Screening for Experiments*

Daehong Min[†a]

[a] *Division of Social Science, New York University Abu Dhabi*[‡]

June 29, 2023

Abstract

I study a problem in which the principal is a decision maker and the agent is an "experimenter." Neither the agent nor the principal observes the true state, but the agent can conduct an experiment that reveals information about the true state. The agent has private information about which experiments are feasible, his type. Before the agent conducts an experiment, the principal commits to a decision rule which is contingent on the experiments and their results. When the first-best outcome is unachievable, the principal faces a trade-off between *the quality of the experiment* and *the ex post optimal decisions given experimental results*. I characterize two kinds of optimal decision rules: one that sacrifices the ex post optimal decisions for the quality of the experiment, and the other that resolves the trade-off the other way around; which one is optimal depends on the properties of each type's set of feasible experiments.

JEL Classification: C61; D82; D83; D86

Keywords: Information design; mechanism design without transfers; information control; Bayesian persuasion; experiments

*I am grateful for comments by Andreas Blume, Martin Dufwenberg, Asaf Plan, Stanley Reynolds, Jaehyon Nahm, Nathan Yoder, John Wooders, and seminar participants at the 28th International Conference on Game Theory at Stony Brook, 2018 Korea's Allied Economic Associations Annual Meeting, and Korean Econometric Society Micro Seminar.

†Corresponding Author: daehong.min@nyu.edu

‡Present Address: New York University Abu Dhabi, A5/1129, Abu Dhabi, Saadiyat Island, United Arab Emirates

1 Introduction

Consider an agent who seeks "approval" for a "project" with unknown quality from a principal. Neither the agent nor the principal can directly observe the quality of the project. The agent can provide some evidence associated with the project's true quality by conducting an experiment or a test. However, while the principal wants to approve the project only if its quality is "*Good*," the agent always wants to get approval regardless of the true quality. Thus, the agent has an incentive to choose an experiment or test which favors him, rather than the most informative one.

There is a simple decision rule to which the principal can commit to resolve this conflict of interests: the principal announces that she will approve the project only if the agent conducts the "best" experiment and the result from it is good; otherwise, she will disapprove. However, when the principal does not know which experiments are feasible for the agent, this decision rule is not implementable as she cannot determine the best experiment; furthermore, it also might backfire on the principal: if the experiment, which the principal believes is the best, is not feasible in fact, the principal disapproves the project even when the agent provides evidence supporting the project by conducting the second-best experiment—the most informative experiment among feasible ones.

Example . *(the FDA and a Pharmaceutical Company)* Consider a pharmaceutical company seeking approval for a new drug from the FDA. The company needs to design and conduct a clinical trial[1] to persuade the FDA. Knowing that the company might select a clinical trial in favor of the approval decision, the FDA might want to declare[2] that it approves only if the trial is of high quality and the results are good. The quality of a clinical trial depends on the number of participants and the recruited participants' characteristics. However, the FDA cannot verify how many participants the company can recruit or whether it selectively recruits the participants with a certain characteristic. Hence, it is hard to determine the best design and the FDA might set an unrealistically high quality requirement. It especially backfires on the FDA when the new drug is for a rare disease[3] or its decision needs to be made in time.[4]

[1]The clinical trials are often conducted by third-parties called CRO (Contract Research Organization). However, their studies are funded by pharmaceutical companies; Beckelman, Li, and Gross (2003) show that company-sponsored studies favor the companies' interests.

[2]21 CFR §314.126 articulates the characteristics of *adequate and well-controlled* clinical trials.

[3]The FDA rejected the new drug application for Waylivra which is a treatment for a rare disease. In response, a bill called the HEART Act of 2020 which intends to make the approval process "smoother" has been initiated with the support of the patient community (See https://www.cnn.com/2020/07/19/health/rare-disease-fda-drug-review-process-bill-wellness/index.html).

[4]Facing the COVID-19 pandemic, the FDA granted EUA (Emergency Use Authorization) for some vaccines instead of requiring the usual stringent standard for the clinical trials due to the urgent need for vaccines.

Then, the question becomes "facing uncertainty about which experiments are feasible for the agent, what is the optimal decision rule to which the principal wants to commit?" To address this question, I study a principal-agent model in which the principal is a decision maker and the agent is an "experimenter." The principal wants the agent to conduct the most informative experiment so that she can make a right decision contingent on the true state. However, the agent, who is only interested in getting approval from the principal, prefers a less informative experiment which has the highest false positive rate up to the degree to which the positive outcome is still convincing enough to get approval. The principal can observe what kind of experiment has been conducted and the experimental outcomes, *but* she does not know what experiments are feasible for the agent. Hence, the principal cannot tell whether the agent does not conduct the most informative experiment due to the feasibility constraint or not. In the model, the principal moves first and commits to a decision rule which is contingent on what she can observe, an experiment and the experimental outcomes. The principal's decision rule is publicly announced. The agent privately learns his type, a set of feasible experiments, and conducts an experiment among all feasible experiments for him. Then, the principal makes her decision according to the decision rule to which she committed. I assume that there is no transfer from the principal to the agent.

The model is in a binary environment; there are two states, $\{Good, Bad\}$, two types of the agent, and the principal has two actions, $\{Approval, Disapproval\}$. Furthermore, I only consider the experiments which have binary outcomes, $\{good, bad\}$, for simplicity. Then a binary experiment can be summarized by two probabilities, $P(g|G)$ and $P(b|B)$: two probabilities of generating the *true positive* and the *true negative* outcomes. More importantly, I assume that one type ("big" type) has a larger set of feasible experiments than the other ("small" type). The principal knows that there are two types of the agent and the set of feasible experiments for each type *but* does not know the true type of the agent. Then the principal's problem becomes that of designing an incentive compatible menu offer (which does *not* involve transfers): a menu offer consisting of two options, where each option is destined for each type. Each option consists of three items, an experiment and two approval probabilities respectively associated with *good* and *bad* outcomes. In short, the principal writes a menu offer which specifies (i) what experiment she wants each type of the agent to conduct and (ii) how she will make her decision based on the outcomes when the agent conducts an experiment specified in the menu offer. Note that, since the big type has a larger set of feasible experiments, the principal can easily make the menu offer "incentive compatible" for the small type by asking the big type to conduct an experiment that is feasible for him but not for the small type. However, to make the menu offer incentive compatible for the big type, the principal needs to make the big type's option more attractive than the small type's since any experiment feasible for the small type is always feasible for the big type.

While I consider this binary environment, I allow the set of feasible experiments for each type of the agent to be as general as possible. The main result of this paper is a characterization of the principal's optimal decision rule which is robust to a certain class of the structures on the sets of feasible experiments. I start by characterizing the first-best outcome, the best outcome the principal could achieve if she could observe the type of the agent. In the first-best case, as the principal knows what experiments are feasible for the agent, the principal can force the agent to conduct the experiment that the principal most prefers among the feasible experiments for the agent; the principal can simply achieve this by declaring she will approve *only if* the agent conducts her "favorite" experiment among all feasible experiments *and* the good outcome is realized. Thus, no matter what the type of the agent is, the agent ends up conducting the principal's favorite experiment, and the principal makes a "right" decision based on the experimental outcomes.

First I show that, despite the information asymmetry, the principal can still achieve the first-best outcome under certain set structures. If the principal's favorite experiment in the big type's set generates the positive outcome more frequently than that in the small type's set, the following decision rule can induce the first-best outcome: the decision rule (or the menu offer) which assigns the principal's favorite experiment in each type's set to each type and approve only if the positive outcome is realized. Note that, under this decision rule, the higher the probability of generating the positive outcome, the higher the probability of getting the *Approval* action. Thus, when a set structure satisfies the condition above, the big type does not have an incentive to conduct the experiment designated to the small type (the principal's favorite experiment in the small type's set). Furthermore, the small type cannot conduct the principal's favorite experiment in the big type's set since the big type's set is bigger than the small type's. Thus the principal's decision rule is incentive compatible, and she can achieve the first-best outcome.

Then I consider the cases in which the condition above does not hold: the principal's favorite experiment in the small type's set generates the positive outcome more frequently than that in the big type's set. In this case, the previous menu offer is not incentive compatible anymore; the big type would conduct the principal's favorite experiment in the small type's set which more frequently gives him the approval decision. Thus any incentive compatible menu offer requires the principal to give up either (i) assigning her favorite experiment to each type or (ii) the *ex post* optimal decisions (approve only if the positive outcome is realized). In other words, the principal faces a trade-off between *information quality* and *decision quality.*

The optimal way to resolve this trade-off depends on a given set structure. I mainly focus on the decision rule which can achieve one of two desirable outcomes in the first-best case, assigning the favorite experiments or making the ex post optimal decisions. First, I show that the principal's optimal decision rule is to assign her favorite experiment to each type at the cost of giving up the ex post optimal decisions if the favorite experiment in each set satisfies some *quality requirements* in terms of informativeness measured by the *positive* or *negative likelihood ratios.* In particular, if the principal's favorite experiment in each set Blackwell-dominates all experiments in that set,[5] the favorite experiment coincides with the most informative experiment in terms of the positive and negative likelihood ratios and, thus, meets the quality requirements.

Under this kind of optimal decision rule, the principal incentivises the big type to conduct her favorite experiment in the big type's set by adjusting the *ex post decisions* based on the outcomes. There are two ways to achieve this: (i) the principal takes the *Approval* action with a positive probability even when the *negative* outcome is realized from the principal's favorite experiment in the big type's set, and (ii) the principal takes the *Disapproval* action with a positive probability even when the *positive* outcome is realized from the principal's favorite experiment in the small type's set. I also show that it is optimal to make this adjustment *only* to the type whose probability of being true is relatively lower than the other: if the probability that the agent is the big type is relatively lower than the other, the principal *only* distorts the ex post decisions associated with the experiment assigned to the big type, and vice versa. Note that this contrasts to the classical result often referred to as "no distortion at the top but distortion everywhere": if the big type is relatively unlikely, distortion occurs at the "top" (i.e., the ex post decisions for the big type.)

Secondly, I show that the other way of resolving the trade-off can also be optimal: the principal's optimal decision rule is to make the ex post optimal decisions at the cost of giving up the favorite experiments if the principal's favorite experiment in each type's set *sufficiently* deviates from the most informative experiment in terms of the positive or negative likelihood ratios. Here the principal's ex post decision is simple and clear: she takes the approval action only if the outcome is positive. Then this decision rule assigns an experiment to each type so that this decision rule is incentive compatible: an experiment assigned to each type has the same *ex ante* probability of generating the positive outcome. This decision rule does not incur a loss in terms of ex post optimality; but the loss occurs as the assigned experiments are not the first-best choices.

[5] That is, the experiment that the principal most prefers is the *most informative* one in the sense of Blackwell (1953).

1.1 Related Literature

This paper relates to the literature which studies a problem between a decision maker (a receiver) and an information provider (a sender). The cheap-talk literature studies information transmission without commitment (c.f. Crawford and Sobel (1982)) or with commitment to a mediation scheme (Blume, Board, and Kawamura (2007) and Goltsman, Hörner, Pavlov, and Squintani (2009)). Kamenica and Gentzkow (2011) have looked at a model in which a sender provides information by committing to an information structure (or an experiment). Herresthal (2017) studies a problem in which a sender conducts a given experiment in a dynamic setting. She provides conditions under which a decision maker benefits from non-transparency in the sender's information acquisition process. The agent in this paper also provides information by conducting an experiment; but here the agent can select an experiment he wants to conduct and the decision maker observes both the experiment conducted and the outcomes from it. Thus, the way that the agent provides information is essentially the same as how the sender in Kamenica and Gentzkow (2011) does. However, the main focus of this paper is on the decision maker's side. The model in this paper can be thought of as an "upside down version" of the model in Kamenica and Gentzkow (2011) in the sense that, here, the decision maker moves first and commits to a decision rule while, in their paper, the sender moves first and commits to an information structure.

This paper also relates to the mechanism design literature which studies principal-agent problems. Most studies assume that different types of the agent have the same set of actions (c.f. Myerson (1979)). The model in this paper is different from these studies since types differ in the set of actions available to them. Green and Laffont (1986) consider a model in which types of the agent have different message spaces. This structure on the message space prevents a type from mimicking other types if that type does not have messages which other types have. They provide a necessary and sufficient condition for the revelation principle to be valid in their setting. The model in this paper satisfies their condition and thus the revelation principle can be applied. Among the mechanism design literature, Carroll (2017) looks at a problem in which the agent collects information for the principal. Then he shows that a simple contract which approximates a linear contract is optimal for the principal. In both Carroll (2017) and this paper, the principal is uncertain about which experiments are feasible for the agent. However, he departs from the traditional Bayesian approach which this paper takes: in his setting, the principal does not have a prior belief over the different sets of feasible experiments and looks for a contract which gives her the maximum payoff among all possible worst-case scenarios. Furthermore, in his work, the agent privately observes the experimental outcomes; but here the experimental outcomes are publicly disclosed.

Recently, there have been many studies on information design (c.f. Bergemann and Morris (2016a)), and Kamenica and Gentzkow (2011) is a special case of the information design problem as Bergemann and Morris (2016b) address. Thus the agent in this paper can be considered as an information designer. Furthermore, the principal in this paper can be considered as a mechanism designer whose problem is to design her decision rule. This paper connects these two design problems; here I look at a mechanism design problem in which the agent is an information designer. To my knowledge, there are two papers which also connect these two fields, Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), and Yoder (2022). Kolotilin et al. (2017) consider a situation in which the information designer cannot observe the private information of a decision maker. In their setting, the information designer designs not only an information structure but also a mechanism which elicits the private information of the decision maker. Yoder (2022) is more closely related to this paper as his work also considers a similar setting in which the agent is an information designer and the principal is a mechanism designer. The principal in his setting writes a contract which involves a *transfer* to the agent. However, I consider a setting in which such a transfer is impossible. Furthermore, the agent's private information here is about the set of feasible experiments; but in his work, the agent is privately informed about the costs of conducting experiments.[6]

The remainder of this paper is structured as follows. In section 2, I formally introduce the model. In section 3, I characterize the optimal decision rules. Finally, section 4 concludes this paper.

2 The Model

There are an agent (he) and a principal (she). There are two states, $\Omega = \{G, B\}$, with a common prior, $P(G) = p \in (0, 1/2)$ (G for "*Good*" state and B for "*Bad*" state). The principal has two actions, $\mathcal{A} = \{A, D\}$ (A for "*Approval*" and D for "*Disapproval*"). The principal's payoff depends on both the action taken by her and the true state while the agent's payoff only depends on the principal's action. While the agent always wants the principal to take action A, the principal wants to take action A only if the state is G. Table 1 summarizes the payoffs of both the agent and the principal, where the first entry is the agent's payoff and the second one is the principal's payoff in each cell.[7]

[6]Though I assume that there is no cost of conducting experiments for both types, one might think that the small type's cost of conducting experiments that are only feasible for the big type is prohibitive.

[7]For simplicity I assume all payoffs are either 1 or 0. But, for example, replacing 1 with 10 and 0 with -5 will not affect the main results in this paper. That is, all essential results in this paper remain valid under any

	G	B
A	1,1	1,0
D	0,0	0,1

Table 1: Payoff Matrix

Neither the agent nor the principal can directly observe the true state. However, the agent has a "tool" to investigate the true state while the principal does not. Namely, the agent can conduct an experiment that provides information about the true state and the principal only can learn about the true state via the agent's experiment.

The agent can conduct an experiment only if that experiment is available to him. I assume that the agent has different sets of feasible experiments depending on his type. Namely, there are two types of the agent, $\Theta = \{\theta_1, \theta_2\}$, which determine the set of feasible experiments. I denote type θ_i's set of feasible experiments by S_i for $i = 1, 2$. The agent's type is private information but the prior distribution over the agent's type space, $P(\theta_1) = t$, is commonly known. I assume that a typical element in S_i is a binary experiment, π_i^k, that has binary outcomes, $\{g, b\}$. Then $\pi_i^k \in S_i$ is characterized by the following two distributions over the outcome space, $\{g, b\}$, conditioning on the true states:

$$\pi_i^k(g|G, \theta_i) = y_i^k \qquad \pi_i^k(g|B, \theta_i) = 1 - x_i^k$$
$$\pi_i^k(b|G, \theta_i) = 1 - y_i^k \qquad \pi_i^k(b|B, \theta_i) = x_i^k$$

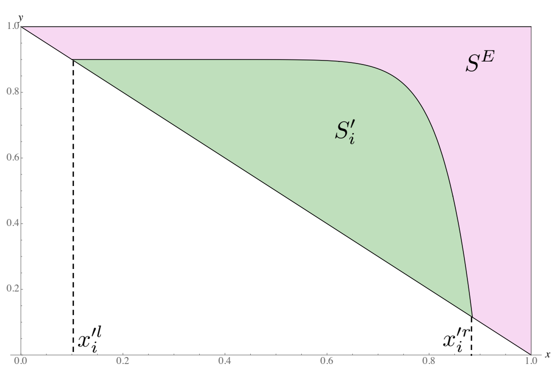


Figure 1: Examples of S_i

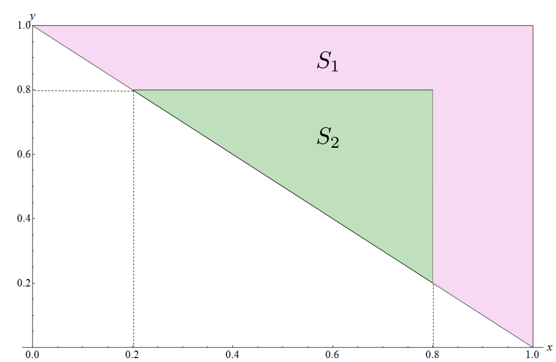


Figure 2: Example of S_1 and S_2

First note that an experiment, π_i^k, is fully characterized by x_i^k and y_i^k, where x_i^k and y_i^k are respectively the probability of generating the *true negative* and the *true positive* outcomes.

preference such that $U^P(A, G) > U^P(D, G)$, $U^P(D, B) > U^P(A, B)$, and $U^A(A, \cdot) > U^A(D, \cdot)$ at both states, where $U^A(\cdot)$ and $U^P(\cdot)$ are the payoffs of the principal and the agent at a cell respectively. Furthermore, one might think that the principal must get strictly lower payoffs in cell (A, B) than in cell (D, G). As long as $U^P(A, G) > U^P(D, G)$ and $U^P(D, B) > U^P(A, B)$, it is possible to incorporate such a preference into the model.

Thus an experiment can be represented by a point, (x, y), in the unit square, $[0,1] \times [0,1]$, in $\mathbb{R}^2$. Then the set of all *possible* experiments is simply the unit square in $\mathbb{R}^2$. It is without loss of generality to assume that the set of all *possible* experiments is the "upper-right" triangle shown as S^E in Figure 1; that is, we can restrict our attention to the experiments such that $y \geq 1 - x$.[8]

Secondly, since an experiment is a point, S_i is a collection of some points in S^E. More specifically, I define S_i as follows:

Definition 1. $S_i := \{(x_i, y_i) | x_i \in [x_i^l, x_i^r]$ and $y_i \in [1 - x_i, \bar{S}_i(x_i)]$ for each $x_i \in [x_i^l, x_i^r]\}$, *where* $\bar{S}_i(x_i)$ *is a single-valued function such that* $\bar{S}_i(x_i^l) = 1 - x_i^l$ *and* $\bar{S}_i(x_i) \in [1 - x_i, 1]$ for each $x_i \in (x_i^l, x_i^r]$.

According to Definition 1, the set of all possible experiments, S^E, is S_i with $x_i^l = 0$, $x_i^r = 1$, and $\bar{S}_i(x_i) = 1$ for $x_i \in [0,1]$. Note that $\bar{S}_i(x_i)$ determines the shape of S_i. I focus on a certain set of S_i by assuming the properties of $\bar{S}_i(x_i)$ as follows.

Assumption 1. $\bar{S}_i(x_i)$ *is continuous on* $[x_i^l, x_i^r]$*, and it satisfies the following:*

(*a*) *for any* $x_i', x_i'' \in [x_i^l, x_i^r]$ *and* $\alpha \in (0,1)$*,* $\bar{S}_i(\alpha x_i' + (1-\alpha)x_i'') \geq \alpha \bar{S}_i(x_i') + (1-\alpha)\bar{S}_i(x_i'')$*,*

(*b*) *for* $x_i' \leq x_i''$*,* $\bar{S}_i(x_i') \geq \bar{S}_i(x_i'')$*,*

(*c*) *there exist some* $x_i \in [x_i^l, x_i^r]$ *such that* $p\bar{S}_i(x_i) > (1-p)(1-x_i)$*,*

(*d*) *there is no subinterval* $(l, h) \subset [x_i^l, x_i^r]$ *on which* $\bar{S}_i(x_i)$ *is differentiable everywhere and* $\frac{\partial \bar{S}_i(x_i)}{\partial x_i} = -\frac{1-p}{p}$ *for all* $x_i \in (l, h)$*, where* $l \neq h$*.*

The continuity of $\bar{S}_i(x_i)$ guarantees S_i to be a *closed* set. On the one hand, the concave $\bar{S}_i(x_i)$ guarantees S_i to be a *convex* set: for any $(x_i', y_i'), (x_i'', y_i'') \in S_i$ and $\alpha \in (0,1)$, $\alpha x_i' + (1-\alpha)x_i'' \in (x_i^l, x_i^r)$; $\alpha y_i' + (1-\alpha)y_i'' \in (1 - (\alpha x_i' + (1-\alpha)x_i''), \alpha\bar{S}_i(x_i') + (1-\alpha)\bar{S}_i(x_i'')] \subset [1 - (\alpha x_i' + (1-\alpha)x_i''), \bar{S}_i(\alpha x_i' + (1-\alpha)x_i'')]$ by Assumption 1.(a). On the other hand, at the boundary of S_i, the decreasing $\bar{S}_i(x_i)$ captures the usual property of an experiment: any attempt to decrease "Type II error" (the false negative rate, $1 - x_i$) increases "Type I error" (the false positive rate, $1 - y_i$). S_i' in Figure 1 is an example which satisfies Assumption 1.(a) and (b). Note that $\bar{S}_i'(x_i)$ is a continuous function on $[x_i'^l, x_i'^r]$, and it is concave and weakly

[8] For example, $\pi = (1,1)$ is qualitatively the same as $\pi' = (0,0)$; simply $\pi = (1,1)$ is the mirror image of $\pi' = (0,0)$ with respect to the line segment, $y = 1 - x$. It is worth noting that this restriction imposes a "meaning" to each outcome. Under the restriction, $y \geq 1 - x$, $\mu(G|g) \geq \mu(G|b)$, where $\mu(\cdot|j)$ is the posterior beliefs over the state space when outcome j is realized for $j = g, b$: the principal's posterior belief that the state is G is higher after observing g than b. That is, g is the good news and b is the bad news as in Milgrom (1981).

decreasing in $x_i \in [x_i'^l, x_i'^r]$. S_i' is indeed a convex and closed subset in the upper-right triangle of the unit square in $\mathbb{R}^2$. Assumption 1.(c) simply excludes some trivial cases. Assumption 1.(d) says that any part of $\bar{S}_i(x_i)$ cannot be a linear function with a slope of $-\frac{1-p}{p}$; this assumption is only for expositional convenience as will be clear later.

