at that equilibrium, and thus, $\kappa = v(s_P^*;\lambda) > g(s_P^*, \theta_u^*(s_P^*;\lambda), \theta_q^*(s_P^*))$.

If $\kappa \leq \max_{s\in[s_P^*,1]}\left\{g(s, \theta_u^*(s;\lambda), \theta_q^*(s))\right\}$, then there exists $s \in (s_P^*, 1]$ such that $\kappa = g(s, \theta_u^*(s;\lambda), \theta_q^*(s))$. Thus, $(s, \theta_u^*(s;\lambda), \theta_q^*(s))$ is the BNE of $\Gamma_0(\underline{\theta}_q, \lambda)$. If $\kappa > \max_{s\in[s_P^*,1]}\left\{g(s, \theta_u^*(s;\lambda), \theta_q^*(s))\right\}$, then $\kappa > g(1, \theta_u^*(1;\lambda), \theta_q^*(1))$, and thus, $(1, \underline{\theta}, \underline{\theta}_q)$ is the BNE of $\Gamma_0(\underline{\theta}_q, \lambda)$. □

Proofs and Ancillary material of Sect. 7

Optimality of (s, p) mechanisms We now show that allowing for positive probabilities that (i) a candidate reporting to be unfit is tested, and (ii) a candidate reporting to be fit is outright hired or outright rejected, cannot decrease the manager's expected loss beyond what he can attain within the (s, p) class:

(i) Any mechanism that, with a strictly positive probability, asks a candidate reporting to be unfit to take a test with standard s can be improved by other mechanism that increases the probability of outright rejection of that candidate by $F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))$ times the probability that she is subjected to the test in the former mechanism. Such a change would not affect the unfit candidate's expected payoff from reporting unfit and would decrease the fit candidate's expected payoff from reporting unfit, because $F(s, \theta_u^*(s)) + C_u(\theta_u^*(s)) > F(s, \theta_q^*(s)) + C_q(\theta_q^*(s))$. That is, incentive compatibility would still hold. Finally, the manager would be strictly better-off due to the higher probability of rejecting the unfit candidate.

(ii) Now we show that allowing for a strictly positive probability of outright rejection or outright hiring of a candidate reporting to be fit cannot make the manager better-off. Since the incentive compatibility constraint for the unfit candidate is binding, the probability of rejecting a candidate who claims to be unfit is $p = p_1 + (1 - p_1 - p_2)\left(F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))\right)$, where p_1 (p_2) is the probability of outright rejecting (hiring) a candidate reporting to be fit, and s is the standard of the test. Hence, the

[37] The sum of the first three terms on the right-hand side of (20) is greater than in the proof of Corollary 1 because $\theta_u^*(\cdot;1)$ is now replaced by $\theta_u^*(\cdot;\lambda)$ with $\lambda < 1$.

manager's expected loss is an affine transformation of

$$p_1 + (1 - p_1 - p_2)F(s, \theta_q^*(s)) - \kappa p = p_1 V_M(1) + p_2 V_M(0) + (1 - p_1 - p_2)V_M(s), \tag{21}$$

for all (s, p_1, p_2) with $s \in [0, 1]$ and $p_1, p_2, 1 - p_1 - p_2 \geq 0$. An argument parallel to the one used in the proof of Lemma 3 proves that 0 is the unique minimizer (corresp., is a minimizer, is not a minimizer) of V_M for all $\kappa < \underline{\kappa}_M$ (corresp., for $\kappa = \underline{\kappa}_M$, for all $\kappa > \underline{\kappa}_M$). Similarly, 1 is the unique minimizer (corresp., is a minimizer, is not a minimizer) of V_M for all $\kappa > \overline{\kappa}_M$ (corresp., for $\kappa = \overline{\kappa}_M$, for all $\kappa < \overline{\kappa}_M$). Thus, (21) implies that the mechanism $(s_M, p_M) = (0, 0)$ is optimal for all $\kappa \leq \underline{\kappa}_M$, $(s_M, p_M) = (1, 1)$ is optimal for all $\kappa \geq \overline{\kappa}_M$, and, for all $\kappa \in (\underline{\kappa}_M, \overline{\kappa}_M)$, there exists $s \in (0, 1)$ such that $(s_M, p_M) = (s, F(s, \theta_u^*(s)) + C_u(\theta_u^*(s)))$ is optimal.