Lastly, I assume that $S_2 \subset S_1$. That is, different types of the agent have different S_i in the sense that type θ_1 can conduct any experiment that is feasible for type θ_2 while type θ_2 cannot. More specifically, I assume that $\pi_2 \in S_2$ is in the interior of S_1 unless π_2 is on the line segment, $y = 1 - x$.

Assumption 2. $S_2 \subset S_1$. *Moreover, let* $\underline{S}_2 := \{(x_2, y_2) \in S_2 | y_2 = 1 - x_2\}$. *Then, for every* $\pi_2 = (x_2, y_2) \in S_2 \backslash \underline{S}_2$, *there exists some* $\epsilon > 0$ *such that*

$$\{(x, y) \in [0,1]^2 | \sqrt{(x_2 - x)^2 + (y_2 - y)^2} < \epsilon\} \subset S_1.$$

Assumption 2 simply excludes cases in which S_1 and S_2 share their boundaries except the southwest one, a line segment of $y = 1 - x$. Figure 2 shows an example that satisfies Assumption 2: $S_1 = S^E$, and S_2 is the smaller triangle that is nested in S_1. In this example, type θ_1 can conduct all possible experiments, and type θ_2 only can conduct experiments in the smaller triangle. Furthermore, for any $\pi_2 \in S_2 \backslash \underline{S}_2$, there exists an ϵ-ball around it which is completely contained in S_1.

I assume that the principal can observe both the experiment that the agent conducts *and* the experimental outcomes. That is, the principal knows the structure of the experiment chosen by the agent ($\{\pi_i(\cdot|\omega)\}_{\omega \in \{G,B\}}$, the probability distributions over the outcome space, $\{g, b\}$, conditioning on the true state). In addition, the principal also observes the outcomes of an experiment. Observing an experiment and an experimental outcome, the principal updates her beliefs about the true state by using Bayes' rule.

The timeline of the interaction between the principal and the agent is as follows: (1) the principal moves first and commits to a decision rule which she will follow later to make her decision; (2) the agent observes the principal's decision rule and privately learns his type (i.e., the set of feasible experiments for him); (3) the agent conducts an experiment and an outcome of that experiment is realized; (4) finally, the principal makes a decision according to the decision rule to which she commits based on what she observed, the experiment chosen by the agent and an outcome of that experiment.

The principal's decision rule specifies what action she will take when she observes an experiment and each outcome of the experiment. I allow the principal to randomize over her action space, $\mathcal{A} = \{A, D\}$. I denote by A_j^k the probability that the principal takes action A when outcome j is realized from an experiment, $\pi^k = (x^k, y^k)$, for $j = g, b$. Accordingly

(A_g^k, A_b^k) is a pair of probabilities that the principal takes action A when the outcome is either g or b given an experiment, π^k. Then a decision rule of the principal specifies (A_g^k, A_b^k) to every experiment that is *possibly feasible* for the agent (every feasible experiment for either type θ_1 or θ_2: every $\pi^k \in S_1 \cup S_2 = S_1$). Formally, a decision rule is $M : S_1 \times \{g, b\} \to \Delta\{A, D\}$. A simple example of the principal's decision rule is as follows: for all $\pi^k \in S_1$, $(A_g^k, A_b^k) = (1, 0)$. This decision rule simply says that, for any experiment, the principal will take action A with probability 1 if she observes outcome g and take action D with probability 1 otherwise.

Observing the principal's decision rule, the agent learns his type and conducts an experiment which is feasible for him. Thus, a strategy of each type of the agent is simply $s(\theta_i) : \mathcal{M} \to S_i$, where $\mathcal{M}$ is the set of all possible decision rules of the principal.[9]

2.1 The Principal's Problem

In this subsection, I formally state the principal's problem. Suppose that the principal commits to a decision rule, $M \in \mathcal{M}$. Given this decision rule, M, the agent chooses an experiment in S_i after learning his type. That is, each type chooses $\pi_i'(M, \theta_i) = (x_i', y_i') \in S_i$ such that

$$\pi_i'(M, \theta_i) \in \underset{\pi_i^k \in S_i}{\arg\max}\ EU^A(\pi_i^k | M, \theta_i),$$

where $EU^A(\pi_i^k | M, \theta_i)$ is the (*ex ante*) expected payoff that type θ_i can get by conducting an experiment, $\pi_i^k \in S_i$. Recall that the agent gets 0 for action D and 1 for action A. Thus $EU^A(\pi_i^k | M, \theta_i)$ is simply the expected probability that the principal takes action A *before* the realization of the experimental outcomes. More precisely, given a decision rule M,

$$\begin{aligned} EU^A(\pi_i^k | M, \theta_i) &= A_g^k P(g|\pi_i^k) + A_b^k P(b|\pi_i^k), \\ &= A_g^k (p y_i^k + (1-p)(1 - x_i^k)) + A_b^k (p(1 - y_i^k) + (1-p) x_i^k). \end{aligned}$$

It is worth noting that $P(g|\pi_i^k)$ is the probability that π_i^k generates the positive outcome (g) and $P(b|\pi_i^k)$ is the probability that π_i^k generates the negative outcome (b).

Now suppose that type θ_i conducts $\pi_i'(M, \theta_i)$ and outcome j is realized for $j = g, b$. As the principal follows the decision rule, M, she takes action A with probability A_j' as specified in the decision rule, M, if she observes outcome j from $\pi_i'(M, \theta_i) = (x_i', y_i')$. Then, following

[9] Later I focus on the direct mechanisms. Then the agent's strategy becomes a report of his type after observing a direct mechanism (or simply conducts an experiment within a "menu offer" provided by the principal). If one takes the game form of this problem seriously, the solution concept is the Subgame Perfect Equilibrium. The principal's choice of a decision rule induces a subgame; in that subgame, the agent learns his type and conducts an experiment; then, the principal behaves according to what she committed to. Lastly, after the agent's equilibrium behavior is described in every subgame, the principal chooses a decision rule which leads her to the subgame that gives her the highest payoff.

the decision rule, the principal obtains the (*ex post*) payoff below:

$$\begin{aligned} EU^P(M|\pi_i', j) &= A_j'\{U^P(A,G)\mu(G|j,\pi_i') + U^P(A,B)\mu(B|j,\pi_i')\} \\ &\quad +(1-A_j')\{U^P(D,G)\cdot\mu(G|j,\pi_i') + U^P(D,B)\mu(B|j,\pi_i')\}, \\ &= A_j'\cdot\mu(G|j,\pi_i') + (1-A_j')\cdot\mu(B|j,\pi_i'),^{10} \\ &= A_j'\cdot\mu(G|j,\pi_i') + (1-A_j')\cdot(1-\mu(G|j,\pi_i')), \end{aligned}$$

where $EU^P(\cdot)$ denotes the principal's expected payoff and $\mu(G|j,\pi_i')$ is the principal's posterior belief after observing outcome j of π_i'. Then, the principal's *interim* expected payoff before observing an outcome from π_i' is simply

$$\begin{aligned} EU^P(M|\pi_i') &= P(g|\pi_i')\cdot EU^P(M|\pi_i', g) + P(b|\pi_i')\cdot EU^P(M|\pi_i', b), \\ &= (1-p) - (1-2p)A_b' + (A_g' - A_b')(py_i' - (1-p)(1-x_i')).^{11} \end{aligned}$$

Based on the principal's preference represented by $EU^P(M|\pi_i^k)$, I introduce two notions, *ex post optimality* and *favorite experiments*, where ex post optimality relates to the principal's preference over the action space while favorite experiments relate to that over the set of feasible experiments.

Definition 2. *(Ex post Optimality) Given $\pi_i^k \in S_i$, (A_g^k, A_b^k) is* ex post optimal *if (A_g^k, A_b^k) maximizes the principal's* (interim) *expected payoff.*

Ex post optimality simply means that the principal needs to make an optimal decision given an experiment. An experiment, $\pi_i^k \in S_i$, generates either outcome g or b which induces a posterior belief, $\mu(G|j,\pi_i^k)$ for $j = g$ or b. If $\mu(G|j,\pi_i^k) \geq 1/2$, action A is optimal; otherwise, action D is optimal. For example, given $\pi^f = (1,1)$, outcome g induces $\mu(G|g,\pi^f) = 1$ and outcome b does $\mu(G|b,\pi^f) = 0$; thus $(A_g^f, A_b^f) = (1,0)$ is ex post optimal. Note that there are experiments that fail to convince the principal to take action A. These experiments cannot induce $\mu(G|j,\pi_i) \geq 1/2$ with any of its outcomes because they are *uninformative*. I call such experiments *non-convincing experiments*. As an extreme example, $\pi^u = (1,0)$ generates outcome b with probability 1 regardless of the true state. Thus, it is fully uninformative; it always induces $\mu(G|b,\pi^u) = p < 1/2$ and fails to convince the principal to take action A. An experiment $\pi_i^k = (x_i^k, y_i^k)$ is not convincing if and only if either $py_i^k < (1-p)(1-x_i^k)$ or $\pi_i^k = (1,0)$.[12] Note that any non-convincing experiment induces $\mu(G|j,\pi_i^k) < 1/2$ for $j = g, b$

[10]Recall that $U^P(A,G) = U^P(D,B) = 1$ and $U^P(A,B) = U^P(D,G) = 0$.

[11]One can easily get the expression in the second line by substituting the explicit form of $\mu(G|\pi_i', j)$ for $j = g, b$.

[12]Note that $\pi^{u'} = (0,1)$ is the other extreme case; it always generates outcome g regardless of the true state; it is also a non-convincing experiment because it satisfies $p\cdot 1 < (1-p)(1-0)$.

(e.g., suppose $p = P(G) = 0.3$, and consider $\pi_i^k = (0.8, 0.2)$.) Thus, for any non-convincing experiment π_i^k, $(A_g^k, A_b^k) = (0,0)$ is optimal. Any experiment that is *not* non-convincing is convincing, and it induces either $\mu(G|g, \pi_i^k) \geq 1/2$ or $\mu(G|b, \pi_i^k) < 1/2$. Thus for any convincing experiment π_i^k, $(A_g^k, A_b^k) = (1,0)$ is optimal. It is also worth noting that S_i must have some convincing experiments under Assumption 1.(c).[13]

Let $\bar{M}$ denote a simple decision rule under which the principal makes the ex post optimal decisions for every $\pi_i^k \in S_i$. More precisely, let $\bar{M}$ map every convincing π_i^k to $(A_g^k, A_b^k) = (1,0)$ and every non-convincing π_i^k to $(A_g^k, A_b^k) = (0,0)$.

Definition 3. *(Favorite Experiment) The principal's favorite experiment is denoted by $\hat{\pi}_i \in S_i$, and it maximizes the principal's* (interim) *expected payoff under $\bar{M}$, i.e.,*

$$\hat{\pi}_i \in \underset{\pi_i^k \in S_i}{\arg\max}\ EU^P(\bar{M}|\pi_i^k).$$

A favorite experiment, $\hat{\pi}_i$, is simply an experiment that the principal most prefers among all experiments in S_i when she always makes the ex post optimal decisions. Assuming that the principal employs $\bar{M}$,

$$EU^P(\bar{M}|\pi_i^k) = \max\{1-p,\ py_i^k + (1-p)x_i^k\}.$$

For a non-convincing experiment, $(A_g^k, A_b^k) = (0,0)$, and the principal obtains a constant payoff of $1-p$. For a convincing experiment, $(A_g^k, A_b^k) = (1,0)$, and the principal's payoff is $py_i^k + (1-p)x_i^k$. For any convincing π_i^k, $py_i^k + (1-p)x_i^k \geq 1-p$. Hence, whenever there exists a convincing $\pi_i^k \in S_i$, $\hat{\pi}_i$ is simply a maximizer of $py_i^k + (1-p)x_i^k$.[14]

Figure 3 demonstrates $\hat{\pi}_i \in S_i$ for $i = 1,2$. In Figure 3, the experiments below the line, $y = \frac{1-p}{p}(1-x)$, are non-convincing, and the experiments on or above the line are convincing (except for $\pi_i^k = (1,0)$). Note that the principal's payoff increases as the indifference curve moves toward the *northeast*; as x_i and y_i increase, π_i becomes more informative in the sense of Blackwell (1953)[15]; the principal makes a better decision with a more informative π_i; thus, the principal's payoff increases. Hence the principal's favorite experiment in S_1 is $\hat{\pi}_1 = (1,1)$, and that in S_2 is $\hat{\pi}_2 = (0.8, 0.8)$ in Figure 3.

Note that there exists a unique $\hat{\pi}_i$ in S_i under Assumption 1. The slope of the principal's indifference curve is $-\frac{1-p}{p}$. If $\bar{S}_i(x_i)$ has any part which is a linear function with a slope of

[13] If both types only have non-convincing experiments in S_i, the principal's problem has a simple solution, which is always to take action D with probability one for any outcome from every experiment.

[14] It is worth noting that $\hat{\pi}_i$ can be defined under other classes of decision rules other than $\bar{M}$. Note that, holding $A_g^k > A_b^k$ fixed, $EU^P(M|\pi_i^k)$ is proportional to $py_i^k + (1-p)x_i^k$ which is maximized at $\hat{\pi}_i$. Hence, $\hat{\pi}_i$ is also an experiment that the principal most prefers under a class of decision rules such that $A_g^k = A_g > A_b^k = A_b$ for every π_i^k.

[15] Weber (2010) shows that $\pi_i = (x_i, y_i)$ is at least informative as $\pi_i' = (x_i', y_i')$ in the sense of Blackwell if and only if $x_i \geq x_i'$ and $y_i \geq y_i'$.

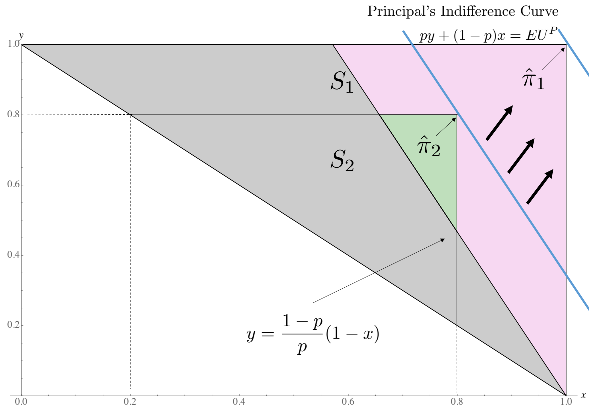


Figure 3: Favorite Experiments

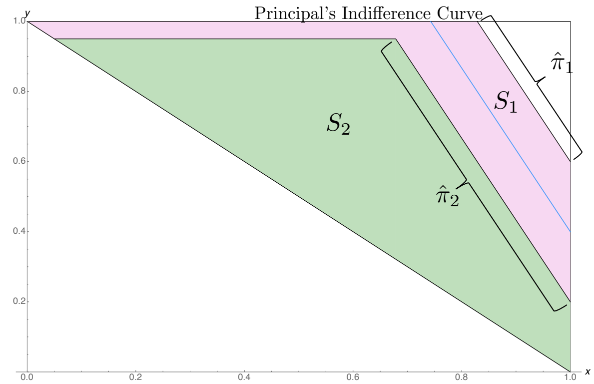


Figure 4: Multiple $\hat{\pi}_i$

$-\frac{1-p}{p}$, any experiment on that part will be a favorite experiment as in Figure 4. Assumption 1.(d) guarantees that we do not have such a case. The uniqueness of $\hat{\pi}_i$ is not necessary but makes it easier to state the results succinctly. I discuss how to deal with the multiple $\hat{\pi}_i$ in the Appendix.[16]

Returning to the discussion on the principal's problem, the principal's *ex ante* payoff from a decision rule, M, is

$$EU^P(M) = t \cdot EU^P(M|\pi_1') + (1-t) \cdot EU^P(M|\pi_2'),$$

and the principal's problem becomes

$$\max_{M \in \mathcal{M}} \; t \cdot EU^P(M|\pi_1') + (1-t) \cdot EU^P(M|\pi_2'),$$
$$s.t. \quad \pi_i' \in \arg\max_{\pi_i^k \in S_i} EU^A(\pi_i^k|M, \theta_i) \quad \text{for } i = 1, 2.$$

The problem seems complicated since a decision rule is a mapping from $S_1 \times \{g, b\}$ to $\Delta\{A, D\}$. However, *the revelation principle* can be applied to this problem: it can be shown that an outcome can be achieved via a decision rule $M \in \mathcal{M}$ if and only if the same outcome can be achieved via a menu offer, $M' = \{(\pi_1, A_g^1, A_b^1), (\pi_2, A_g^2, A_b^2)\}$, such that each triple, (π_i, A_g^i, A_b^i), is destined for type θ_i. In the rest of this paper, a decision rule means a *menu offer* unless I mention that I use the original form of the decision rule.

A menu offer, M', has two triples where each triple consists of an experiment, π_i, and two probabilities, A_g^i and A_b^i. For example, a triple, (π_i, A_g^i, A_b^i), means that, if the agent conducts π_i, the principal will take action A with probability A_j^i when the outcome of π_i is j for $j = g, b$. The revelation principle also implies that it suffices to only consider incentive

[16] There are two possible cases: (i) when we can construct a first-best decision rule by exploiting the multiplicity of $\hat{\pi}_i$ or (ii) when we can choose a specific $(\hat{\pi}_1, \hat{\pi}_2)$ to which we can apply the results in this paper.

compatible menu offers; i.e., type θ_i prefers (π_i, A_g^i, A_b^i) to (π_j, A_g^j, A_b^j) for $i \neq j$. Hence we can write the principal's problem as follows:

$$\begin{aligned}
&\max_{m_1,m_2} tEU^P(m_1) + (1-t)EU^P(m_2) \\
s.t.\quad & EU^A(m_1|\theta_1) \geq EU^A(m_2|\theta_1) \;\; \text{(IC constraint for type } \theta_1\text{)}, \\
& EU^A(m_2|\theta_2) \geq EU^A(m_1|\theta_2) \;\; \text{(IC constraint for type } \theta_2\text{)}, \\
& \pi_1 \in S_1, \;\pi_2 \in S_2,
\end{aligned}$$

where $m_i = (\pi_i, A_g^i, A_b^i)$ for $i = 1, 2$, and the IC constraint for type θ_2 is valid only if $\pi_1 \in S_2$.

Note that there are cases in which the IC constraint for type θ_2 can be ignored. If π_1 in m_1 is chosen so that it does not belong to S_2 (i.e., $\pi_1 \in S_1 \backslash S_2$), type θ_2 only has one option which is to conduct $\pi_2 \in S_2$. Thus the IC constraint for type θ_2 is redundant in this case. In short, depending on which menu offer we are considering, the IC constraint for type θ_2 can be "on or off" in the maximization problem. The set of possible menu offers can be divided into two classes: one that assigns $\pi_1 \in S_1 \backslash S_2$ and $\pi_2 \in S_2$ and the other that assigns $\pi_1 \in S_2$ and $\pi_2 \in S_2$. The following proposition tells us that it is innocuous to restrict our attention to the menu offers that assign $\pi_1 \in S_1 \backslash S_2$ and $\pi_2 \in S_2$.

Proposition 1. *There always exists an incentive compatible menu offer with $\pi_1 \in S_1 \backslash S_2$ and $\pi_2 \in S_2$ that gives the principal a higher payoff than every menu offer with $\pi_1 \in S_2$ and $\pi_2 \in S_2$.*

The proof of Proposition 1 is relegated to the Appendix. A rough intuition is that, if a menu offer assigns $\pi_1 \in S_2$ and $\pi_2 \in S_2$, the optimal choice for both π_1 and π_2 is $\hat{\pi}_2$; Assumption 2 guarantees that a set of experiments around $\hat{\pi}_2$ is feasible for type θ_1; then, I can always find $\pi_1' = (\hat{x}_2 + \epsilon_x, \hat{y}_2 + \epsilon_y) \in S_1 \backslash S_2$ that (i) the principal strictly prefers over $\hat{\pi}_2$ and (ii) gives as much payoff to type θ_1 as $\hat{\pi}_2$.

By Proposition 1, we can exclude the menu offers with $\pi_1 \in S_2$ and $\pi_2 \in S_2$. Then, the IC constraint for type θ_2 can be ignored, and the principal's problem is simplified as follows:

$$\begin{aligned}
&\max_{m_1,m_2} tEU^P(m_1) + (1-t)EU^P(m_2) \\
s.t.\quad & EU^A(m_1|\theta_1) \geq EU^A(m_2|\theta_1) \;\; \text{(IC constraint for type } \theta_1\text{)}, \\
& \pi_1 \in S_1 \backslash S_2, \;\pi_2 \in S_2,
\end{aligned}$$

where $m_i = (\pi_i, A_g^i, A_b^i)$ for $i = 1, 2$.

2.2 The Benchmark: the First-best Outcome

In this subsection, I establish the benchmark: the best outcome that the principal can achieve when there is no uncertainty about the agent's type. Here I use the *original form of the decision rule* instead of a menu offer to discuss the first-best outcome.

Suppose that the principal can observe the agent's type. Then the principal can exert her commitment power contingent on the types of the agent. If the agent's type is θ_i, the principal simply looks over all experiments which are feasible for that type (every $\pi_i^k \in S_i$), identifies $\hat{\pi}_i$ in S_i, and demands the agent to conduct it. The principal can do this by saying that she will take action D regardless of experimental outcomes if the agent conducts any $\pi_i^k \neq \hat{\pi}_i$ (i.e., by setting (A_g^k, A_b^k) to $(0,0)$ for any $\pi_i^k \neq \hat{\pi}_i$). Then, the agent does not have any better alternative than conducting $\hat{\pi}_i$ among feasible experiments for him. Thus, the agent will conduct $\hat{\pi}_i$, and the principal can simply make an optimal decision based on an experimental outcome of $\hat{\pi}_i$. In short, if the principal can observe the type of the agent, the principal can (i) assign her *favorite experiment* to the agent and (ii) make the *ex post optimal decisions* based on the experimental outcomes of her favorite experiment. This is the first-best outcome and denoted by $[(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)]$.[17] For example, when S_1 and S_2 are given as in Figure 3, the first-best outcome is simply $[\big((1,1),1,0\big), \big((0.8,0.8),1,0\big)]$.

3 Optimal Decision Rules

In this section, I characterize the principal's optimal decision rules when she cannot observe the agent's type. A reasonable question to start with is "Is it possible to achieve the first-best outcome even when the principal does not know the agent's type?" The answer is that it depends on the properties of two favorite experiments, $\hat{\pi}_1$ and $\hat{\pi}_2$.

Remark 1. *The principal can achieve the first-best outcome if and only if $\hat{\pi}_1$ generates the positive outcome more frequently than $\hat{\pi}_2$ (that is, $p\hat{y}_1 + (1-p)(1-\hat{x}_1) \geq p\hat{y}_2 + (1-p)(1-\hat{x}_2)$, where $\hat{\pi}_1 = (\hat{x}_1, \hat{y}_1)$ and $\hat{\pi}_2 = (\hat{x}_2, \hat{y}_2)$).*

The following example in Figure 5 is helpful to understand Remark 1 above. In the example, I assume $p = P(G) = 0.3$. Suppose that each type of the agent has S_1 and S_2, respectively. Then $\hat{\pi}_1 = (1,1)$ and $\hat{\pi}_2 = (0.8, 0.8)$. Now consider the following menu offer, $\{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)\}$. Given this menu offer, the agent's payoff is equal to the probability that $\hat{\pi}_i$ generates the positive outcome. It is easy to see that type θ_1 would conduct $\hat{\pi}_2$ rather

[17] To avoid possible confusions, I use $[(\cdot),(\cdot)]$ for an outcome and $\{(\cdot),(\cdot)\}$ for a menu offer. An outcome, $[(\pi_1, A_g^1, A_b^1), (\pi_2, A_g^2, A_b^2)]$, means that, if the agent's type is θ_i, then that type conducts π_i and the principal takes action A with probability A_j^i if the outcome of π_i is j, where $j = g, b$.

than $\hat{\pi}_1$ as the former has a higher probability to generate the positive outcome than the latter. However, if type θ_2 has S_2', $\hat{\pi}_2' = (0.9, 0.7)$ and it generates the positive outcome less frequently than $\hat{\pi}_1$. Thus, $\{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2', 1, 0)\}$ is incentive compatible and the principal can achieve the first-best outcome.

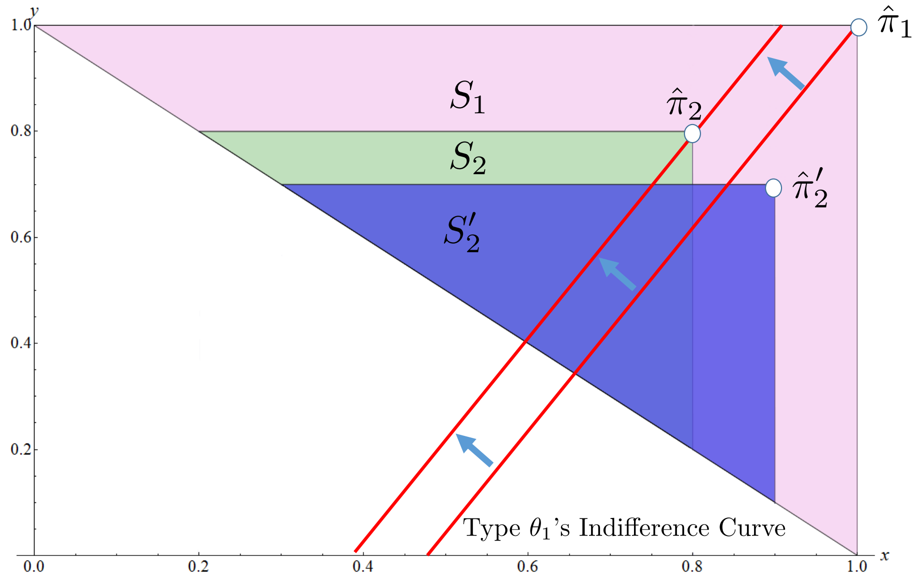


Figure 5: Example

3.1 Optimal Decision Rules

From now on, I focus only on the cases in which the principal cannot achieve the first-best outcome. That is, I assume that $p\hat{y}_1 + (1-p)(1-\hat{x}_1) < p\hat{y}_2 + (1-p)(1-\hat{x}_2)$ for the rest of the paper. I start by curtailing the set of experiments, S_i.

3.1.1 Curtailing Sets of Experiments

First I can exclude the non-convincing experiments. Recall that an experiment, $\pi_i = (x_i, y_i)$, is non-convincing if either $py_i < (1-p)(1-x_i)$ or $\pi_i = (1, 0)$. These non-convincing experiments are not informative. Hence, the principal who wants an informative experiment does not assign these non-convincing experiments under an optimal decision rule. Figure 6 shows how non-convincing experiments in S_i are excluded: in both examples, $\pi_i = (1, 0) \notin S_i$ and all experiments below the line, $y = \frac{1-p}{p}(1-x)$, are excluded. I denote by Π_i the set of convincing experiments in S_i (the shaded areas in Figure 6).

From now on, I only consider Π_i as non-convincing experiments would not be assigned under an optimal decision rule. Note that given any $\pi_i \in \Pi_i$, $py_i - (1-p)(1-x_i) \geq 0$. Thus,

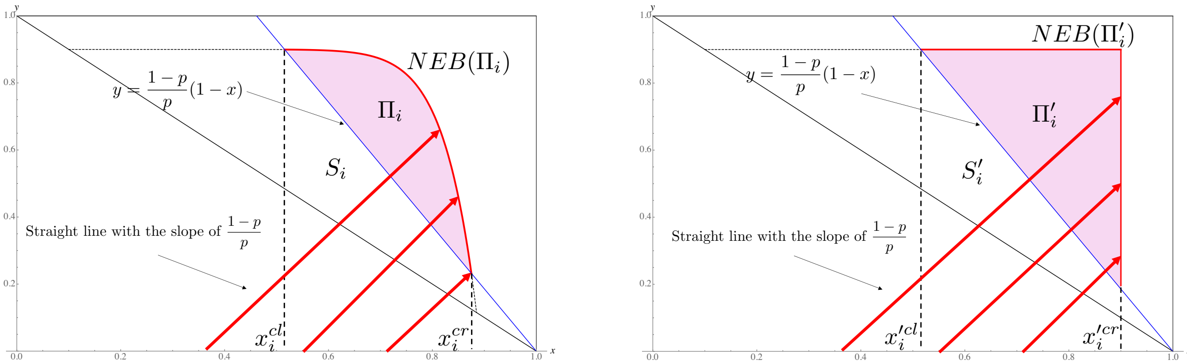


Figure 6: Π_i and $NEB(\Pi_i)$

the principal's *ex ante* payoff increases in A_g^i but decreases in A_b^i for $i = 1, 2$. This helps us to understand more about the optimal action-probability pairs, (A_g^i, A_b^i) for $i = 1, 2$.

Remark 2. *For every* $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, *(i)* $A_g^1 = 1$ *and* $A_b^2 = 0$ *are optimal, and (ii)* $A_b^1 < 1$ *and* $A_g^2 > 0$ *are optimal.*

The first result is immediate. Given any $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, the IC constraint is

$$\begin{aligned} &A_g^1(py_1 + (1-p)(1-x_1)) + A_b^1(1 - (py_1 + (1-p)(1-x_1))) \\ &\qquad \geq A_g^2(py_2 + (1-p)(1-x_2)) + A_b^2(1 - (py_2 + (1-p)(1-x_2))). \end{aligned}$$

As the left-hand side of the IC constraint increases in A_g^1 and the right-hand side of it does in A_b^2, the principal can be better off by increasing A_g^1 and decreasing A_b^2 *without* violating the IC constraint. Hence, $A_g^1 = 1$ and $A_b^2 = 0$ are optimal. Given that $A_g^1 = 1$ and $A_b^2 = 0$, the IC constraint reduces to

$$A_b^1 + (1 - A_b^1)(py_1 + (1-p)(1-x_1)) \geq A_g^2(py_2 + (1-p)(1-x_2)).$$

Note that if either $A_b^1 = 1$ or $A_g^2 = 0$, the IC constraint is vacuously satisfied with strict inequality. Thus, the principal can be better off by decreasing A_b^1 from 1 or increasing A_g^2 from 0 without violating the IC constraint.[18]

Now recall that the principal prefers one experiment to the other if the former is in the *"northeast"* of the latter. By using this property, I can further reduce the possible candidates for experiments that are assigned under an optimal decision rule. I denote the *"northeast boundary"* of Π_i by $NEB(\Pi_i)$.

Definition 4. *An experiment,* $\pi_i = (x_i, y_i) \in \Pi_i$, *is on the* $NEB(\Pi_i)$ *if, for any* $\epsilon > 0$, $(x_i + p\epsilon, y_i + (1-p)\epsilon)$ *is not in* Π_i.

[18]The detailed proof for the second result in Remark 2 is relegated to the Appendix.

Figure 6 visualizes how to find $NEB(\Pi_i)$. It is straightforward to tell whether an experiment is on the $NEB(\Pi_i)$; (1) choose an experiment in Π_i, (2) draw a straight line that has a slope of $\frac{1-p}{p}$ from the chosen experiment to the northeast of that experiment, (3) if no part of this line segment except the chosen experiment belongs to Π_i, the chosen experiment is on the $NEB(\Pi_i)$; otherwise, it is not.

With the precise definition of $NEB(\Pi_i)$, I state the following lemma.

Lemma 1. *An optimal decision rule assigns an experiment on $NEB(\Pi_i)$.*

The intuition of Lemma 1 is simple. Note that, given any decision rule, type θ_1's payoff depends on the probability that π_i generates the positive outcome, $py_i + (1-p)(1-x_i)$. Then type θ_1 is indifferent among experiments on a line through $\pi_i = (x_i, y_i)$ with a slope of $\frac{1-p}{p}$ as they have the same probability of generating the positive outcome. Hence, the principal can choose any experiment on the line segment without altering the IC constraint of type θ_1. Then, the principal would choose the one at the most "northeast".

The procedure so far gives us two "lines," $NEB(\Pi_i)$ for $i = 1, 2$, rather than two convex sets. Note that $\bar{S}_i(x_i)$ *dictates* the shape of $NEB(\Pi_i)$. Let $N_i(x_i)$ be the part of $\bar{S}_i(x_i)$ above $y = \frac{1-p}{p}(1-x)$: $N_i(x_i) := \{\bar{S}_i(x_i) | p\bar{S}_i(x_i) \geq (1-p)(1-x_i)\}$. Then, I say $NEB(\Pi_i)$ is non-increasing (or concave) if and only if $N_i(x_i)$ is non-increasing (or concave) in x_i. One caveat here is that $N_i(x_i)$ does not always coincide with $NEB(\Pi_i)$. For example, in the left panel of Figure 6, $N_i(x_i) : [x_i^{cl}, x_i^{cr}] \to [\bar{S}_i(x_i^{cl}), \bar{S}_i(x_i^{cr})]$ exactly describes $NEB(\Pi_i)$. However, in the right panel, $N_i(x_i') = 0.9$ for $x_i' \in [x_i'^{cl}, x_i'^{cr}]$ does not include the vertical line at $x_i'^{cr}$, which is a part of $NEB(\Pi_i')$. Nonetheless, $NEB(\Pi_i')$ is "non-increasing" in the sense that for any $\pi_i' = (x_i', y_i') \in NEB(\Pi_i')$, y_i' does not increase as x_i' increases; it is also "concave" in the sense, that for any $\pi_i', \pi_i'' \in NEB(\Pi_i')$ and $\alpha \in (0, 1)$, $\alpha\pi_i' + (1-\alpha)\pi_i''$ is "below" $NEB(\Pi_i')$.

Under Assumption 1, we have a decreasing and concave $NEB(\Pi_i)$ which is always non-empty. Then, under Assumption 2, we have $NEB(\Pi_1)$ in the northeast of $NEB(\Pi_2)$ as shown in Figure 7. It is worth mentioning that $\hat{\pi}_i$ is always included in $NEB(\Pi_i)$.

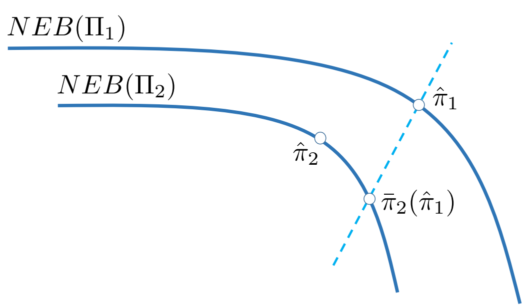


Figure 7: $NEB(\Pi_i)$

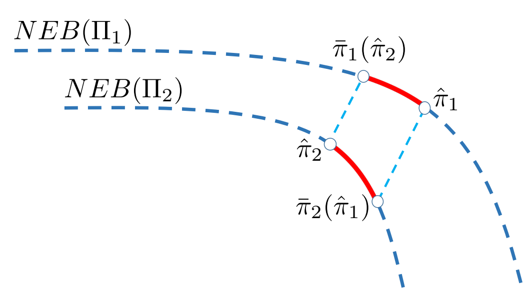


Figure 8: Proposition 2

Given $NEB(\Pi_i)$ for $i = 1, 2$, we can be even more precise about the experiments assigned under an optimal decision rule. I introduce two definitions used to state this next result.

Definition 5. *Given an experiment,* $\pi_i = (x_i, y_i)$*, on* $NEB(\Pi_i)$*,* $\bar{\pi}_j(\pi_i) = (\bar{x}_j, \bar{y}_j)$ *is the experiment on* $NEB(\Pi_j)$ *such that* $py_i + (1-p)(1-x_i) = p\bar{y}_j + (1-p)(1-\bar{x}_j)$ *for* $i \neq j$*.*

Given an experiment, π_i, on $NEB(\Pi_i)$, $\bar{\pi}_j(\pi_i)$ is the experiment on $NEB(\Pi_j)$ that can constitute an *incentive compatible* decision rule along with the given π_i while keeping ex post optimality in both types. For example, consider $\hat{\pi}_1$ and $\bar{\pi}_2(\hat{\pi}_1)$. Geometrically, $\bar{\pi}_2(\hat{\pi}_1)$ is the point on $NEB(\Pi_2)$ which intersects with the straight line drawn from $\hat{\pi}_1$ on $NEB(\Pi_1)$ with a slope of $\frac{1-p}{p}$ as shown in Figure 7. It is easy to see that $\{(\hat{\pi}_1, 1, 0), (\bar{\pi}_2(\hat{\pi}_1), 1, 0)\}$ is incentive compatible.

Definition 6. *Relative locations of experiments,* π_i *and* π_i'*, on* $NEB(\Pi_i)$*:*

(*a*) $\pi_i \succsim \pi_i'$*:* π_i *is in the northwest of or equal to* π_i' *(i.e.,* $y_i \geq y_i'$ *and* $x_i \leq x_i'$*),*

(*b*) $\pi_i \succ \pi_i'$*:* π_i *is in the northwest of* π_i' *(i.e.,* $y_i > y_i'$ *and* $x_i < x_i'$*).*

It is worth noting that, given the optimal action-probability pairs in Remark 2, the agent prefers π_i to π_i' if and only if $\pi_i \succsim \pi_i'$. Furthermore, note that, as we are interested in the cases such that $p\hat{y}_1 + (1-p)(1-\hat{x}_1) < p\hat{y}_2 + (1-p)(1-\hat{x}_2)$, we have $\hat{\pi}_2 \succ \bar{\pi}_2(\hat{\pi}_1)$ (or, equivalently, $\bar{\pi}_1(\hat{\pi}_2) \succ \hat{\pi}_1)$).

With two definitions above, I state Proposition 2 which concludes the "curtailing process."

Proposition 2. *An optimal decision rule assigns* (π_1, π_2) *such that* $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$ *and* $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$*.*

Figure 8 visualizes Proposition 2. It is easy to find $\hat{\pi}_i$ on $NEB(\Pi_i)$. Then each $\hat{\pi}_i$ determines $\bar{\pi}_j(\hat{\pi}_i)$ on $NEB(\Pi_j)$. Proposition 2 says that these four experiments, $\hat{\pi}_1, \hat{\pi}_2, \bar{\pi}_1(\hat{\pi}_2)$ and $\bar{\pi}_2(\hat{\pi}_1)$, confine the candidates for experiments assigned under an optimal decision rule.[19] The intuition behind it is simple: if a decision rule assigns π_i which is "far from" $\hat{\pi}_i$, we can improve the principal's payoff by moving π_i closer to $\hat{\pi}_i$.

3.1.2 Simplifying the Problem

In this subsection, I introduce statements that further simplify our problem.

[19] Note that the existence of $\bar{\pi}_1(\hat{\pi}_2)$ is guaranteed. However, there are cases in which $\bar{\pi}_2(\hat{\pi}_1)$ is excluded and not on the $NEB(\Pi_2)$ as it is a non-convincing experiment. I do not explicitly discuss such cases in this paper. However, in those cases, I only need to replace $\bar{\pi}_2(\hat{\pi}_1)$ with $\underline{\pi}_2$, where $\underline{\pi}_2 = (x_2^{cr}, \bar{S}_2(x_2^{cr}))$. Then, one can easily see that every statement which involves $\bar{\pi}_2(\hat{\pi}_1)$ in this paper (i.e., Propositions 2 or 5) is still valid after replacing $\bar{\pi}_2(\hat{\pi}_1)$ with $\underline{\pi}_2$.

Lemma 2. *The IC constraint for type θ_1 binds under an optimal decision rule given that $p\hat{y}_1 + (1-p)(1-\hat{x}_1) < p\hat{y}_2 + (1-p)(1-\hat{x}_2)$.*

By Lemma 2, we do not need to consider any decision rule with a non-binding IC constraint. Furthermore, any decision rule with (π_1, π_2) such that $py_1+(1-p)(1-x_1) > py_2+(1-p)(1-x_2)$ cannot be optimal since there are no optimal action-probability pairs that make type θ_1's IC constraint bind (i.e., the optimal action-probability pairs in Remark 2).

Lemma 3. *For any decision rule with $(\pi_1, \pi_2) \in NEB(\Pi_1) \times NEB(\Pi_2)$ such that $\pi_2 \succ \bar{\pi}_2(\pi_1)$ (i.e., $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$), it is optimal to give up ex post optimality just for one type.*

The core intuition behind Lemma 3 is the fact that the principal's payoff is linear in A_b^1 and A_g^2 given a pair of two experiments, (π_1, π_2): it is better to put the cost of distortion in the decisions to one type than to distribute the cost to both types.

Finally, lemmas above altogether lead us to Proposition 3 which pins down the optimal action-probability pairs given any pair of experiments, (π_1, π_2), under our consideration and, thus, further reduces the set of candidates for optimal decision rules.

Proposition 3. *Given any $(\pi_1, \pi_2) \in NEB(\Pi_1) \times NEB(\Pi_2)$ such that $py_1+(1-p)(1-x_1) \leq py_2 + (1-p)(1-x_2)$,*

(a) if $\tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$, $M^2(\pi_1, \pi_2) := \{(\pi_1, 1, 0), (\pi_2, A_g^2(\pi_1, \pi_2), 0)\}$ is optimal,

(b) if $\tau(\pi_1, \pi_2) \leq \frac{1-t}{t}$, $M^1(\pi_1, \pi_2) := \{(\pi_1, 1, A_b^1(\pi_1, \pi_2)), (\pi_2, 1, 0)\}$ is optimal ,

where $A_g^2(\pi_1, \pi_2) = \frac{py_1+(1-p)(1-x_1)}{py_2+(1-p)(1-x_2)}$, $A_b^1(\pi_1, \pi_2) = \frac{(py_2+(1-p)(1-x_2))-(py_1+(1-p)(1-x_1))}{1-(py_1+(1-p)(1-x_1))}$, and

$$\tau(\pi_1, \pi_2) = \left(\frac{py_2 + (1-p)(1-x_2)}{py_2 - (1-p)(1-x_2)}\right) \cdot \left(\frac{1 - 2p + py_1 - (1-p)(1-x_1)}{1 - (py_1 + (1-p)(1-x_1))}\right).$$

Remark 2 and Lemma 2 imply that an optimal decision rule assigns (π_1, π_2) such that $py_1 + (1-p)(1-x_1) \leq py_2+(1-p)(1-x_2)$, which is the pair of experiments considered in Proposition 3. Then, Proposition 3 implies that we can focus on either $M^1(\pi_1, \pi_2)$ or $M^2(\pi_1, \pi_2)$ with (π_1, π_2) such that $py_1 + (1-p)(1-x_1) \leq py_2 + (1-p)(1-x_2)$.