Proof of Proposition 5. We proved part (i) in the previous paragraph. For part (ii), we have that $V_M(s) < \mathcal{V}(s)$ for all $s \in (0, 1)$. Notice that $\underline{\kappa}_M \leq \underline{\kappa}_q(1) < \overline{\kappa}_q(1) \leq \overline{\kappa}_M$. An argument parallel to the one in the proof of Lemma 3 reveals that, in Γ, 0 is the unique minimizer (corresp., is a minimizer, is not a minimizer) of $\mathcal{V}$ for all $\kappa < \underline{\kappa}_q(1)$ (corresp., for $\kappa = \underline{\kappa}_q(1)$, for all $\kappa > \underline{\kappa}_q(1)$). Similarly, 1 is the unique minimizer (corresp., is a minimizer, is not a minimizer) of $\mathcal{V}$ for all $\kappa > \overline{\kappa}_q(1)$ (corresp., for $\kappa = \overline{\kappa}_q(1)$, for all $\kappa < \overline{\kappa}_q(1)$). Thus, $\min_{s \in [0,1]} V_M(s) < \min_{s \in [0,1]} \mathcal{V}(s)$ for all $\kappa \in (\underline{\kappa}_M, \overline{\kappa}_M)$.

Finally, we show that $\mathcal{S}_M^*$ is increasing. Consider $\kappa' > \kappa$, $s \in \mathcal{S}_M^*(\kappa)$, and $s' \in \mathcal{S}_M^*(\kappa')$. Then,

$$F(s, \theta_q^*(s)) - \kappa \left(F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))\right) \leq F(s', \theta_q^*(s')) - \kappa \left(F(s', \theta_u^*(s')) + C_u(\theta_u^*(s'))\right)$$
$$F(s', \theta_q^*(s')) - \kappa' \left(F(s', \theta_u^*(s')) + C_u(\theta_u^*(s'))\right) \leq F(s, \theta_q^*(s)) - \kappa' \left(F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))\right).$$

Adding these inequalities yields

$$(\kappa' - \kappa) \left[\left(F(s, \theta_u^*(s)) + C_u(\theta_u^*(s))\right) - \left(F(s', \theta_u^*(s')) + C_u(\theta_u^*(s'))\right)\right] \leq 0,$$

which implies $F(s, \theta_u^*(s)) + C_u(\theta_u^*(s)) \leq F(s', \theta_u^*(s')) + C_u(\theta_u^*(s'))$. By the Envelope Theorem, $F(\cdot, \theta_u^*(\cdot)) + C_u(\theta_u^*(\cdot))$ is strictly increasing; thus, $s \leq s'$. □

Proof of Proposition 6. Let $v(\cdot; \underline{\theta}_q, \lambda) := d_q(\cdot; \underline{\theta}_q)/d_u(\cdot; \lambda)$, for all $\underline{\theta}_q \in \Theta^\circ$ and $\lambda \in (0, 1)$. We denote explicitly the dependence dependence of $\underline{\kappa}_q(\lambda)$ and $\overline{\kappa}_q(\lambda)$ on $\underline{\theta}_q$, writing $\underline{\kappa}_q(\underline{\theta}_q, \lambda)$ and $\overline{\kappa}_q(\underline{\theta}_q, \lambda)$, respectively, for all $\underline{\theta}_q \in \Theta^\circ$ and $\lambda \in (0, 1)$.

Let $v_M(s; \lambda)$ be defined as $v_M(s)$ with $\theta_u^* = \theta_u^*(\cdot; \lambda)$. By the log-supermodularity assumption on the density of F and the Maximum Theorem,

$$\lim_{(\underline{\theta}_q, \lambda) \to (\overline{\theta}, 0)} \min_{s \in [0,1]} \frac{dv(s; \underline{\theta}_q, \lambda)}{ds} = \lim_{(\underline{\theta}_q, \lambda) \to (\overline{\theta}, 0)} \min_{s \in [0,1]} \frac{dv_M(s; \lambda)}{ds} = \min_{s \in [0,1]} \frac{dg(s, \underline{\theta}, \overline{\theta})}{ds} > 0.$$

Thus, there exist $\underline{\underline{\theta}}_q$ and $\overline{\lambda}$ such that $v(\cdot; \underline{\theta}_q, \lambda)$ and $v_M(\cdot; \lambda)$ are strictly increasing for all $(\underline{\theta}_q, \lambda)$ such that $\underline{\theta}_q > \underline{\underline{\theta}}_q$ and $\lambda < \overline{\lambda}$. Further, $v(s; \underline{\theta}_q, \lambda) = v_M(s; \lambda)$ at $s = 0, 1$.