First suppose that the principal wants to assign (π_1, π_2) such that $py_1 + (1-p)(1-x_1) = py_2 + (1-p)(1-x_2)$. Then, the principal does not need to sacrifice ex post optimality. Note that, given (π_1, π_2) such that $py_1 + (1-p)(1-x_1) = py_2 + (1-p)(1-x_2)$, the optimal action-probability pairs in each decision rule are ex post optimal: $A_g^2(\pi_1, \pi_2) = 1$ and $A_b^1(\pi_1, \pi_2) = 0$, and $M^1(\pi_1, \pi_2) = M^2(\pi_1, \pi_2) = \{(\pi_1, 1, 0), (\pi_2, 1, 0)\} := \bar{M}(\pi_1, \pi_2)$. Thus, regardless of the values of $\tau(\pi_1, \pi_2)$ and t, it is optimal for the principal to make the ex post optimal decisions.

Now suppose that the principal wants to assign (π_1, π_2) such that $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. Then, she needs to give up the ex post optimal decisions. By Lemma 3, either $M^1 = \{(\pi_1, 1, A_b^1 > 0), (\pi_2, 1, 0)\}$ or $M^2 = \{(\pi_1, 1, 0), (\pi_2, A_g^2 < 1, 0)\}$ is optimal. Then, by Lemma 2, we can immediately obtain A_b^1 and A_g^2 from the binding IC constraint under each decision rule. Thus, both A_b^1 and A_g^2 are the probabilities of making an ex post sub-optimal decision to optimally incentivize type θ_1 to conduct the experiment destined to him, i.e., π_1. Note that both A_b^1 and A_g^2 are functions of (π_1, π_2) and are in $(0, 1)$ given that $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. By comparing the principal's two different ex ante payoffs under two decision rules, the sufficient condition for each decision rule to payoff-dominate the other is immediate. Intuitively, if $t = P(\theta_1)$ is relatively high so that $\tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$, $M^2(\pi_1, \pi_2)$ is optimal, i.e., it is better to sacrifice the ex post optimal decisions for type θ_2; otherwise, $M^1(\pi_1, \pi_2)$ is optimal.

3.1.3 Optimal Decision Rules: Favorite Experiments vs. Ex post Optimality

As we focus on the cases in which the first-best outcome is not feasible, the principal needs to give up either the *favorite experiments* or/and *ex post optimality.*[20] I focus on two kinds of optimal decision rules which possess at least one of two desirable properties: (i) decision rule that assigns the *favorite experiments* at the cost of giving up *ex post optimality* and (ii) decision rule that achieves *ex post optimality* at the cost of giving up the *favorite experiments.* I characterize sufficient conditions under which each decision rule is optimal.

I first introduce two special experiments which relate to the sufficient conditions.

Definition 7. *(Two Special Experiments)*

(a) $\tilde{\pi}_i = (\tilde{x}_i, \tilde{y}_i)$ *is the experiment on* $NEB(\Pi_i)$ *which maximizes* $\frac{y_i}{1-x_i}$ *(the positive likelihood ratio):*

$$\tilde{\pi}_i \in \underset{\pi_i \in NEB(\Pi_i)}{\arg\max} \frac{y_i}{1-x_i},$$

(b) $\mathring{\pi}_i = (\mathring{x}_i, \mathring{y}_i)$ *is the experiment on* $NEB(\Pi_i)$ *which minimizes* $\frac{1-y_i}{x_i}$ *(the negative likelihood ratio):*

$$\mathring{\pi}_i \in \underset{\pi_i \in NEB(\Pi_i)}{\arg\min} \frac{1-y_i}{x_i}.$$

[20]It is worth mentioning that the trade-off between the favorite experiments and ex post optimality is meaningful due to the restrictions on S_i: S_i contains only some but *not all* binary experiments. Consider $(\hat{\pi}_i, A_g^i < 1, A_b^i > 0)$. By reflecting the randomness of $(A_g^i < 1, A_b^i > 0)$ in π_i', one might be able to construct π_i' so that $(\pi_i', A_g'^i = 1, A_b'^i = 0)$ gives as much payoff to both the principal and the agent as $(\hat{\pi}_i, A_g^i < 1, A_b^i > 0)$ does. However, the constructed experiment, π_i', might not be in S_i due to the restrictions on S_i. Even if S_i includes all possible *binary* experiments, this kind of construction might require more than two outcomes, and thus, π_i' might not be in S_i.

Note that $\frac{y_i}{1-x_i}$ is the ratio of the frequency of the true positive outcome to that of the false positive outcome; $\frac{1-y_i}{x_i}$ is the ratio of the frequency of the false negative outcome to that of the true negative outcome. It is easy to see that both $\frac{y_i}{1-x_i}$ and $\frac{1-y_i}{x_i}$ are reasonable measures for the "quality" of experiments: given π_i and π_i', π_i Blackwell-dominates π_i' if and only if $y_i \geq y_i'$ and $x_i \geq x_i'$, which implies that $\frac{y_i}{1-x_i} \geq \frac{y_i'}{1-x_i'}$ and $\frac{1-y_i}{x_i} \leq \frac{1-y_i'}{x_i'}$. Then, according to the first measure, $\frac{y_i}{1-x_i}$, $\tilde{\pi}_i$ is the best experiment in S_i as $\tilde{\pi}_i$ has the highest value of $\frac{y_i}{1-x_i}$. If we employ the second measure, $\frac{1-y_i}{x_i}$, $\mathring{\pi}_i$ is the best experiment in S_i as $\mathring{\pi}_i$ has the lowest value of $\frac{1-y_i}{x_i}$. Figure 9 summarizes a geometrical property of $\tilde{\pi}_i$ and $\mathring{\pi}_i$ in $[0,1]^2$.

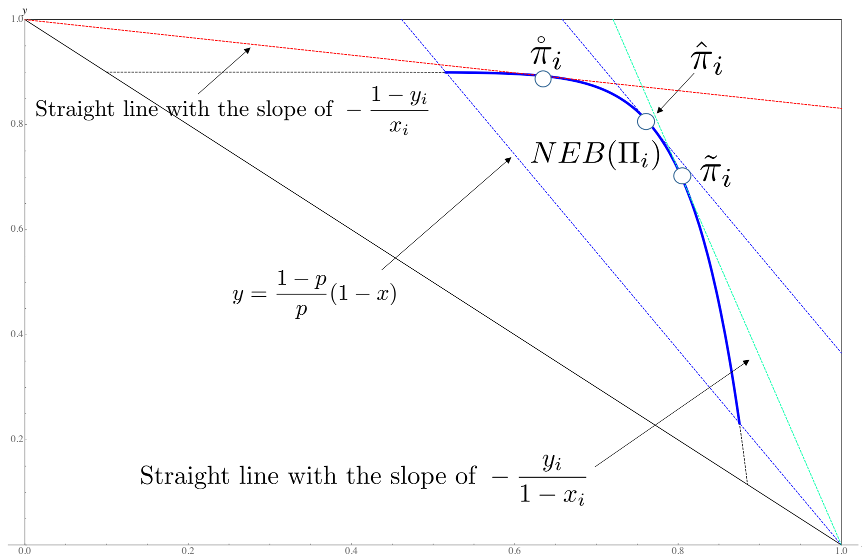


Figure 9: $\tilde{\pi}_i$ and $\mathring{\pi}_i$ on $NEB(\Pi_i)$

Note that $\frac{y_i}{1-x_i}$ is merely the absolute value of the slope of a straight line connecting an experiment $\pi_i = (x_i, y_i)$ and $(1,0)$; $\frac{1-y_i}{x_i}$ is merely the absolute value of the slope of a straight line connecting an experiment $\pi_i = (x_i, y_i)$ and $(0,1)$. Thus, $\tilde{\pi}_i$ is the point that maximizes $|-\frac{y_i}{1-x_i}|$; $\mathring{\pi}_i$ is the point that minimizes $|-\frac{1-y_i}{x_i}|$.

Lastly, the properties of $\bar{S}_i(x_i)$ determine the relative locations of *three special experiments*, $\mathring{\pi}_i$, $\tilde{\pi}_i$, and $\hat{\pi}_i$, as shown in Figure 9.

Lemma 4. *(Relative Locations of $\mathring{\pi}_i, \hat{\pi}_i$, and $\tilde{\pi}_i$)*

(a) $\mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i$*:* $\mathring{y}_i \geq \hat{y}_i \geq \tilde{y}_i$ *(and, thus,* $\mathring{x}_i \leq \hat{x}_i \leq \tilde{x}_i$*),*

(b) if $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$ *and* $N_i(x_i)$ *is twice-differentiable at every* $x_i \in [x_i^{cl}, x_i^{cr}]$, $\mathring{\pi}_i \neq \hat{\pi}_i$ *and* $\tilde{\pi}_i \neq \hat{\pi}_i$*.*

The proof of Lemma 4 is relegated to the online Appendix as the results are evident as shown in Figure 9. Given the relative locations of $\mathring{\pi}_i$, $\hat{\pi}_i$, and $\tilde{\pi}_i$, it is worth discussing the changes in the quality of π_i along $NEB(\Pi_i)$. Note that, as π_i moves along $NEB(\Pi_i)$, the quality improvement in one dimension necessarily involves the quality worsening in the other dimension due to Assumption 1.(b): if x_i increases, y_i decreases, and vice versa. Hence it is impossible to tell whether π_i becomes more informative or not in the sense of Blackwell (1953) unless $N_i(x_i)$ is a horizontal line. However, it is still possible to discuss the quality changes of π_i in terms of three quality measures introduced so far, $\frac{1-y_i}{x_i}$, the principal's preference over $NEB(\Pi_i)$, and $\frac{y_i}{1-x_i}$. Note that as π_i approaches one of three "best" experiments, $\mathring{\pi}_i, \hat{\pi}_i$, and $\tilde{\pi}_i$, the quality of π_i increases in terms of the quality measure that defines the corresponding "best" experiment. For example, as π_i approaches $\tilde{\pi}_i$ from either side of $\tilde{\pi}_i$, $\frac{y_i}{1-x_i}$ increases. I say the quality improvement of π_i is *unambiguous* if π_i becomes closer to *all* of three "best" experiments, $\mathring{\pi}_i$, $\hat{\pi}_i$, and $\tilde{\pi}_i$; if π_i becomes closer to one or two of three "best" experiments, I say the quality improvement of π_i is *ambiguous*.[21]

The unambiguous improvement occurs (i) when π_i approaches $\mathring{\pi}_i$ from the *northwest* (i.e., $\pi_i \to \mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i$) and (ii) when π_i approaches $\tilde{\pi}_i$ from the *southeast* (i.e., $\mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i \leftarrow \pi_i$). The quality improvement is ambiguous in any other cases. More importantly, note that the unambiguous improvement is always *superior* to the ambiguous one in terms of the decrease (or increase) of y_i due to the changes in x_i since $NEB(\Pi_i)$ is concave. For example, as clear in Figure 9, given the same increase in x_i, the unambiguous improvement ($\pi_i \to \mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i$) requires a smaller decrease in $y_i = N_i(x_i)$ than any other ambiguous improvements ($\mathring{\pi}_i \succsim \pi_i \to \hat{\pi}_i \succsim \tilde{\pi}_i$ or $\mathring{\pi}_i \succsim \hat{\pi}_i \succsim \pi_i \to \tilde{\pi}_i$).[22]

Now I state Proposition 4 which provides sufficient conditions for the favorite experiments to be assigned under an optimal decision rule.

Proposition 4. *If (i) $\mathring{\pi}_1 = \hat{\pi}_1$ and (ii) $\tilde{\pi}_2 = \hat{\pi}_2$, an optimal decision rule assigns $\hat{\pi}_1$ to type θ_1 and $\hat{\pi}_2$ to type θ_2.*

An immediate observation is that the sufficient conditions in Proposition 4 cannot be satisfied if $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$ and $N_i(x_i)$ is twice-differentiable as one can see in Lemma 4. That is, a "non-smooth" $NEB(\Pi_i)$ is a necessary condition for the sufficient conditions in Proposition 4 to be satisfied. Figures 10 and 11 show two examples when the sufficient condi-

[21]Weber (2010) proposes the confidence order over the set of binary experiments, and it is related to the unambiguous improvement. Given two experiments, $\pi_i = (x_i, y_i)$ and $\pi_i' = (x_i', y_i')$, a Bayesian decision maker is at least *confident* in π_i' as in π_i if $\frac{y_i'}{1-x_i'} \geq \frac{y_i}{1-x_i}$ and $\frac{1-y_i'}{x_i'} \leq \frac{1-y_i}{x_i}$. Then, according to his terminology, if π_i' is a result of an unambiguous improvement of π_i, a Bayesian decision maker is at least confident in π_i' as in π_i. For example, $\mathring{\pi}_i$ is a result of the unambiguous improvement of any $\pi_i \succsim \mathring{\pi}_i$. Thus, the principal is at least confident in $\mathring{\pi}_i$ as in any $\pi_i \succsim \mathring{\pi}_i$.

[22]In the Appendix, I present a simple exercise with a differentiable $N_i(x_i)$ which demonstrates that a decrease in y_i has a lower upper bound and an increase in y_i has a higher lower bound for the unambiguous improvement than an ambiguous improvement.

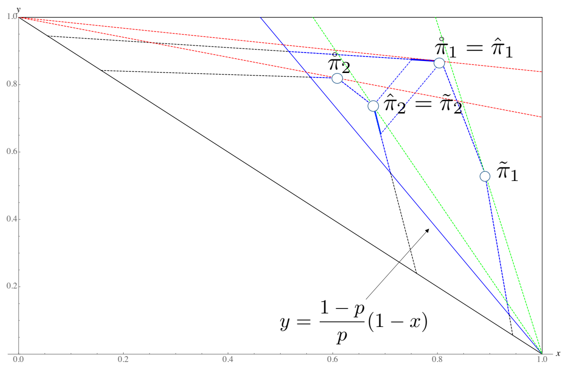


Figure 10: $\mathring{\pi}_1 = \hat{\pi}_1 \succ \tilde{\pi}_1$ and $\mathring{\pi}_2 \succ \hat{\pi}_2 = \tilde{\pi}_2$

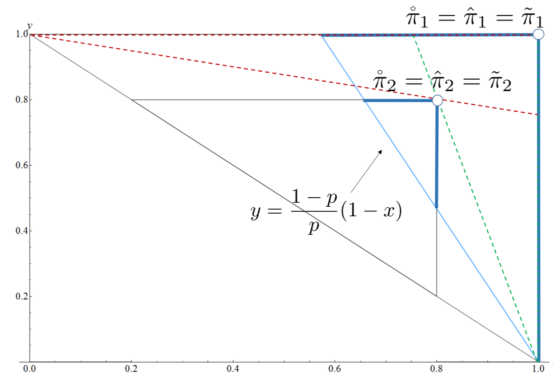


Figure 11: $\mathring{\pi}_i = \hat{\pi}_i = \tilde{\pi}_i$ for $i = 1, 2$

tions in Proposition 4 are satisfied. In Figure 10, $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$ but $N_i(x_i)$ is not differentiable at some x_i; in Figure 11, $N_i(x_i)$ is differentiable but $\{(x_i, N_i(x_i))\} \not\equiv NEB(\Pi_i)$.

Note that, in Figure 11, both S_1 and S_2 take the form of an upper-right triangle. Then, $\mathring{\pi}_i = \hat{\pi}_i = \tilde{\pi}_i$ must hold. Thus, the conditions in Proposition 4 are automatically satisfied. It is worth noting that $\hat{\pi}_i$ Blackwell-dominates every $\pi_i \in S_i$ in this upper-right triangle case.[23] Thus, we have the following Corollary.

Corollary 1. *If $\hat{\pi}_i$ Blackwell-dominates every $\pi_i \in S_i$ for $i = 1, 2$, an optimal decision rule assigns the favorite experiment to each type of the agent.*

When the conditions in Proposition 4 hold, we know what experiments are assigned under an optimal decision rule. For a complete characterization of an optimal decision rule, we need to find the optimal action-probability pairs. We can use Proposition 3 to find these pairs as $(\hat{\pi}_1, \hat{\pi}_2)$ satisfies $p\hat{y}_1 + (1-p)(1-\hat{x}_1) < p\hat{\pi}_2 + (1-p)(1-\hat{x}_2)$. We can simply calculate $\tau(\hat{\pi}_1, \hat{\pi}_2)$ and compare it with $\frac{1-t}{t}$. If t is low enough, $(1-t)/t > \tau(\hat{\pi}_1, \hat{\pi}_2)$ will hold. Then, it is optimal to have $A_b^1(\hat{\pi}_1, \hat{\pi}_2) > 0$. If t is high enough, then it is optimal to have $A_g^2(\hat{\pi}_1, \hat{\pi}_2) < 1$.

To see a rough intuition behind Proposition 4, choose and fix any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$ which then defines $\bar{\pi}_1(\pi_2)$ such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \bar{\pi}_1(\pi_2) \succsim \hat{\pi}_1$. Given the chosen π_2, the principal's choice set for π_1 is $\{\pi_1 | \bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1\}$ (Proposition 2). As π_1 approaches $\bar{\pi}_1(\pi_2)$ from $\bar{\pi}_1(\hat{\pi}_2)$ (i.e., $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \to \bar{\pi}_1(\pi_2) \succsim \hat{\pi}_1$), the principal trades off an increase in x_1 for the associated decrease in y_1. This always benefits the principal because π_1 becomes closer to $\hat{\pi}_1$ while satisfying type θ_1's IC constraint with the ex post optimal decisions. Thus, $\bar{\pi}_1(\pi_2)$ payoff-dominates every π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succ \bar{\pi}_1(\pi_2)$, and type θ_1's IC constraint binds at $\bar{\pi}_1(\pi_2)$ with the principal's ex post optimal decisions by definition of $\bar{\pi}_1(\pi_2)$ (Lemma 2). As

[23] Under Assumption 1, S_i takes the form of an upper-right triangle if and only if $\hat{\pi}_i$ Blackwell-dominates every $\pi_i \in S_i$.

π_1 approaches $\hat{\pi}_1$ from $\bar{\pi}_2(\pi_1)$ (i.e., $\bar{\pi}_1(\pi_2) \succsim \pi_1 \to \hat{\pi}_1$), the principal trades off an increase in x_1 for the associated decrease in y_1 *and distortion in the ex post optimal decisions.* Note that for a given π_2 and any π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1$, $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. Thus, distortion in the ex post optimal decisions is necessary to satisfy type θ_1's IC constraint. This distortion imposes an additional cost for having π_1 closer to $\hat{\pi}_1$. Furthermore, note that the distortion becomes severer as π_1 becomes closer to $\hat{\pi}_1$, i.e., $A_b^1(\pi_1, \pi_2) > 0$ increases and $A_g^2(\pi_1, \pi_2) < 1$ decreases as $\pi_1 \to \hat{\pi}_1$ from $\bar{\pi}_1(\pi_2)$ (Lemma 3 and Proposition 3).[24] Hence, for the principal to give up ex post optimality and prefer π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1$ to $\bar{\pi}_1(\pi_2)$, the benefit from an increase in x_1 must be high enough to compensate for the associated decrease in y_1 and the additional cost of making ex post sub-optimal decisions.

Recall that the unambiguous improvement from the northwest (i.e., $\pi_1 \to \mathring{\pi}_1 \succsim \hat{\pi}_1 \succsim \tilde{\pi}_1$) has a large increase in x_1 *relative to* the associated decrease in y_1. In fact, an increase in x_1 by any unambiguous improvement of π_1 is large enough so that its benefit covers the associated decreases in y_1 and the additional cost imposed by ex post sub-optimality. To see it more clearly, consider $M^1(\pi_1; \pi_2)$ given any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$ as an example. Furthermore, assume that $\{(x_1, N_1(x_1))\} \equiv NEB(\Pi_1)$ and $N_1(x_1)$ is differentiable for expositional convenience. Under $M^1(\pi_1; \pi_2)$, the principal's expected payoff is

$$\begin{aligned} EU^P(M^1(\cdot)) &= (1-t)EU^P(M^1|\pi_2) + t\Big[1 - p + pN_1(x_1) - (1-p)(1-x_1) \\ &\qquad - A_b^1(\pi_1;\pi_2)(1 - 2p + pN_1(x_1) - (1-p)(1-x_1))\Big] \\ &= (1-t)EU^P(M^1|\pi_2) + t\Big[p + (1-p)x_1 - p(1-N_1(x_1)) \\ &\qquad \underbrace{-\left(1 - \frac{1-u_A(\pi_2)}{(1-p)x_1 + p(1-N_1(x_1))}\right)\{(1-p)x_1 - p(1-N_1(x_1))\}}_{:=C_{A_b}(\pi_1;\pi_2),\ \text{additional cost due to } A_b^1>0}\Big], \end{aligned}$$

where $u_A(\pi_2) = py_2 + (1-p)(1-x_2)$. Note that, if there is no additional cost, the principal can always benefit from having π_1 closer to $\hat{\pi}_1$ since its marginal gain is always positive, i.e., $\partial[(1-p)x_1 - p(1-N_1(x_1))]/\partial x_1 = (1-p) + pN_1'(x_1) \geq 0$ for any $\pi_1 \succsim \hat{\pi}_1$. Taking the partial derivative of the additional cost with respect to x_1, we have

$$\frac{\partial C_{A_b}(\cdot)}{\partial x_1} = (1-p) + pN_1'(x_1) - \frac{2p(1-p)(1-u_A(\pi_2))}{[(1-p)x_1 + p(1-N_1(x_1))]^2}\left(N_1'(x_1)x_1 + (1-N_1(x_1))\right).$$

For any $\pi_1 \succsim \mathring{\pi}_1$, $N_1'(x_1)x_1 + (1-N_1(x_1)) \geq 0$ by definition of $\mathring{\pi}_1$.[25] Thus, as far as the quality

[24] If the distortion in the ex post optimal decisions is independent of the choice of π_1, the optimal choice for π_1 is always $\hat{\pi}_1$. It is immediate by the principal's payoff from an experiment π_i', $EU^P(M|\pi_i') = (1-p) - (1-2p)A_b' + (A_g' - A_b')(py_i' - (1-p)(1-x_i'))$, and definition of $\hat{\pi}_i$.

[25] As $\pi_1 \to \mathring{\pi}_1$ from the northwest, $(1 - N_i(x_i))/x_i$ decreases, which implies $\partial[(1 - N_1(x_1))/x_1]/\partial x_1 = [-N_1'(x_1)x_1 - (1 - N_1(x_1))]/x_1^2 \leq 0$ for any $\pi_1 \succsim \mathring{\pi}_1$.

improvement of π_1 is unambiguous (i.e., $\pi_1 \to \mathring{\pi}_1$), the marginal gain, $pN_1'(x_1) + (1-p)$, is always greater than the marginal cost, $\partial C_{A_b}(\cdot)/\partial x_1$.[26]

Given the relative locations of the three "best" experiments (i.e., $\mathring{\pi}_1 \succsim \hat{\pi}_1 \succsim \tilde{\pi}_1$) by Lemma 4, the condition, $\mathring{\pi}_1 = \hat{\pi}_1$, guarantees that having π_1 closer to $\hat{\pi}_1$ from the northwest is always an unambiguous improvement (i.e., $\pi_1 \to \mathring{\pi}_1 = \hat{\pi}_1 \succsim \tilde{\pi}_1$). This, in turn, implies that the benefit from an increase in x_1 is relatively large to compensate for the associated decrease in y_1 and the additional cost due to ex post sub-optimality until $\pi_1 = \hat{\pi}_1 (= \mathring{\pi}_1)$. Thus, given any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$, the principal is willing to assign $\hat{\pi}_1$ to type θ_1 even though the distortion is maximized at $\hat{\pi}_1$. A similar intuition can be found for the other condition, $\hat{\pi}_2 = \tilde{\pi}_2$. Now the condition guarantees that having π_2 closer to $\hat{\pi}_2$ from the southeast is always an unambiguous improvement (i.e., $\mathring{\pi}_2 \succsim \hat{\pi}_2 = \tilde{\pi}_2 \leftarrow \pi_2$). Thus, given any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$, the principal is willing to assign $\hat{\pi}_2$ to type θ_2 at the cost of giving up ex post optimality.

Propositions 4 and 3 fully characterize the optimal decision rule in which the favorite experiments are assigned to each type at the cost of giving up ex post optimality. If type θ_1 is highly likely compared to type θ_2, it is optimal to distort the action-probability pair associated with $\hat{\pi}_2$, and vice versa. Here the principal incentivises type θ_1 to conduct $\hat{\pi}_1$ by increasing the payoff from $\hat{\pi}_1$ (i.e., $A_b^1 > 0$) or decreasing that from $\hat{\pi}_2$ (i.e., $A_g^2 < 1$). It is worth mentioning that this contrasts to the classical "no distortion at the top" result: here distortion is optimally made at type θ_1 (or "high" type) if type θ_2 is highly likely than type θ_1.

Now I provide the sufficient conditions for the decision rule which achieves ex post optimality at the cost of giving up the favorite experiments to be optimal. The sufficient conditions again relate to the relative locations of $\mathring{\pi}_1$ and $\tilde{\pi}_2$.

Proposition 5. *If (i) $\mathring{\pi}_1 \succsim \bar{\pi}_1(\hat{\pi}_2)$ and (ii) $\bar{\pi}_2(\hat{\pi}_1) \succsim \tilde{\pi}_2$, an optimal decision rule achieves ex post optimality; that is, $(A_g^i, A_b^i) = (1, 0)$ for $i = 1, 2$.*

Figure 12 visualizes the conditions in Proposition 5. Recall that, in Proposition 4, we have $\mathring{\pi}_1 = \hat{\pi}_1$ and $\tilde{\pi}_2 = \hat{\pi}_2$. In Proposition 5, $\mathring{\pi}_1$ and $\tilde{\pi}_2$ are "sufficiently far from" $\hat{\pi}_1$ and $\hat{\pi}_2$, respectively.

As we have seen in Proposition 4, the principal is willing to trade off ex post optimality for an unambiguous quality improvement of π_i. However, given (π_1, π_2) specified in Proposition

[26] It is possible to show that $\partial EU^P(M^2(\cdot))/\partial x_1 > 0$ for any $\pi_1 \succsim \mathring{\pi}_1$ with a similar exercise. However, it requires an additional assumption on the likelihood of types (i.e., $\tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$), which is natural because the distortion occurs in the decisions for type θ_2 while the benefit occurs in the decisions for type θ_1. If $\tau(\pi_1, \pi_2) < \frac{1-t}{t}$, then $EU^P(M^1(\cdot)) > EU^P(M^2(\cdot))$ by Proposition 3. Hence, the principal employs $M^1(\cdot)$, and we return to the discussion on $M^1(\cdot)$ above.

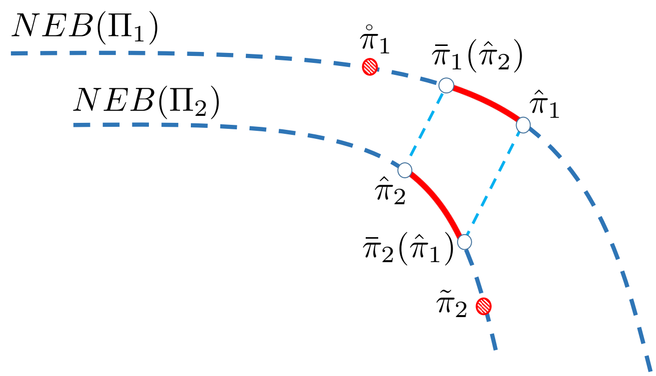


Figure 12: Proposition 5

2, the conditions in Proposition 5 make it that any change in π_i is an ambiguous improvement. Hence, given a decision rule with ex post optimality, $\bar{M}(\cdot)$, there is no quality improvement of π_i for which the principal is willing to trade off ex post optimality. To be more specific, consider two decision rules, $M^1(\pi_1, \pi_2)$ and $M^2(\pi_1, \pi_2)$ with (π_1, π_2) such that $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. By Proposition 3, one of these two decision rules must be optimal *if* it is optimal for the principal to give up ex post optimality. However, each of these two decision rules is payoff-dominated by a decision rule with ex post optimality. For example, $\bar{M}(\bar{\pi}_1(\pi_2), \pi_2)$ payoff-dominates $M^1(\pi_1, \pi_2)$. Note that they share the same π_2. However, $\bar{\pi}_1(\pi_2) \succ \pi_1 (\precsim \hat{\pi}_1)$ since $p\bar{y}_1 + (1-p)(1-\bar{x}_1) = py_2 + (1-p)(1-x_2) > py_1 + (1-p)(1-x_1)$ by definition of $\bar{\pi}_1(\pi_2)$. Thus, under $M^1(\pi_1, \pi_2)$, the principal trades off ex post optimality for an ambiguous improvement of π_1 toward $\hat{\pi}_1$. However, the net benefit of this ambiguous improvement is negative. To see this, we can use the same functional forms for the marginal gain and cost from the discussion on Proposition 4. Now, $\partial C_{A_b}(\cdot)/\partial x_1 > pN_1'(x_1) + (1-p)$ for any π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1$ since (i) $N_1'(x_1)x_1 + (1 - N_1(x_1)) < 0$ for any π_1 such that $\mathring{\pi}_1 \succ \pi_1$ by definition of $\mathring{\pi}_1$ and (ii) we have $\mathring{\pi}_1 \succsim \bar{\pi}_1(\hat{\pi}_2) \succsim \bar{\pi}_1(\pi_2) \succ \pi_1 \succsim \hat{\pi}_1$. Hence $EU^P(\bar{M}(\bar{\pi}_1(\pi_2), \pi_2)) > EU^P(M^1(\pi_1, \pi_2))$ for any π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1 \succsim \hat{\pi}_1$. Similarly, $\bar{M}(\pi_1, \bar{\pi}_2(\pi_1))$ payoff-dominates $M^2(\pi_1, \pi_2)$. Since both $M^1(\pi_1, \pi_2)$ and $M^2(\pi_1, \pi_2)$ we consider here are not optimal, it must be optimal for the principal to achieve ex post optimality.

Note that Proposition 5 *does not* completely characterize an optimal decision rule as Propositions 3 and 4 do. However, Proposition 5 is still helpful. As we know the optimal (A_g^i, A_b^i) for $i = 1, 2$, we only need to find an optimal $\pi_i = (x_i, y_i)$ for $i = 1, 2$. Furthermore, we know that the IC constraint for type θ_1 must be binding. Thus, if $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$ and $y_i = N_i(x_i)$ is differentiable, it is relatively easy to formulate a maximization problem with the binding constraint and two choice variables, x_1 and x_2.[27]

Propositions 4 and 5 tell us an optimal way to resolve the trade-off between the favorite experiments and ex post optimality. Whether an optimal decision rule assigns the favorite experiments or achieves ex post optimality depends on the relative locations of two "best" experiments, $\mathring{\pi}_1$ and $\tilde{\pi}_2$, which determines whether having π_i closer to $\hat{\pi}_i$ is an unambiguous or ambiguous improvement of π_i. Under the conditions in Proposition 4, having π_i closer to $\hat{\pi}_i$ is always an unambiguous improvement. Thus, the principal is willing to trade off ex post optimality for the favorite experiment. Under the conditions in Proposition 5, having π_i closer to $\hat{\pi}_i$ is always an ambiguous improvement. Thus, the principal does not trade off ex post optimality for any quality improvement of π_i toward $\hat{\pi}_i$.

Lastly, I do not exhaustively consider all possible relative locations of $\mathring{\pi}_1$ and $\tilde{\pi}_2$. While one can complete the analysis for every possible case, the results would not provide intuition beyond that given in this paper.[28] For example, if $\bar{\pi}_1(\hat{\pi}_2) \succ \mathring{\pi}_1 \succ \hat{\pi}_1$, $\hat{\pi}_2 \succ \tilde{\pi}_2 \succ \bar{\pi}_2(\hat{\pi}_1)$, and $\bar{\pi}_1(\tilde{\pi}_2) \succsim \mathring{\pi}_1$, the principal will be willing to give up ex post optimality to have π_i as closer to $\hat{\pi}_i$ as possible until such an improvement is *ambiguous*.

4 Conclusion

I study a principal-agent problem in which the agent's action is to conduct an experiment that reveals information about the true state and the principal only can infer the true state via the agent's experiment. While the principal wants to take an appropriate action that depends on the true state, the agent always wants to induce the same action regardless of the true state. On top of these misaligned interests between the agent and the principal, there is information asymmetry: the agent privately observes his type which determines the set of feasible experiments for him. The agent has two types, big and small types: the big type has a larger set of feasible experiments than the small type. The principal can commit to a decision rule before the agent chooses an experiment he wants to conduct. While the principal cannot observe the type of the agent, she can observe the experiment conducted by the agent and its outcome. Thus, the principal's decision rule is contingent on both an experiment and the experimental outcomes.

The main result of this paper is a partial characterization of the principal's optimal

[27] I do not discuss the first-order conditions in the maximization problem for the optimal π_i here as they do not give much information without an explicit form of $N_i(x_i)$.

[28] In a note that is available upon request, I characterize optimal decision rules for other cases in which the relative locations of $\mathring{\pi}_1$ and $\tilde{\pi}_2$ are different from the sufficient conditions in Propositions 4 or 5. However, the results do not add much to those in this paper while the analysis requires the differentiability of $N_i(x_i)$ and involves tediously long proofs.

decision rules. The crucial factor that shapes an optimal decision rule is a trade-off between imposing restrictions on experiments and making the ex post optimal decisions based on the experimental outcomes; to demand each type of the agent to conduct the principal's favorite experiment among the feasible ones, the principal needs to give up making the ex post optimal decisions, and vice versa.

I first characterize the condition under which there is no such a trade-off: if the principal's favorite experiment in the big type's feasible set generates the positive outcome more frequently than that in the small type's feasible set, the principal can assign her favorite experiments and make the ex post optimal decisions.

Then I focus on the cases in which the principal does face the trade-off between assigning her favorite experiments and making the ex post optimal decision. I mainly focus on two kinds of optimal decision rules which possess at least one of two desirable properties in the first-best outcome: (i) a decision rule which assigns the favorite experiments at the cost of giving up the ex post optimal decisions and (ii) a decision rule which achieves the ex post optimal decisions at the cost of giving up the favorite experiments. I provide the sufficient conditions under which each decision rule is optimal.

On the one hand, if the favorite experiments are the results of some unambiguous quality improvements, it is optimal for the principal to assign her favorite experiments to both types. To do so, the principal needs to give up ex post optimality. As it is too costly to give up ex post optimality for both types, the principal deviates from the ex post optimal decisions only for the relatively unlikely type.

On the other hand, it is optimal for the principal to achieve ex post optimality *only if* the favorite experiments are always results of ambiguous quality improvements. To achieve ex post optimality, the principal needs to give up assigning the favorite experiments to both types. In this case, it is not easy to pin down the experiments assigned under an optimal decision rule without specific functional forms for $NEB(\Pi_i)$. However, I provide possible candidates for the experiments that are assigned under an optimal decision rule (Proposition 2); those candidates cannot be arbitrarily far from the favorite experiment. Furthermore, two experiments assigned under this optimal decision rule must be paired in a specific way; they should have the same *ex ante* probability of generating the positive outcomes.

References

Beckelman, J., Y. Li, and C. P. Gross (2003): "Scope and impact of financial conflicts of interest in biomedical research," *Journal of the American Medical Association*, 289, 454–465.

Bergemann, D., and S. Morris (2016a): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11(2), 487–522.

——— (2016b): "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *The American Economic Review*, 106(5), 586–591.

Blackwell, D. (1953): "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 24(2), 265–272.

Blume, A., O. J. Board, and K. Kawamura (2007): "Noisy talk," *Theoretical Economics*, 2(4), 395–440.

Carroll, G. (2017): "Robust incentives for information acquisition," *Working paper.*

Crawford, V. P., and J. Sobel (1982): "Strategic information transmission," *Econometrica: Journal of the Econometric Society*, pp. 1431–1451.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, arbitration and negotiation," *Journal of Economic Theory*, 144(4), 1397–1420.

Green, J. R., and J.-J. Laffont (1986): "Partially verifiable information and mechanism design," *The Review of Economic Studies*, 53(3), 447–456.

Herresthal, C. (2017): "Hidden Testing and Selective Disclosure of Evidence," *Working paper.*

Kamenica, E., and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101(6), 2590–2615.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85(6), 1949–1964.

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

Myerson, R. B. (1979): "Incentive compatibility and the bargaining problem," *Econometrica*, pp. 61–73.

Weber, T. A. (2010): "Simple methods for evaluating and comparing binary experiments," *Theory and decision*, 69(2), 257–288.

Yoder, N. (2022): "Designing incentives for heterogeneous researchers," *Journal of Political Economy*, 130(8), 2018–2054.

Screening for Experiments: Appendix

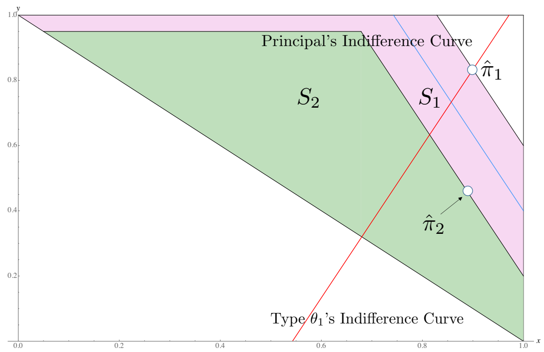
Daehong Min

June 29, 2023

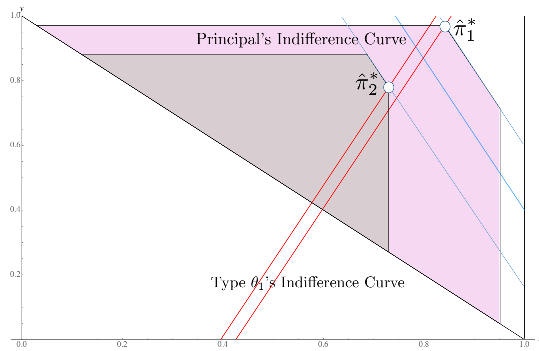
1 Appendix

The Multiplicity of the Favorite Experiments

Without Assumption 1.(d), a part of $\bar{S}_i(x_i)$ can be linear and share the same slope with the principal's indifference curve. Thus, there might exist multiple $\hat{\pi}_i$. We have already seen such a case in Figure 4 given in the left panel of the figure below.



Example A



Example B

If there are multiple $\hat{\pi}_i$ in S_i, we can either (i) exploit the multiplicity to construct a decision rule that can achieve the first-best outcome or (ii) choose $(\hat{\pi}_1, \hat{\pi}_2)$ to which we can apply the results in this paper.

When we have the cases such as Example A, we can exploit the multiplicity of $\hat{\pi}_i$ to construct a decision rule that can achieve the first-beset outcome as follows. In example A, the principal's indifference curve is parallel to both line segments at the northeast of S_1 and S_2. Thus every point on the line segment at the northeast of S_i is $\hat{\pi}_i$. Furthermore, the principal is indifferent among these experiments. Then, we can always find two favorite experiments, $\hat{\pi}_1$ and $\hat{\pi}_2$, which satisfy the condition in Remark 1. In Example A, $\hat{\pi}_1 = (0.9, \frac{5}{6})$ and $\hat{\pi}_2 = (0.9, \frac{13}{30})$ are chosen in such a way assuming that $p = P(G) = 0.3$; $p \cdot \frac{5}{6} + (1-p) \cdot (1-0.9) = 0.32 > p \cdot \frac{13}{30} + (1-p) \cdot (1-0.9) = 0.20$. Then by setting both (A_g^1, A_b^1) and (A_g^2, A_b^2) to $(1, 0)$ we have the following menu offer, $\{((0.9, \frac{5}{6}), 1, 0), ((0.9, \frac{13}{30}), 1, 0)\}$, which can achieve the first-best outcome.

However, there are cases in which we cannot do the construction above. One of such cases is given in Example B above. In these cases, we can choose a unique pair of favorite experiments satisfying a certain condition. By treating the chosen pair as if it is the pair of the unique favorite experiments in the main text, we can apply the results in the main text to these cases. Denote by F_i the set of favorite experiments given Π_i; denote by $\hat{\pi}_i^k$ a favorite experiment in F_i. Formally, it is not possible to construct a decision rule that achieves the first-best outcome as above if

$$\max_{\hat{\pi}_1^k \in F_1} p\hat{y}_1^k + (1-p)(1-\hat{x}_1^k) < \min_{\hat{\pi}_2^k \in F_2} p\hat{y}_2^k + (1-p)(1-\hat{x}_2^k).$$

Denote by $\hat{\pi}_1^*$ the favorite experiment that maximizes $p\hat{y}_1^k + (1-p)(1-\hat{x}_1^k)$; denote by $\hat{\pi}_2^*$ the favorite experiment that minimizes $p\hat{y}_2^k + (1-p)(1-\hat{x}_2^k)$. Example B demonstrates how to find $(\hat{\pi}_1^*, \hat{\pi}_2^*)$. Note that $(\hat{\pi}_1^*, \hat{\pi}_2^*)$ minimizes $|p\hat{y}_1^k + (1-p)(1-\hat{x}_1^k) - (p\hat{y}_2^k + (1-p)(1-\hat{x}_2^k))|$, and it is *uniquely* defined. Given $(\hat{\pi}_1^*, \hat{\pi}_2^*)$, we can apply the results in the main text to these cases by treating $\hat{\pi}_i^*$ as if it is the unique $\hat{\pi}_i$. For example, consider Example B. Given $(\hat{\pi}_1^*, \hat{\pi}_2^*)$, one can easily apply Proposition 2 and check that it is true with similar logic in its proof. Then Lemma 2 and other results can be applied as well. One can check the sufficient conditions in Proposition 4 are satisfied in Example B. Thus an optimal decision rule assigns $\hat{\pi}_i^*$ to type θ_i. Lastly, the optimal action-probability pairs are determined by Proposition 3. To see a rough intuition why we focus on $(\hat{\pi}_1^*, \hat{\pi}_2^*)$, note that $p\hat{y}_i^k + (1-p)(1-\hat{x}_i^k)$ is the agent's payoff when he conducts $\hat{\pi}_i^k$ and the principal makes the ex post optimal decisions. With $(\hat{\pi}_1^*, \hat{\pi}_2^*)$ that minimizes the difference between the agent's payoffs from conducting these two experiments, the principal can have the IC constraint satisfied with the minimum distortion in the ex post optimal decisions when she needs to sacrifice them.

These two ways to handle the multiplicity of $\hat{\pi}_i$ can be applied to the cases in which S_i has multiple $\hat{\pi}_i$ but S_j has a unique $\hat{\pi}_j$.