Suppose $\underline{\theta}_q > \underline{\theta}_q$ and $\lambda < \overline{\lambda}$. For all $\kappa \leq \underline{\kappa}_M(\underline{\theta}_q, \lambda)$, $s_M = 0$ and for all $\kappa \geq \overline{\kappa}_M(\underline{\theta}_q, \lambda)$, $s_M = 1$. Similarly, for all $\kappa \leq \underline{\kappa}_q(\underline{\theta}_q, \lambda)$, $s_P^* = 0$ and for all $\kappa \geq \overline{\kappa}_q(\underline{\theta}_q, \lambda)$, $s_P^* = 1$. Furthermore, for all $\kappa \in (\underline{\kappa}_M(\underline{\theta}_q, \lambda), \overline{\kappa}_M(\underline{\theta}_q, \lambda))$, s_M is the root of $v_M(\cdot; \lambda) = \kappa$. Similarly, for all $\kappa \in (\underline{\kappa}_q(\underline{\theta}_q, \lambda), \overline{\kappa}_q(\underline{\theta}_q, \lambda))$, s_P^* is the root of $v(\cdot; \underline{\theta}_q, \lambda) = \kappa$. Since $\underline{\kappa}_M(\underline{\theta}_q, \lambda) \leq \underline{\kappa}_q(\underline{\theta}_q, \lambda) < \overline{\kappa}_q(\underline{\theta}_q, \lambda) \leq \overline{\kappa}_M(\underline{\theta}_q, \lambda)$, and over $(0, 1)$, $v(s; \underline{\theta}_q, \lambda) > (=, <) v_M(s; \lambda)$ if $s < (=, >) \hat{s}_u(\lambda)$, we conclude $s_P^* < (>) s_M$ if and only if $\kappa \in (\underline{\kappa}_M(\underline{\theta}_q, \lambda), v(\hat{s}_u(\lambda); \underline{\theta}_q, \lambda))$ $(\kappa \in (v(\hat{s}_u(\lambda); \underline{\theta}_q, \lambda), \overline{\kappa}_M(\underline{\theta}_q, \lambda)))$. □

References

Ball, I.: Scoring Strategic Agents, Working Paper, Yale University (2020)

Ben-Porath, E., Dekel, E., Lipman, B.L.: Mechanisms with evidence: commitment and robustness. Econometrica **87**, 529–566 (2019)

Boleslavsky, R., Cotton, C.: Grading standards and education quality. Am. Econ. J. Microecon. **7**, 248–79 (2015)

Butcher, J.N.: Psychological assessment of airline pilot applicants with the MMPI-2. J. Pers. Assess. **62**, 31–44 (1994)

Butcher, J.N.: Assessing pilots with the wrong stuff: a call for research on emotional health factors in commercial aviators. Int. J. Sel. Assess. **10**, 168–184 (2002)

Caska, B.: Psychometrics at Work: How to Ensure Test Results You Can Trust, DBS Business Review, 3 (2020)

Cunningham, T., Moreno de Barreda, I.: Effective Signal-Jamming. University of Oxford, Mimeo (2019)

Darrow, J.J., Avorn, J., Kesselheim, A.S.: The FDA Breakthrough-Drug Designation-Four Years of Experience. N. Engl. J. Med. **378**(15), 1444–1453 (2018)

Dattner, B.: How to use psychometric testing in hiring. Harv. Bus. Rev. **12**, 2013 (2013)

de Haan, T., Offerman, T., Sloof, R.: Noisy signaling: theory and experiment. Games Econ. Behav. **73**, 402–428 (2011)

Degan, A., Li, M.: Persuasion with costly precision. Econ. Theor. **72**, 869–908 (2021)

DeGroot, M.H.: Optimal Statistical Decisions. John Wiley & Sons (2005)

Demougin, D., Fluet, C.: Rules of proof, courts, and incentives. Rand J. Econ. **39**, 20–40 (2008)

Di Tillio, A., Ottaviani, M., Sørensen, P.N.: Persuasion bias in science: can economics help? Econ. J. **127**, F266–F304 (2017)

Di Tillio, A., Ottaviani, M., Sørensen, P.N.: Strategic sample selection. Econometrica **89**, 911–953 (2021)