Comparisons of unambiguous and ambiguous improvements

For expositional convenience, assume that $NEB(\Pi_i) \equiv \{(x, N_i(x_i))\}$ and $N_i(x_i)$ is twice-differentiable. First consider the case when $\pi_i \to \mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i$ which is an unambiguous improvement. In this case, x_i increases, and, thus, $y_i = N_i(x_i)$ decreases. As $\pi_i \to \mathring{\pi}_i$ from the northwest, $\nu(\pi_i) := \frac{1-N_i(x_i)}{x_i}$ decreases by definition of $\mathring{\pi}_i$. Thus, we must have

$$\frac{\partial \nu(\pi_i)}{\partial x_i} = \frac{-N_i'(x_i)x_i - (1-N_i(x_i))}{x_i^2} \leq 0 \iff -N_i'(x_i) \leq \frac{1-N_i(x_i)}{x_i} \leq \frac{1-p}{p},$$

where $\frac{1-N_i(x_i)}{x_i} \leq \frac{1-p}{p}$ holds since any $\pi_i \in NEB(\Pi_i)$ is convincing.[1] Now consider an ambiguous improvement of π_i while π_i moves in the same direction (toward the southeast), i.e., $\mathring{\pi}_i \succsim \pi_i \to \hat{\pi}_i \succsim \tilde{\pi}_i$. In this case, as π_i is already in the southeast of $\mathring{\pi}_i$, $\nu(\pi_i)$ must

[1] For any convincing π_i, $pN_i(x_i) - (1-p)(1-x_i) \geq 0$. Then it must be true that $pN_i(x_i) - (1-p)(1-x_i) \geq 0 \geq 2p - 1$ since $p < 1/2$. This implies that $1 - 2p + pN_i(x_i) - (1-p)(1-x_i) \geq 0$, which is equivalent to $\frac{1-N_i(x_i)}{x_i} \leq \frac{1-p}{p}$.

increase. Furthermore, as π_i approaches $\hat{\pi}_i$ from the northwest, $pN_i(x_i)-(1-p)(1-x_i)$ must increase. Then we must have

$$\frac{1-N_i(x_i)}{x_i} \leq -N_i'(x_i) \leq \frac{1-p}{p},$$

where the last inequality comes from $\partial[pN_i(x_i)-(1-p)(1-x_i)]/\partial x_i \geq 0$. Thus, $-N_i'(x_i)$ has a tighter upper bound in the unambiguous improvement. Note that $-N_i'(x_i)$ measures the absolute value of the decrease in $y_i = N_i(x_i)$ associated with an infinitesimal increase in x_i because $N_i'(x_i) \leq 0$. In other words, the quality worsening in $y_i = N_i(x_i)$ associated with the same amount of the quality improvement in x_i must be less in the unambiguous improvement.

In the comparison between the unambiguous and ambiguous improvement from the southeast (i.e., $\mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i \leftarrow \pi_i$ and $\mathring{\pi}_i \succsim \hat{\pi}_i \leftarrow \pi_i \succsim \tilde{\pi}_i$), the quality improvement in $y_i = N_i(x_i)$ associated with the same amount of the quality worsening in x_i is always bigger in the unambiguous improvement. That is, for the unambiguous improvement (i.e., $\mathring{\pi}_i \succsim \hat{\pi}_i \succsim \tilde{\pi}_i \leftarrow \pi_i$), we have

$$-N_i'(x_i) \geq \frac{N_i(x_i)}{1-x_i} \geq \frac{1-p}{p}.$$

For the ambiguous improvement (i.e., $\mathring{\pi}_i \succsim \hat{\pi}_i \leftarrow \pi_i \succsim \tilde{\pi}_i$), we have

$$\frac{N_i(x_i)}{1-x_i} \geq -N_i'(x_i) \geq \frac{1-p}{p}.$$

Hence, $-N_i'(x_i)$ which represents the quality increase in $y_i = N_i(x_i)$ associated with the same amount of the quality decrease in x_i has a tighter lower bound in the unambiguous improvement.

Proof of Proposition 1

Proof. Consider menu offers such that $\pi_1 \in S_2$ and $\pi_2 \in S_2$. Among these menu offers, the menu offer which maximizes the principal's *ex ante* payoffs is $\{(\hat{\pi}_2, 1, 0), (\hat{\pi}_2, 1, 0)\}$. Now I show that there always exists a menu offer such that $\pi_1 \in S_1 \backslash S_2$ and $\pi_2 \in S_2$ which is incentive compatible and payoff-dominates $\{(\hat{\pi}_2, 1, 0), (\hat{\pi}_2, 1, 0)\}$.

Note that $\hat{\pi}_2 \in S_2 \backslash \underline{S}_2$ under Assumption 1.(c): if $\hat{\pi}_2 \in \underline{S}_2$, every $\pi_2 \in S_2$ is non-convincing, which implies that there is no x_i such that $p\bar{S}_i(x_i) > (1-p)(1-x_i)$. Thus, by Assumption 2, there exists some $\epsilon > 0$ such that

$$\{(x,y) | \sqrt{(\hat{x}_2 - x)^2 + (\hat{y}_2 - y)^2} < \epsilon\} \subset S_1.$$

Then, note that $\pi_1' = (\hat{x}_2 + p\epsilon, \hat{y}_2 + (1-p)\epsilon) \in S_1 \backslash S_2$. First note that $\pi_1' \notin S_2$ as π_1' gives the principal a higher payoff than $\hat{\pi}_2$. That is, if $\pi_1' \in S_2$, $\hat{\pi}_2$ cannot be the favorite experiment in S_2. Secondly, note that $\pi_1' \in \{(x,y) | \sqrt{(\hat{x}_2 - x)^2 + (\hat{y}_2 - y)^2} < \epsilon\} \subset S_1$. Thus, $\pi_1' \in S_1$.

Then, consider a menu offer, $\{(\pi_1', 1, 0), (\hat{\pi}_2, 1, 0)\}$, where $\pi_1' = (\hat{x}_2 + p\epsilon, \hat{y}_2 + (1-p)\epsilon)$. This is a menu offer such that $\pi_1 \in S_1 \backslash S_2$ and $\pi_2 \in S_2$. Furthermore, $\{(\pi_1', 1, 0), (\hat{\pi}_2, 1, 0)\}$ is incentive compatible and payoff-dominates $\{(\hat{\pi}_2, 1, 0), (\hat{\pi}_2, 1, 0)\}$.

For the incentive compatibility, we only need to consider type θ_1's IC constraint which is $EU^A(\pi_1', 1, 0|\theta_1) \geq EU^A(\hat{\pi}_2, 1, 0|\theta_1)$. Note that

$$\begin{aligned} EU^A(\pi_1', 1, 0|\theta_1) &= p(\hat{y}_2 + (1-p)\epsilon) + (1-p)(1 - (\hat{x}_2 + p\epsilon)) = p\hat{y}_2 + (1-p)(1-\hat{x}_2) \\ &= EU^A(\hat{\pi}_2, 1, 0|\theta_1). \end{aligned}$$

Hence, $\{(\pi_1', 1, 0), (\hat{\pi}_2, 1, 0)\}$ is incentive compatible.

Furthermore, the principal can obtain a strictly higher payoffs since type θ_1 conducts an experiment that is more preferred by the principal: π_1' has higher values for both x_1 and y_1 than $\hat{\pi}_2$. More specifically, under $\{(\hat{\pi}_2, 1, 0), (\hat{\pi}_2, 1, 0)\}$, the principal's payoff is

$$EU^P(\cdot) = (1-p) + t(p\hat{y}_2 - (1-p)(1-\hat{x}_2)) + (1-t)(p\hat{y}_2 - (1-p)(1-\hat{x}_2)),$$

but, under $\{(\pi_1', 1, 0), (\hat{\pi}_2, 1, 0)\}$, the principal's payoff is

$$EU^P(\cdot) = (1-p) + t(p(\hat{y}_2 + (1-p)\epsilon) - (1-p)(1 - (\hat{x}_2 + p\epsilon))) + (1-t)(p\hat{y}_2 - (1-p)(1-\hat{x}_2))$$

which is strictly higher than the payoff from $\{(\hat{\pi}_2, 1, 0), (\hat{\pi}_2, 1, 0)\}$. □

Proof of Remark 1

Proof. Suppose $\hat{\pi}_1$ and $\hat{\pi}_2$ satisfy $p\hat{y}_1 + (1-p)(1-\hat{x}_1) \geq p\hat{y}_2 + (1-p)(1-\hat{x}_2)$. Then the following menu offer, $\{(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)\}$, implements the first-best outcome, $[(\hat{\pi}_1, 1, 0), (\hat{\pi}_2, 1, 0)]$. Note that under this menu offer, type θ_1's IC constraint holds: if type θ_1 conducts $\hat{\pi}_1$, he gets $p\hat{y}_1 + (1-p)(1-\hat{x}_1)$ while he gets $p\hat{y}_2 + (1-p)(1-\hat{x}_2)$ by conducting $\hat{\pi}_2$. Thus type θ_1 does not have an incentive to conduct $\hat{\pi}_2$ instead of $\hat{\pi}_1$. □

Proof of Remark 2

Proof. The first result in Remark 2 is proved in the main text. Here I prove the second result, $A_b^1 < 1$ and $A_g^2 > 0$ are optimal. Given the first result, $A_g^1 = 1$ and $A_b^2 = 0$ are optimal, we can focus on the incentive compatible decision rules with $A_g^1 = 1$ and $A_b^2 = 0$, i.e., $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ with $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ and (A_b^1, A_g^2) satisfying the IC constraint below:

$$py_1 + (1-p)(1-x_1) + A_b^1(1 - (py_1 + (1-p)(1-x_1))) \geq A_g^2(py_2 + (1-p)(1-x_2)).$$

Note that $py_i + (1-p)(1-x_i) \in (0,1)$ for $i = 1, 2$ since $\pi_i \in \Pi_i$ implies that $\pi_i \neq (1,0)$ and $\pi_i \neq (0,1)$.

First, I show that $A_b^1 < 1$ is optimal. Choose any $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ and any $A_g^2 \in [0,1]$. Holding the chosen (π_1, π_2) and A_g^2 fixed, the IC constraint gives the lower bound for A_b^1:

$$A_b^1 \geq \max\left\{0, \frac{A_g^2(py_2 + (1-p)(1-x_2)) - (py_1 + (1-p)(1-x_1))}{1 - (py_1 + (1-p)(1-x_1))}\right\}.$$

Since the principal's expected payoff decreases in A_b^1, the optimal solution for A_b^1 is the lower bound, the lowest value among all A_b^1 satisfying the IC constraint. If we have (π_1, π_2) and A_g^2 such that $A_g^2(py_2 + (1-p)(1-x_2)) \leq py_1 + (1-p)(1-x_1)$, $A_b^1 = 0$ is optimal; otherwise, $A_b^1 = \frac{A_g^2(py_2+(1-p)(1-x_2))-(py_1+(1-p)(1-x_1))}{1-(py_1+(1-p)(1-x_1))}$ is optimal. In either case, the optimal solution for A_b^1 is less than 1: $py_1 + (1-p)(1-x_1) \in (0,1)$ for all $\pi_1 \in \Pi_1$ and $A_g^2(py_2 + (1-p)(1-x_2)) < 1$ for all $A_g^2 \in [0,1]$ and for all $\pi_2 \in \Pi_2$.

Second, now choose and fix any $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ and any $A_b^1 \in [0,1]$. The IC constraint gives the upper bound for A_g^2:

$$\min\left\{1, \frac{py_1 + (1-p)(1-x_1) + A_b^1(1 - (py_1 + (1-p)(1-x_1)))}{py_2 + (1-p)(1-x_2)}\right\} \geq A_g^2.$$

Since the principal's expected payoff (weakly) increases in A_g^2, an optimal solution for A_g^2 is the upper bound, the highest value among all A_g^2 satisfying the IC constraint. If we have (π_1, π_2) and A_b^1 such that $py_1 + (1-p)(1-x_1) + A_b^1(1-(py_1 + (1-p)(1-x_1))) \geq py_2 + (1-p)(1-x_2)$, $A_g^2 = 1$ is optimal; otherwise, $A_g^2 = \frac{py_1+(1-p)(1-x_1)+A_b^1(1-(py_1+(1-p)(1-x_1)))}{py_2+(1-p)(1-x_2)}$ is optimal. In either case, the optimal solution for A_g^2 is greater than 0: for all $\pi_i \in \Pi_i$, $py_i + (1-p)(1-x_i) \in (0,1)$ for $i = 1, 2$. $\square$

Proof of Lemma 1

Proof. I show that a decision rule is not optimal if it does not assign $\pi_i \in NEB(\Pi_i)$ for $i = 1, 2$. By Remark 2, it suffices to consider incentive compatible decision rules such that $A^1_g = 1, A^1_b < 1, A^2_g > 0$, and $A^2_b = 0$, i.e., $\{(\pi_1, 1, A^1_b < 1), (\pi_2, A^2_g > 0, 0)\}$. Among these decision rules, consider $\{(\pi_1, 1, A^1_b < 1), (\pi_2, A^2_g > 0, 0)\}$ with either $\pi_i \notin NEB(\Pi_i)$ for $i = 1, 2$ or $\pi_i \notin NEB(\Pi_i)$ and $\pi_j \in NEB(\Pi_j)$ for $i \neq j$. Type θ_1's IC constraint is

$$\big(py_1 + (1-p)(1-x_1)\big) + A^1_b\big(1 - (py_1 + (1-p)(1-x_1))\big) \geq A^2_g\big(py_2 + (1-p)(1-x_2)\big).$$

Fix A^1_b and A^2_g as they are given. Then the inequality above always holds as far as there is no change in the values of $py_1 + (1-p)(1-x_1)$ and $py_2 + (1-p)(1-x_2)$.

Let $py_1 + (1-p)(1-x_1) = h_1$ and $py_2 + (1-p)(1-x_2) = h_2$. Then every π_i on the line, $y_i = \frac{1-p}{p}x_i + \frac{h_i-(1-p)}{p}$, has the same value of $py_i + (1-p)(1-x_i)$.

Since π_i is not on $NEB(\Pi_i)$, an experiment, $\pi'_i = (x_i + p\epsilon, y_i + (1-p)\epsilon)$, should be available for type θ_i; that is, given an experiment, $\pi_i = (x_i, y_i)$, if we slightly move to the "*northeast*" of it along the line, $y_i = \frac{1-p}{p}x_i + \frac{h_i-(1-p)}{p}$, there should be an experiment which is on the line and within Π_i.

Note that $p(y_i + (1-p)\epsilon) + (1-p)(1-(x_i + p\epsilon)) = py_i + (1-p)(1-x_i)$ for $i = 1, 2$. Thus replacing π_i with $\pi'_i = (x_i + p\epsilon, y + (1-p)\epsilon)$ does not make any change in type θ_1's original IC constraint; that is, a decision rule which assigns π'_i to type θ_i instead of π_i is incentive compatible.

Furthermore the decision rule with π'_i gives a strictly higher payoff to the principal than that under the decision rule with π_i since $EU^P(\cdot)$ is increasing in both x_i and y_i. Note that $\pi'_i = (x_i + p\epsilon, y_i + (1-p)\epsilon)$ is a strictly better experiment than $\pi_i = (x_i, y_i)$ because $P(b|B; \pi'_i) = x_i + p\epsilon > P(b|B; \pi_i) = x_i$ and $P(g|G; \pi'_i) = y_i + (1-p)\epsilon > P(g|G; \pi_i) = y_i$.

Thus, if π_i is not on $NEB(\Pi_i)$, we can always find a π'_i which can improve the principal's payoffs without violating the IC constraint. This completes the proof. □

Proof of Proposition 2

Proof. By Remark 2 and Lemma 1, it suffices to consider the following class of decision rules, $\{(\pi_1, 1, A^1_b < 1), (\pi_2, A^2_g > 0, 0)\}$ with $(\pi_1, \pi_2) \in NEB(\Pi_1) \times NEB(\Pi_2)$. For expositional convenience, I use $\{(\pi_1, 1, A^1_b), (\pi_2, A^2_g, 0)\}$ to denote a decision rule in this class.

I start by proving the following argument which I frequently invoke in this proof:

> Given $M = \{(\pi_1, 1, A^1_b), (\pi_2, A^2_g, 0)\}$, $EU^P(M)$ increases as π_i approaches $\hat{\pi}_i$ for $i = 1, 2$, holding A^1_b and A^2_g fixed.

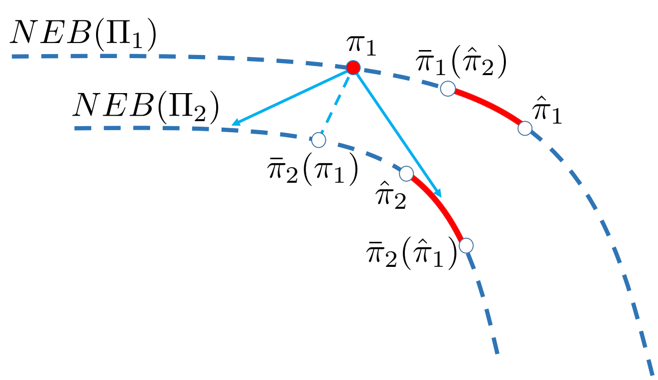
Recall that $\hat{\pi}_i$ is the unique maximizer of $(1-p)+py_i-(1-p)(1-x_i) = py_i+(1-p)x_i$ which is the principal's *interim* expected payoff when making the ex post optimal decisions given a convincing π_i. Given that $NEB(\Pi_i)$ is concave and decreasing, $py_i + (1-p)x_i$ increases before $\hat{\pi}_i$ and decreases after $\hat{\pi}_i$. Hence $py_i + (1-p)x_i$ increases as π_i approaches $\hat{\pi}_i$. Under $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, the principal's ex ante payoff is

$$\begin{aligned} EU^P(M) &= t\Big[(1-p) - (1-2p)A_b^1 + (1-A_b^1)\big(py_1 - (1-p)(1-x_1)\big)\Big] \\ &\qquad\qquad +(1-t)\Big[(1-p) + A_g^2\big(py_2 - (1-p)(1-x_2)\big)\Big] \\ &= t\Big[(1-p)A_b^1 - (1-2p)A_b^1 + (1-A_b^1)\big(py_1 + (1-p)x_1\big)\Big] \\ &\qquad\qquad +(1-t)\Big[(1-p)(1-A_g^2) + A_g^2\big(py_2 + (1-p)x_2\big)\Big]. \end{aligned}$$

Since $(1-A_b^1) > 0$ and $A_g^2 > 0$, $EU^P(M)$ increases as $py_i + (1-p)x_i$ increases. Thus, $EU^P(M)$ increases as π_i approaches $\hat{\pi}_i$, holding A_b^1 and A_g^2 fixed.

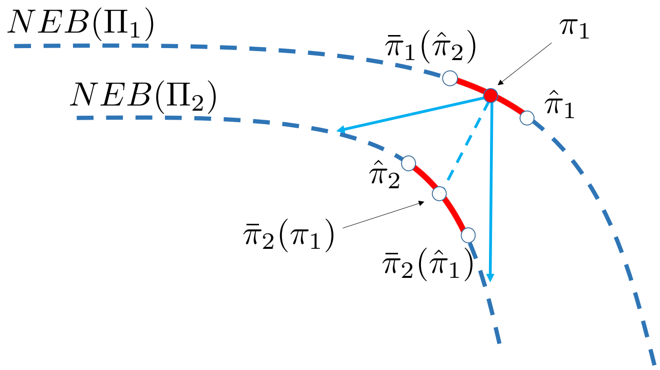
There are three cases to consider. For each case, I start with an incentive compatible decision rule, $M = \{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, such that either π_1 or π_2 does *not* satisfy the condition in Proposition 2. Then I show that there exists an alternative incentive compatible decision rule that payoff-dominates the decision rule I start with.

Case 1: $\pi_1 \succ \bar{\pi}_1(\hat{\pi}_2)$ (π_1 is located in the *northwest* of $\bar{\pi}_1(\hat{\pi}_2)$ as shown below)



Consider any incentive compatible decision rule, $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, which assign $\pi_1 \succ \bar{\pi}_1(\hat{\pi}_2)$ and any π_2 on $NEB(\Pi_2)$. Consider $\{(\pi_1, 1, 0), (\hat{\pi}_2, 1, 0)\}$. First it is incentive compatible as $\bar{\pi}_2(\pi_1) = (\bar{x}_2, \bar{y}_2) \succ \hat{\pi}_2 = (\hat{x}_2, \hat{y}_2)$: $py_1 + (1-p)(1-x_1) = p\bar{y}_2 + (1-p)(1-\bar{x}_2) > p\hat{y}_2 + (1-p)(1-\hat{x}_2)$ holds. Second, it payoff-dominates any decision rule we consider here: (i) the principal makes the ex post optimal decisions and (ii) type θ_2 conducts $\hat{\pi}_2$ that the principal most prefers. Then, lastly, $\{(\pi_1, 1, 0), (\hat{\pi}_2, 1, 0)\}$ is payoff-dominated by $\{(\bar{\pi}_1(\hat{\pi}_2), 1, 0), (\hat{\pi}_2, 1, 0)\}$ as $\bar{\pi}_1(\hat{\pi}_2)$ is closer to $\hat{\pi}_1$ than π_1.

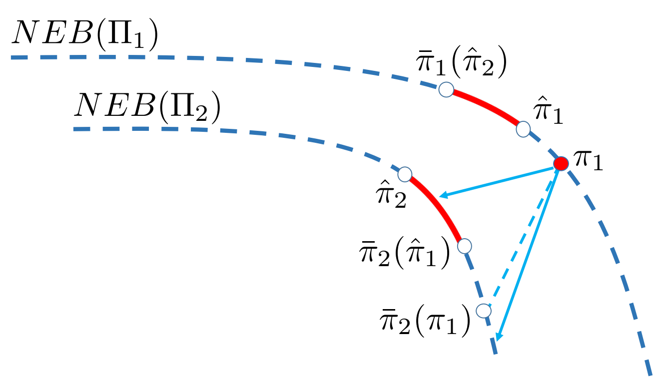
Case 2: $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$ (π_1 is located "between" $\bar{\pi}_1(\hat{\pi}_2)$ and $\hat{\pi}_1$ as shown below)



First, consider any incentive compatible decision rule, $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, which assign π_1 assumed above and any $\pi_2 \succ \hat{\pi}_2$. Then, any such a decision rule is payoff-dominated by $\{(\pi_1, 1, A_b^1), (\hat{\pi}_2, A_g^2, 0)\}$ again as the principal prefers $\hat{\pi}_2$ to any other π_2. Furthermore, $\{(\pi_1, 1, A_b^1), (\hat{\pi}_2, A_g^2, 0)\}$ is incentive compatible. Note that $A_b^1 + (1-A_b^1)(py_1+(1-p)(1-x_1)) \geq A_g^2(py_2 + (1-p)(1-x_2))$ because $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ is incentive compatible. Then, $A_g^2(py_2 + (1-p)(1-x_2)) > A_g^2(p\hat{y}_2 + (1-p)(1-\hat{x}_2))$ because $\pi_2 \succ \hat{\pi}_2$.

Second, consider any incentive compatible decision rule, $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, which assigns π_1 assumed above and any π_2 such that $\bar{\pi}_2(\hat{\pi}_1) \succ \pi_2$. Note that we have $\bar{\pi}_2(\pi_1) \succsim \bar{\pi}_2(\hat{\pi}_1) \succ \pi_2$ which implies $py_1+(1-p)(1-x_1) = p\bar{y}_2+(1-p)(1-\bar{x}_2) \geq p\bar{\bar{y}}_2+(1-p)(1-\bar{\bar{x}}_2) > py_2 + (1-p)(1-x_2)$, where $\bar{\pi}_2(\hat{\pi}_1) = (\bar{\bar{x}}_2, \bar{\bar{y}}_2)$. Thus, the principal can make the ex post optimal decisions without violating the IC constraint. Hence, $\{(\pi_1, 1, 0), (\pi_2, 1, 0)\}$ is incentive compatible and payoff-dominates $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$. Lastly, $\{(\pi_1, 1, 0), (\pi_2, 1, 0)\}$ is payoff-dominated by $\{(\pi_1, 1, 0), (\bar{\pi}_2(\hat{\pi}_1), 1, 0)\}$ as $\bar{\pi}_2(\hat{\pi}_1)$ is closer to $\hat{\pi}_2$ than π_2. The incentive compatibility of $\{(\pi_1, 1, 0), (\bar{\pi}_2(\hat{\pi}_1), 1, 0)\}$ is immediate as $\bar{\pi}_2(\pi_1) \succ \bar{\pi}_2(\hat{\pi}_1)$.

Case 3: $\hat{\pi}_1 \succ \pi_1$ (π_1 is located at the *southeast* of $\hat{\pi}_1$ as shown below)



Consider any incentive compatible decision rule, $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$, which assigns π_1 assumed above and any π_2 on $NEB(\Pi_2)$. Then, $\{(\hat{\pi}_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ payoff-dominates $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ as the principal prefers $\hat{\pi}_1$ to any other π_1. Furthermore, the incentive compatibility of $\{(\hat{\pi}_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ is immediate as $\hat{\pi}_1 \succ \pi_1$: the incentive compatibility of $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ implies $A_b^1 + (1 - A_b^1)(py_1 + (1-p)(1-x_1)) \geq A_g^2(py_2 + (1-p)(1-x_2))$; then, since $\hat{\pi}_1 \succ \pi_1$, $A_b^1 + (1 - A_b^1)(p\hat{y}_1 + (1-p)(1-\hat{x}_1)) > A_b^1 + (1 - A_b^1)(py_1 + (1-p)(1-x_1))$. □

Proof of Lemma 2

Proof. I show that, if the IC constraint is not binding under a decision rule, we can always improve the principal's payoff.

Proposition 2 and Remark 2 imply that we can focus on the following class of decision rules, $\{(\pi_1, 1, A_b^1), (\pi_2, A_g^2, 0)\}$ with $A_b^1 < 1$, $A_g^2 > 0$, and $(\pi_1, \pi_2) \in NEB(\Pi_1) \times NEB(\Pi_2)$ such that $\bar{\pi}_1(\hat{\pi}_2) \precsim \pi_1 \precsim \hat{\pi}_1$ and $\hat{\pi}_2 \precsim \pi_2 \precsim \bar{\pi}_2(\hat{\pi}_1)$. Consider any such a decision rule with a non-binding IC constraint,

$$\big(py_1 + (1-p)(1-x_1)\big) + A_b^1\big(1 - (py_1 + (1-p)(1-x_1))\big) > A_g^2\big(py_2 + (1-p)(1-x_2)\big).$$

The principal's payoff is

$$EU^P(\cdot) = t\big[(1-p) - (1-2p)A_b^1 + (1 - A_b^1)(py_1 - (1-p)(1-x_1))\big] \\ +(1-t)\big[(1-p) + A_g^2(py_2 - (1-p)(1-x_2))\big].$$

Note that $py_i - (1-p)(1-x_i) \geq 0$ for any $(\pi_1, \pi_2) \in NEB(\Pi_1) \times NEB(\Pi_2)$ and $1 - 2p > 0$ for $p \in (0, 1/2)$. Thus, $EU^P(\cdot)$ decreases in A_b^1. Hence, we can increase $EU^P(\cdot)$ by decreasing A_b^1 until the IC constraint binds.

If the IC constraint is not binding after decreasing A_b^1 to 0, we have

$$py_1 + (1-p)(1-x_1) > A_g^2\big(py_2 + (1-p)(1-x_2)\big).$$

Given any π_2 such that $\hat{\pi}_2 \precsim \pi_2 \precsim \bar{\pi}_2(\hat{\pi}_1)$, we have

$$p\hat{y}_2 + (1-p)(1-\hat{x}_2) \geq py_2 + (1-p)(1-x_2) \geq p\bar{\hat{y}}_2 + (1-p)(1-\bar{\hat{x}}_2).$$

Recall that $EU^P(\cdot)$ increases as π_2 approaches $\hat{\pi}_2$ given $A_g^2 > 0$. Furthermore, as π_2 approaches $\hat{\pi}_2$, the right-hand side of the IC constraint above increases. Since the IC constraint is not binding with the given π_2, now we can increase $EU^P(\cdot)$ by replacing π_2 with π_2' that is closer to $\hat{\pi}_2$ (i.e., $\hat{\pi}_2 \precsim \pi_2' \leftarrow \pi_2$) until the IC constraint binds.

If the IC constraint is not binding after replacing π_2 with $\hat{\pi}_2$, we have

$$py_1 + (1-p)(1-x_1) > A_g^2\big(p\hat{y}_2 + (1-p)(1-\hat{x}_2)\big).$$

Since $p\hat{y}_2 - (1-p)(1-\hat{x}_2) > 0$, $EU^P(\cdot)$ increases in A_g^2. Now we can increase $EU^P(\cdot)$ by increasing A_g^2 until the IC constraint binds.

Now the IC constraint must bind with some $A_g^2 \leq 1$. Note that for any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$, we have

$$py_1 + (1-p)(1-x_1) \leq p\hat{y}_2 + (1-p)(1-\hat{x}_2)$$

because $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1$ implies $py_1 + (1-p)(1-x_1) \leq p\bar{\hat{y}}_1 + (1-p)(1-\bar{\hat{x}}_1)$, and $p\bar{\hat{y}}_1 + (1-p)(1-\bar{\hat{x}}_1) = p\hat{y}_2 + (1-p)(1-\hat{x}_2)$ by the definition of $\bar{\pi}_1(\hat{\pi}_2)$. Hence, for any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$, there always exists $A_g^2(\pi_1, \hat{\pi}_2) = \frac{py_1+(1-p)(1-x_1)}{p\hat{y}_2+(1-p)(1-\hat{x}_2)} \leq 1$ that makes the IC constraint bind. This completes the proof.

□

Proof of Lemma 3

Proof. The result is immediate by the fundamental theorem of linear programing. By Remark 2, we have $A_g^1 = 1$ and $A_b^2 = 0$. Once we choose and fix (π_1, π_2) such that $py_1+(1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$, then the principal's payoff is linear in A_b^1 and A_g^2. Furthermore, all constraints in the principal's problem are also linear in A_b^1 and A_g^2. Then, by the fundamental theorem of linear programming, if there exists an optimal solution, there is an optimal solution at an extreme point of the feasible set (a basic feasible solution). Given that the IC constraint must be binding (by Lemma 2) and $(A_b^1, A_g^2) \in [0,1) \times (0,1]$ (by Remark 2), the feasible set for (A_b^1, A_g^2) is

$$\left\{(A_b^1, A_g^2) \in [0,1) \times (0,1] \middle| A_g^2 = \frac{\big(1-\big(py_1+(1-p)(1-x_1)\big)\big)}{py_2+(1-p)(1-x_2)}A_b^1 + \frac{py_1+(1-p)(1-x_1)}{py_2+(1-p)(1-x_2)}\right\}.$$

Then, there are two extreme points on this feasible set, which are

$$\left(A_b^{1'} = \frac{py_2+(1-p)(1-x_2)-(py_1+(1-p)(1-x_1))}{1-(py_1+(1-p)(1-x_1))}, A_g^{2'} = 1\right) \text{ and}$$
$$\left(A_b^{1''} = 0, A_g^{2''} = \frac{py_1+(1-p)(1-x_1)}{py_2+(1-p)(1-x_2)}\right),$$

where $A_b^{1'} \in (0,1)$ and $A_g^{2''} \in (0,1)$ given that $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. Thus, one of these two is an optimal solution. □

Proof of Proposition 3

Proof. By Lemma 3, we know that given any pair of π_1 and π_2 such that $py_1+(1-p)(1-x_1) < py_2+(1-p)(1-x_2)$, either

$$M^1 := \{(\pi_1, 1, A_b^1 > 0), (\pi_2, 1, 0)\} \quad \text{or} \quad M^2 := \{(\pi_1, 1, 0), (\pi_2, A_g^2 < 1, 0)\}$$

must be optimal, where $A_b^1 = \frac{py_2+(1-p)(1-x_2)-(py_1+(1-p)(1-x_1))}{1-(py_1+(1-p)(1-x_1))}$ and $A_g^2 = \frac{py_1+(1-p)(1-x_1)}{py_2+(1-p)(1-x_2)}$. Then,

$$\begin{aligned} & EU^P(M^2) - EU^P(M^1) \\ = \; & t\big((1-2p+py_1-(1-p)(1-x_1))A_b^1\big) + \big(1-t\big)(A_g^2-1)(py_2-(1-p)(1-x_2)) \end{aligned}$$

Note that $EU^P(M^2) - EU^P(M^1) \gtreqless 0$ if and only if

$$\begin{aligned} & t\big((1-2p+py_1-(1-p)(1-x_1))A_b^1\big) \gtreqless \big(1-t\big)(1-A_g^2)(py_2-(1-p)(1-x_2)) \\ \Leftrightarrow \; & \frac{1-2p+py_1-(1-p)(1-x_1)}{1-A_g^2} \cdot \frac{A_b^1}{py_2-(1-p)(1-x_2)} := \tau(\pi_1,\pi_2) \gtreqless \frac{1-t}{t}. \end{aligned}$$

Lastly, given (π_1, π_2) such that $py_1+(1-p)(1-x_1) = py_2+(1-p)(1-x_2)$, $A_b^1 = 0$ and $A_g^2 = 1$. Hence, $M^1(\pi_1,\pi_2) = M^2(\pi_1,\pi_2) := \bar{M}(\pi_1,\pi_2) = \{(\pi_1,1,0),(\pi_2,1,0)\}$. □

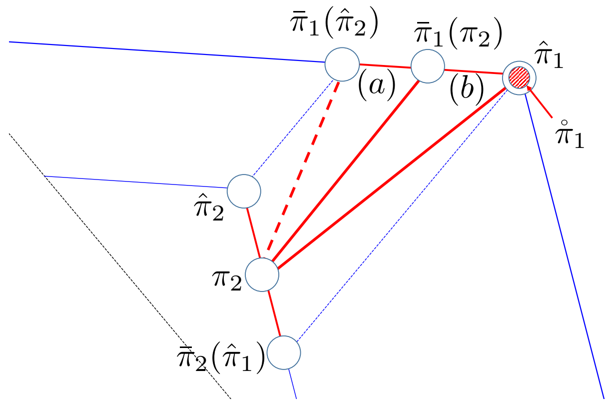
Proof of Proposition 4

Proof. Proposition 4 consists of two claims: (1) if $\mathring{\pi}_1 = \hat{\pi}_1$, it is optimal to assign $\hat{\pi}_1$ to type θ_1 and (2) if $\tilde{\pi}_2 = \hat{\pi}_2$, it is optimal to assign $\hat{\pi}_2$ to type θ_2. I prove each claim below.

Lemma 5. *If $\mathring{\pi}_1 = \hat{\pi}_1$, either $M^1(\hat{\pi}_1, \pi_2)$ or $M^2(\hat{\pi}_1,\pi_2)$ is optimal given any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$.*

Proof. By Proposition 2, we know that an optimal decision rule assigns (π_1, π_2) such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$ and $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. Thus we only need to consider decision rules which assign such a pair of π_1 and π_2.

The proof strategy is as follows. I first choose an arbitrary π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. Then I consider all decision rules which assign the chosen π_2 along with different π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Among these decision rules, I show that the decision rule which assigns the chosen π_2 and $\hat{\pi}_1$ payoff-dominates any other decision rules which assigns the chosen π_2 and any other $\pi_1 \neq \hat{\pi}_1$. Finally, since π_2 is arbitrarily chosen, this is true for any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. The following figure is helpful to understand the proof strategy here.



Choose and fix any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succ \bar{\pi}_2(\hat{\pi}_1)$. Then $\bar{\pi}_1(\pi_2)$ is defined. Note that any $\pi_1 \succ \bar{\pi}_1(\pi_2)$ (any π_1 on region (a) in the figure above) cannot be optimal by Lemma 2 because such a π_1 and the chosen π_2 cannot make the IC constraint binding with the optimal action-probability pairs in Remark 2. Thus we only need to consider π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$ (π_1 on region (b) in the figure above). If $\pi_2 = \bar{\pi}_2(\hat{\pi}_1)$, it is immediate that $\hat{\pi}_1$ is optimal.

The chosen π_2 and π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1 \succsim \hat{\pi}_1$ satisfy $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$. Thus, by Proposition 3, either $M^1(\pi_1;\pi_2) = \{(\pi_1, 1, A_b^1(\pi_1;\pi_2)), (\pi_2, 1, 0)\}$ or $M^2(\pi_1;\pi_2) = \{(\pi_1, 1, 0), (\pi_2, A_g^2(\pi_1;\pi_2), 0)\}$ is optimal, where

$$\begin{aligned} A_b^1(\pi_1;\pi_2) &= \frac{py_2 + (1-p)(1-x_2) - (py_1 + (1-p)(1-x_1))}{1 - (py_1 + (1-p)(1-x_1))}, \text{ and} \\ A_g^2(\pi_1;\pi_2) &= \frac{py_1 + (1-p)(1-x_1)}{py_2 + (1-p)(1-x_2)}. \end{aligned}$$

Note that, if $\pi_1 = \bar{\pi}_1(\pi_2)$, $M^1(\pi_1;\pi_2) = M^2(\pi_1;\pi_2) = \bar{M}(\pi_1;\pi_2) = \{(\pi_1, 1, 0), (\pi_2, 1, 0)\}$. Thus, we can write the principal's *ex ante* payoff under each decision rule for any π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$ as follows:

$$\begin{aligned} EU^P(M^1(\pi_1;\pi_2)) &= t\big(1 - p + py_1 - (1-p)(1-x_1) - A_b^1(\pi_1;\pi_2)(1 - 2p + py_1 - (1-p)(1-x_1))\big) \\ &\qquad + \big(1-t\big)\big(1 - p + py_2 - (1-p)(1-x_2)\big), \\ EU^P(M^2(\pi_1;\pi_2)) &= t(1 - p + py_1 - (1-p)(1-x_1)) \\ &\qquad + (1-t)\big(1 - p + A_g^2(\pi_1;\pi_2)(py_2 - (1-p)(1-x_2))\big), \end{aligned}$$

where $EU^P(M^1(\pi_1;\pi_2)) = EU^P(M^2(\pi_1;\pi_2)) = EU^P(\bar{M}(\pi_1;\pi_2))$ if $\pi_1 = \bar{\pi}_1(\pi_2)$. It is also worth noting that $EU^P(M^i(\pi_1;\pi_2))$ for $i = 1, 2$ varies only in the choice of π_1 as π_2 is fixed.

First, consider $M^1(\pi_1;\pi_2)$. With simple algebra, we can write $EU^P(M^1(\pi_1;\pi_2))$ as a function of the negative likelihood ratio of π_1, $\nu := \frac{1-y_1}{x_1}$. Denote $(1-t)(1 - p + py_2 - (1 -$

$p)(1-x_2))$ by $\bar{c}_2$ as this part is invariant with a fixed π_2. Then, we have

$$\begin{aligned}
EU^P(M^1(\pi_1;\pi_2)) &= t\big[p+(1-A_b^1)\big((1-p)x_1-p(1-y_1)\big)\big]+\bar{c}_2, \\
&= t\Big[p+\frac{1-(py_2+(1-p)(1-x_2))}{1-(py_1+(1-p)(1-x_1))}\big((1-p)x_1-p(1-y_1)\big)\Big]+\bar{c}_2, \\
&= t\Big[p+(1-(py_2+(1-p)(1-x_2))\Big(\frac{(1-p)x_1-p(1-y_1)}{(1-p)x_1+p(1-y_1)}\Big)\Big]+\bar{c}_2, \\
&= t\Big[p+(1-(py_2+(1-p)(1-x_2))\Big(\frac{(1-p)-p\big(\frac{1-y_1}{x_1}\big)}{(1-p)+p\big(\frac{1-y_1}{x_1}\big)}\Big)\Big]+\bar{c}_2, \\
&= t\Big[p+(1-(py_2+(1-p)(1-x_2))\Big(\frac{(1-p)-p\cdot\nu}{(1-p)+p\cdot\nu}\Big)\Big]+\bar{c}_2.
\end{aligned}$$

Then, it is immediate to see that $EU^P(M^1(\pi_1;\pi_2))$ decreases in the negative likelihood ratio, $\nu := \frac{1-y_1}{x_1}$. Thus, $EU^P(M^1(\pi_1;\pi_2))$ is maximized at $\mathring{\pi}_1$ which has the lowest $\frac{1-y_1}{x_1}$. Then, as $\hat{\pi}_1=\mathring{\pi}_1$, we have

$$EU^P(M^1(\mathring{\pi}_1=\hat{\pi}_1;\pi_2)) \geq EU^P(M^1(\pi_1;\pi_2)) \;\text{ for any }\; \pi_1 \;\; s.t. \;\; \bar{\pi}_1(\pi_2)\succsim\pi_1\succsim\hat{\pi}_1.$$

Now consider $M^2(\pi_1;\pi_2)$. Note that it is impossible to write $EU^P(M^2(\pi_1;\pi_2))$ as a function of the negative likelihood ratio as for $EU^P(M^1(\pi_1;\pi_2))$. Instead, I directly compare $M^2(\hat{\pi}_1;\pi_2)$ and $M^2(\pi_1;\pi_2)$ with any π_1 such that $\bar{\pi}_1(\pi_2)\succsim\pi_1\succsim\hat{\pi}_1$. By subtracting $EU^P(M^2(\pi_1;\pi_2))$ from $EU^P(M^2(\hat{\pi}_1;\pi_2))$, we have

$$\begin{aligned}
& EU^P(M^2(\hat{\pi}_1;\pi_2))-EU^P(M^2(\pi_1;\pi_2)) \\
= \;& t\big((p\hat{y}_1-(1-p)(1-\hat{x}_1))-(py_1-(1-p)(1-x_1))\big) \\
& \qquad +\big(1-t\big)\big(A_g^2(\hat{\pi}_1;\pi_2)-A_g^2(\pi_1;\pi_2)\big)\big(py_2-(1-p)(1-x_2)\big).
\end{aligned}$$

Note that $EU^P(M^2(\hat{\pi}_1;\pi_2))-EU^P(M^2(\pi_1;\pi_2))\geq 0$ if and only if

$$\frac{(p\hat{y}_1-(1-p)(1-\hat{x}_1))-(py_1-(1-p)(1-x_1))}{(A_g^2(\pi_1;\pi_2)-A_g^2(\hat{\pi}_1;\pi_2))(py_2-(1-p)(1-x_2))}\geq\frac{1-t}{t},$$

where $A_g^2(\pi_1;\pi_2)-A_g^2(\hat{\pi}_1;\pi_2)=\frac{py_1+(1-p)(1-x_1)-(p\hat{y}_1+(1-p)(1-\hat{x}_1))}{py_2+(1-p)(1-x_2)}\geq 0$ as $\pi_1\succsim\hat{\pi}_1$. Denote the left-hand side of the inequality above by $\rho_1(\hat{\pi}_1,\pi_1;\pi_2)$. By rearranging terms, we have

$$\rho_1(\hat{\pi}_1,\pi_1,\pi_2)=\left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right)\left(\frac{(p\hat{y}_1-(1-p)(1-\hat{x}_1))-(py_1-(1-p)(1-x_1))}{(py_1+(1-p)(1-x_1))-(p\hat{y}_1+(1-p)(1-\hat{x}_1))}\right).$$

Note that $\rho_1(\hat{\pi}_1,\pi_1;\pi_2)$ and $\tau(\pi_1,\pi_2)=\left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right)\left(\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}\right)$ share the same fraction, $\left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right)$. Then, with tedious algebra after subtracting $\tau(\pi_1,\pi_2)$ from $\rho_1(\hat{\pi}_1,\pi_1,\pi_2)$, we have

$$\rho_1(\hat{\pi}_1,\pi_1,\pi_2)-\tau(\pi_1,\pi_2)=\left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right)\left(\frac{2p(1-p)}{D_1D_2}\big((1-y_1)\hat{x}_1-(1-\hat{y}_1)x_1\big)\right),$$

where $D_1 = (py_1 + (1-p)(1-x_1)) - (p\hat{y}_1 + (1-p)(1-\hat{x}_1)) > 0$ and $D_2 = 1 - (py_1 + (1-p)(1-x_1)) > 0$. Now note that

$$\rho_1(\hat{\pi}_1, \pi_1, \pi_2) - \tau(\pi_1, \pi_2) \geq 0 \iff \frac{1-y_1}{x_1} \geq \frac{1-\hat{y}_1}{\hat{x}_1},$$

which is true under the assumption that $\mathring{\pi}_1 = \hat{\pi}_1$. Then, if $\tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$, then we must have $\rho_1(\hat{\pi}_1, \pi_1, \pi_2) \geq \tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$, which immediately implies that $EU^P(M^2(\hat{\pi}_1; \pi_2)) \geq EU^P(M^2(\pi_1; \pi_2))$ for any π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$.