Ederer, F., Holden, R., Meyer, M.: Gaming and strategic opacity in incentive provision. RAND J. Econ. **49**, 819–854 (2018)

Espinosa, F., Ray, D.: Too good to be true? Microeconomics, retention rules for noisy agents. Am. Econ. J. **15**(2), 493–535 (2020)

Fanelli, D.: How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE **4**, e5738 (2009)

Frankel, A., Kartik, N.: Improving information from manipulable data. J. Eur. Econ. Assoc. **20**, 79–115 (2022)

Ganuza, J., Gomez, F., Penalva, J.: Minimizing errors, maximizing incentives: optimal court decisions and the quality of evidence, Working Paper, University Pompeu Fabra (2015)

Gerlach, H.: Self-reporting, investigation, and evidentiary standards. J. Law Econ. **56**, 1061–1090 (2013)

Goldacre, B.: Bad pharma: how drug companies mislead doctors and harm patients. Macmillan (2014)

Grove, W.A., Wu, S.: The search for economics talent: doctoral completion and research productivity. Am. Econ. Rev. **97**, 506–511 (2007)

Henry, E., Ottaviani, M.: Research and the approval process: the organization of persuasion. Am. Econ. Rev. **109**, 911–55 (2019)

Herresthal, C.: Hidden testing and selective disclosure of evidence. J. Econ. Theory **200**, 105402 (2022)

Hu, L., Immorlica, N., Vaughan, J.W.: The disparate effects of strategic manipulation. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 259–268 (2019)

Kamenica, E.: Bayesian persuasion and information design. Annu. Rev. Econ. **11**, 249–272 (2019)

Kamenica, E., Kim, K., Zapechelnyuk, A.: Bayesian persuasion and information design: perspectives and open issues. Econ. Theor. **72**, 701–704 (2021)
Kaplow, L.: On the optimal burden of proof. J. Polit. Econ. **119**, 1104–1140 (2011)
Karlin, S., Rubin, H.: The theory of decision procedures for distributions with monotone likelihood ratio. Ann. Math. Stat. **27**, 272–299 (1956)
Li, H.: A theory of conservatism. J. Polit. Econ. **109**, 617–636 (2001)
Li, H., Suen, W.: Delegating decisions to experts. J. Polit. Econ. **112**, S311–S335 (2004)
McClellan, A.: Experimentation and approval mechanisms. Econometrica **90**(5), 2215–2247 (2022)
Milgrom, P.R.: Good news and bad news: representation theorems and applications. Bell J. Econ. **73**, 380–391 (1981)
Milli, S., Miller, J., Dragan, A. D., Hardt, M.: The social cost of strategic classification. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 230–239 (2019)
Mungan, M.C., Samuel, A.: Mimicking, errors, and the optimal standard of proof. Econ. Lett. **174**, 18–21 (2019)
Mussa, M., Rosen, S.: Monopoly and product quality. J. Econ. Theory **18**, 301–317 (1978)
Neyman, J., Pearson, E.S.: IX. On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. R. Soc. Lond. A **231**, 289–337 (1933)
Perez-Richet, E., Skreta, V.: Test design under falsification. Econometrica **90**, 1109–1142 (2022)
Sanchirico, C.W.: Evidence Tampering. Duke Law J. **53**, 1215–1336 (2004)
Sanchirico, C.W.: Detection avoidance and enforcement theory. In: Encyclopedia of Law and Economics. Edward Elgar Publishing Limited (2012)
Stephenson, M.C.: Evidentiary Standards and Information Acquisition in Public Law. Am. Law Econ. Rev. **10**, 351–387 (2008)
Taylor, C.R., Yildirim, H.: Subjective performance and the value of blind evaluation. Rev. Econ. Stud. **78**, 762–794 (2011)
Tsakas, E., Tsakas, N., Xefteris, D.: Resisting persuasion. Econ. Theor. **72**, 723–742 (2021)
Vohra, R., Espinosa, F., Ray, D.: A principal-agent relationship with no advantage to commitment. In: Pure and Applied Functional Analysis, pp. 1043–1064 (2021)
Whitmeyer, M.: Bayesian elicitation, (2019) arXiv preprint arXiv:1902.00976
Zapechelnyuk, A.: Optimal quality certification. Am. Econ. Rev. Insights **2**, 161–76 (2020)