Now choose any π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Suppose $\tau(\pi_1, \pi_2) \leq \frac{1-t}{t}$. Then, we must have $EU^P(M^1(\hat{\pi}_1; \pi_2)) \geq EU^P(M^1(\pi_1; \pi_2)) \geq EU^P(M^2(\pi_1; \pi_2))$. Suppose that $\tau(\pi_1, \pi_2) \geq \frac{1-t}{t}$. Then, we must have $EU^P(M^2(\hat{\pi}_1; \pi_2)) \geq EU^P(M^2(\pi_1; \pi_2)) \geq EU^P(M^1(\pi_1; \pi_2))$. This completes the proof that a decision rule which assigns $\hat{\pi}_1$ payoff-dominates any decision rule which assigns $\pi_1 \neq \hat{\pi}_1$ such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. $\square$

Lemma 6. *If $\tilde{\pi}_2 = \hat{\pi}_2$, either $M^1(\pi_1, \hat{\pi}_2)$ or $M^2(\pi_1, \hat{\pi}_2)$ is optimal given any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$.*

The detailed proof of Lemma 6 above is relegated to the online Appendix as the proof follows almost identical steps in the proof of Lemma 5 above. Basically I repeat the same steps in the proof of Lemma 5 *but* start by fixing a π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Then I show that a decision rule which assigns the chosen π_1 and $\hat{\pi}_2$ payoff-dominates any decision rule which assigns the chosen π_1 and $\pi_2 \neq \hat{\pi}_2$ such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$.

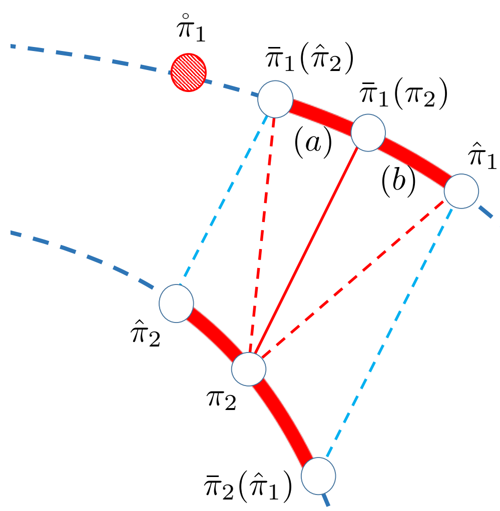
Finally by Lemma 5 and 6, if (1) $\mathring{\pi}_1 = \hat{\pi}_1$ and (2) $\tilde{\pi}_2 = \hat{\pi}_2$, an optimal decision rule assigns $\hat{\pi}_1$ to type θ_1 and $\hat{\pi}_2$ to type θ_2. $\square$

Proof of Proposition 5

Proof. Proposition 5 consists of two claims: (1) if $\mathring{\pi}_1 \succsim \bar{\pi}_1(\hat{\pi}_2)$, then any decision rule with $A_b^1 > 0$ cannot be optimal and (2) if $\bar{\pi}_2(\hat{\pi}_1) \succsim \tilde{\pi}_2$, then any decision rule with $A_g^2 < 1$ cannot be optimal. In the following I prove the first claim and the detailed proof for the second claim is relegated to the online Appendix.

Lemma 7. *If $\mathring{\pi}_1 \succsim \bar{\pi}_1(\hat{\pi}_2)$, any decision rule with $A_b^1(\pi_1, \pi_2) > 0$ cannot be optimal given any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$.*

Proof. The proof strategy is similar to that in Lemma 5. By Proposition 2, it suffices to consider (π_1, π_2) such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$ and $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_2)$. I fix a π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. Then I show that any decision rule with $A_b^1 > 0$ cannot be optimal since such a decision rule is payoff-dominated by a decision rule which assigns $\bar{\pi}_1(\pi_2)$ along with a chosen π_2 without sacrificing *ex post* optimality: $EU^P(\bar{M}(\bar{\pi}_1(\pi_2); \pi_2)) \geq EU^P(M^1(\pi_1; \pi_2))$ for any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. The figure below is similar to that in the proof of Lemma 5 but is different due to the position of $\mathring{\pi}_1$ on $NEB(\Pi_1)$.



Choose and fix any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. Then, $\bar{\pi}_1(\pi_2)$ is defined accordingly. Note that any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succ \bar{\pi}_1(\pi_2)$ (π_1 on region (a) in the figure above) cannot be optimal by Lemma 2: such a π_1 and the chosen π_2 cannot make the IC constraint binding with the optimal action-probability pairs in Remark 2. Hence, we only need to consider π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$ (π_1 on region (b) in the figure above).

Now consider $M^1(\pi_1; \pi_2)$. We know that, for π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$,

$$EU^P(M^1(\pi_1; \pi_2)) = t\left[p + (1 - (py_2 + (1-p)(1-x_2))\left(\frac{(1-p) - p \cdot \nu}{(1-p) + p \cdot \nu}\right)\right] + \bar{c}_2,$$

where $EU^P(M^1(\pi_1; \pi_2)) = EU^P(\bar{M}(\pi_1; \pi_2))$ at $\pi_1 = \bar{\pi}_1(\pi_2)$, $\nu = \frac{1-y_1}{x_1}$, and $\bar{c}_2 = (1-t)(1 - p + py_2 - (1-p)(1-x_2))$. As $EU^P(M^1(\pi_1; \pi_2))$ decreases in $\nu = \frac{1-y_1}{x_1}$, it is optimal to choose π_1 which has the lowest $\frac{1-y_1}{x_1}$. It is easy to see that $\frac{1-y_1}{x_1}$ is minimized at $\mathring{\pi}_1$ then increases as $\pi_1 \to \hat{\pi}_1$ when $NEB(\Pi_1)$ is decreasing and concave. Thus, among π_1 such that $\bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1$, $\bar{\pi}_1(\pi_2)$ has the lowest ν. Thus, we must have

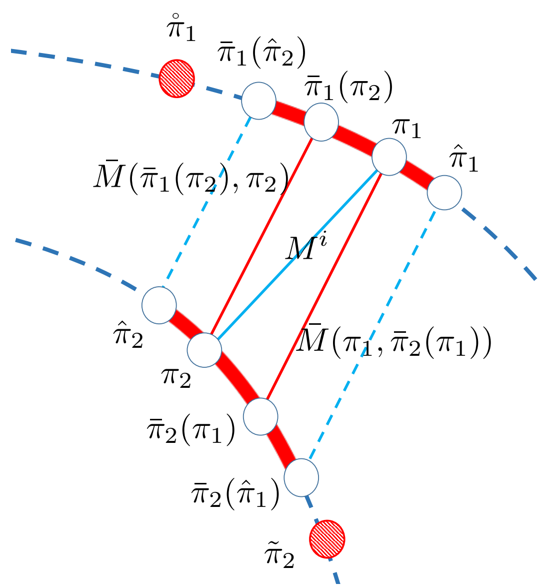
$$EU^P(\bar{M}(\bar{\pi}_1(\pi_2); \pi_2)) \geq EU^P(M^1(\pi_1; \pi_2)) \text{ for any } \pi_1 \;\; s.t. \;\; \bar{\pi}_1(\pi_2) \succsim \pi_1 \succsim \hat{\pi}_1.$$

Note that $A_b^1(\pi_1, \pi_2)$ associated with $M^1(\pi_1; \pi_2)$ is strictly positive if $\bar{\pi}_1(\pi_2) \succ \pi_1$. This completes the proof: $\bar{M}(\bar{\pi}_1(\pi_2); \pi_2)$ always payoff-dominates $M^1(\pi_1; \pi_2)$ with π_1 such that $\bar{\pi}_1(\pi_2) \succ \pi_1 \succsim \hat{\pi}_1$ if $\mathring{\pi}_1 \succsim \bar{\pi}_1(\hat{\pi}_2)$. $\square$

Lemma 8. *If $\bar{\pi}_2(\hat{\pi}_1) \succsim \tilde{\pi}_2$, any decision rule with $A_g^2(\pi_1, \pi_2) < 1$ cannot be optimal given any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$.*

The proof for Lemma 8 is relegated to the online Appendix as this proof also follows almost identical steps in that for Lemma 7 except that I start by choosing and fixing π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$.

Now suppose that it is optimal for the principal to give up ex post optimality, i.e., a decision rule with (π_1, π_2) such that $py_1+(1-p)(1-x_1) < py_2+(1-p)(1-x_2)$ is optimal. Then, by Proposition 3, either $M^1(\pi_1, \pi_2)$ or $M^2(\pi_1, \pi_2)$ with (π_1, π_2) such that $py_1+(1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$ must be optimal. In the figure below, M^i represents these two decision



rules. Note that $M^1(\pi_1, \pi_2)$ has $A_b^1(\pi_1, \pi_2) > 0$. Thus, by Lemma 7, $M^1(\pi_1, \pi_2)$ considered here cannot be optimal and payoff-dominated by $\bar{M}(\bar{\pi}_1(\pi_2), \pi_2)$. Note that $M^2(\pi_1, \pi_2)$ has $A_g^2(\pi_1, \pi_2) < 1$. Thus, by Lemma 8, $M^2(\pi_1, \pi_2)$ considered here cannot be optimal and payoff-dominated by $\bar{M}(\pi_1, \bar{\pi}_2(\pi_1))$. Then, both $M^1(\pi_1, \pi_2)$ and $M^2(\pi_1, \pi_2)$ with (π_1, π_2) such that $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_2)$ are not optimal. Hence it is *not* optimal for the principal to give up ex post optimality, which implies that an optimal decision rule achieves ex post optimality. $\square$

Screening for Experiments: Oneline Appendix

Daehong Min

June 29, 2023

I include the relegated proofs from the main text or the Appendix.

Proof of Lemma 4

Proof. Note that $\hat{\pi}_i = (\hat{x}_i, \hat{y}_i) \in NEB(\Pi_i)$; if $\hat{\pi}_i \notin NEB(\Pi_i)$, there is $\pi_i' = (\hat{x}_i + p\epsilon, \hat{y}_i + (1-p)\epsilon) \in \Pi_i$ for some $\epsilon > 0$; then the principal prefers π_i' to $\hat{\pi}_i$, which is a contradiction. Furthermore, both $\mathring{\pi}_i$ and $\tilde{\pi}_i$ are defined on $NEB(\Pi_i)$. Hence, $\hat{\pi}_i, \mathring{\pi}_i$, and $\tilde{\pi}_i$ are on $NEB(\Pi_i)$.

Now recall that $NEB(\Pi_i)$ does not always coincide with $N_i(x_i)$. That is, if $\bar{S}_i(x_i^{cr}) > \frac{1-p}{p}(1 - x_i^{cr})$, then $N_i(x_i^{cr}) := \bar{S}(x_i^{cr})$ but $NEB(\Pi_i)$ includes the vertical line connecting $\bar{S}_i(x_i^{cr})$ and $\frac{1-p}{p}(1 - x_i^{cr})$:

$$\upsilon := \left\{ \pi_i = (x_i, y_i) \middle| x_i = x_i^{cr}, y_i \in \Big[\frac{1-p}{p}(1 - x_i^{cr}), \bar{S}_i(x_i^{cr})\Big) \right\} \subset NEB(\Pi_i).$$

As I use $N_i(x_i)$ instead of $NEB(\Pi_i)$ in this proof, I first argue that neither of $\hat{\pi}_i$, $\mathring{\pi}_i$ or $\tilde{\pi}_i$ can be on the vertical line of $NEB(\Pi_i)$: $\hat{\pi}_i, \mathring{\pi}_i, \tilde{\pi}_i \notin \upsilon$. Let $\pi_i' = (x_i^{cr}, N_i(x_i^{cr}))$. By definition, $N_i(x_i^{cr}) > y_i$ for any y_i associated with $\pi_i \in \upsilon$. Suppose that $\hat{\pi}_i \in \upsilon$. Then, it is a contradiction as the principal prefers π_i' to $\hat{\pi}_i \in \upsilon$. Similarly, $\tilde{\pi}_i, \mathring{\pi} \notin \upsilon$. Note that π_i' has a higher positive likelihood ratio and a lower negative likelihood ratio than every $\pi_i \in \upsilon$: $\frac{N_i(x_i^{cr})}{1-x_i^{cr}} > \frac{y_i}{1-x_i^{cr}}$ and $\frac{1-N_i(x_i^{cr})}{x_i^{cr}} < \frac{1-y_i}{x_i^{cr}}$ for every $\pi_i = (x_i^{cr}, y_i) \in \upsilon$. Thus when we discuss the relative locations of $\hat{\pi}_i$, $\tilde{\pi}_i$, and $\mathring{\pi}_i$, we only need to focus on $N_i(x_i)$ not $NEB(\Pi_i)$.

For $\mathring{\pi}_i \succsim \hat{\pi}_i$ in (a), suppose that $\hat{\pi}_i \succ \mathring{\pi}_i$ (i.e., $\hat{x}_i < \mathring{x}_i$, and, thus, $N_i(\hat{x}_i) \geq N_i(\mathring{x}_i)$). By definition of $\hat{\pi}_i$,

$$\begin{aligned} & pN_i(\hat{x}_i) - (1-p)(1-\hat{x}_i) \geq pN_i(\mathring{x}_i) - (1-p)(1-\mathring{x}_i), \\ \iff \quad & \frac{N_i(\hat{x}_i) - N_i(\mathring{x}_i)}{\hat{x}_i - \mathring{x}_i} \leq -\frac{1-p}{p}. \end{aligned}$$

By definition of $\mathring{\pi}_i$, we have

$$\frac{1 - N_i(\mathring{x}_i)}{\mathring{x}_i} \leq \frac{1 - N_i(\hat{x}_i)}{\hat{x}_i} \iff -\frac{1 - N_i(\mathring{x}_i)}{\mathring{x}_i} \leq \frac{N_i(\hat{x}_i) - N_i(\mathring{x}_i)}{\hat{x}_i - \mathring{x}_i}.$$

Thus, we must have

$$-\frac{1 - N_i(\mathring{x}_i)}{\mathring{x}_i} \leq \frac{N_i(\hat{x}_i) - N_i(\mathring{x}_i)}{\hat{x}_i - \mathring{x}_i} \leq -\frac{1-p}{p},$$

which implies

$$\frac{1 - N_i(\mathring{x}_i)}{\mathring{x}_i} \geq \frac{1-p}{p} \iff 2p - 1 \geq pN_i(\mathring{x}_i) - (1-p)(1-\mathring{x}_i).$$

As $\mathring{\pi}_i$ is on $NEB(\Pi_i)$, it is a convincing experiment. Hence, $pN_i(\mathring{x}_i) - (1-p)(1-\mathring{x}_i) \geq 0$. This implies that $2p - 1 \geq 0$. This contradicts to the assumption that $p \in (0, 1/2)$. Thus, $\mathring{\pi}_i \succsim \hat{\pi}_i$ (i.e., $\mathring{x}_i \leq \hat{x}_i$, and, thus, $N_i(\mathring{x}_i) \geq N_i(\hat{x}_i)$).

For $\hat{\pi}_i \succsim \tilde{\pi}_i$ in (a), assume that $\tilde{\pi}_i \succ \hat{\pi}_i$ (i.e., $\tilde{x}_i < \hat{x}_i$, and, thus, $N_i(\tilde{x}_i) \geq N_i(\hat{x}_i)$). Again, by definition of $\hat{\pi}_i$, we have

$$\begin{aligned} & pN_i(\hat{x}_i) - (1-p)(1-\hat{x}_i) \geq pN_i(\tilde{x}_i) - (1-p)(1-\tilde{x}_i) \\ \iff \quad & -\frac{1-p}{p} \leq \frac{N_i(\tilde{x}_i) - N_i(\hat{x}_i)}{\tilde{x}_i - \hat{x}_i}. \end{aligned}$$

Then, by definition of $\tilde{\pi}_i$, we have

$$\frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i} \geq \frac{N_i(\hat{x}_i)}{1-\hat{x}_i} \iff \frac{N_i(\tilde{x}_i) - N_i(\hat{x}_i)}{\tilde{x}_i - \hat{x}_i} \leq -\frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i}.$$

Thus, we have

$$-\frac{1-p}{p} \leq \frac{N_i(\tilde{x}_i) - N_i(\hat{x}_i)}{\tilde{x}_i - \hat{x}_i} \leq -\frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i},$$

which implies that

$$\frac{1-p}{p} \geq \frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i} \iff 0 \geq pN_i(\tilde{x}_i) - (1-p)(1-\tilde{x}_i).$$

However, $pN_i(\tilde{x}_i) - (1-p)(1-\tilde{x}_i) \geq 0$ as $\tilde{\pi}_i$ is on $NEB(\Pi_i)$. This implies that $pN_i(\tilde{x}_i) - (1-p)(1-\tilde{x}_i) = 0$, which is equivalent to $\frac{1-p}{p} = \frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i}$. By definition of $\tilde{\pi}_i$, we must have $\frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i} = \frac{1-p}{p} \geq \frac{N_i(x_i)}{1-x_i}$ for all $\pi_i \in NEB(\Pi_i)$. Then, $\frac{N_i(x_i)}{1-x_i} = \frac{1-p}{p}$ for all $\pi_i \in NEB(\Pi_i)$. This is only possible when $NEB(\Pi_i) \equiv \{(x_i, N_i(x_i) = \frac{1-p}{p}(1-x_i))$ for $x_i \in [x_i^{cl}, x_i^{cr}]\}$. This contradicts to Assumption 1.(c). Thus, we conclude that $\hat{\pi}_i \succsim \tilde{\pi}_i$ (i.e. $\hat{x}_i \leq \tilde{x}_i$, and, thus $N_i(\hat{x}_i) \geq N_i(\tilde{x}_i)$).

For (b), suppose that $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$ and $N_i(x_i)$ is twice-differentiable at every $x_i \in [x_i^{cl}, x_i^{cr}]$. Then, under Assumption 1, we have $N_i'(x_i) \leq 0$ and $N_i''(x_i) \leq 0$. Then, by definition of $\hat{\pi}_i = (\hat{x}_i, N_i(\hat{x}_i))$ and differentiability of $N_i(x_i)$, we have

$$\begin{aligned} & \left.\frac{\partial\big[pN_i(x_i) - (1-p)(1-x_i)\big]}{\partial x_i}\right|_{x_i = \hat{x}_i} = pN_i'(\hat{x}_i) + (1-p) = 0 \iff N_i'(\hat{x}_i) = -\frac{1-p}{p}, \\ & \frac{\partial^2\big[pN_i(x_i) - (1-p)(1-x_i)\big]}{\partial x_i^2} = pN_i''(x_i) \leq 0 \quad \text{for any} \quad x_i. \end{aligned}$$

Note that it is impossible to have a corner solution for $pN_i'(x_i) + (1-p) = 0$ when $N_i(x_i)$ is differentiable at every $x_i \in [x_i^{cl}, x_i^{cr}]$: $\hat{x}_i \in (x_i^{cl}, x_i^{cr})$. As $N_i(x_i)$ is differentiable, by the mean value theorem, there exits $x_i^m \in (x_i^{cl}, x_i^{cr})$ such that

$$N_i'(x_i^m) = \frac{N_i(x_i^{cl}) - N_i(x_i^{cr})}{x_i^{cl} - x_i^{cr}} = -\frac{1-p}{p},$$

where the second equality comes from the fact that both x_i^{cl} and x_i^{cr} are on $y = -\frac{1-p}{p}x + \frac{1-p}{p}$ when $\{(x_i, N_i(x_i))\} \equiv NEB(\Pi_i)$. It is immediate that $x_i^m = \hat{x}_i$.

With differentiability of $N_i(x_i)$, we can find $\mathring{\pi}_i$ by studying the first and second derivatives of $\frac{1-N_i(x_i)}{x_i}$:

$$\frac{\partial \frac{1-N_i(x_i)}{x_i}}{\partial x_i} = \frac{-N_i'(x_i)x_i - (1-N_i(x_i))}{x_i^2},$$
$$\frac{\partial^2 \frac{1-N_i(x_i)}{x_i}}{\partial x_i^2} = \frac{-N_i''(x_i)x_i^3 - \big(-N_i'(x_i)x_i - (1-N_i(x_i))\big)2x_i}{x_i^4}.$$

Suppose that $\partial\frac{1-N_i(x_i)}{x_i}/\partial x_i \geq 0$ for all $x_i \in [x_i^{cl}, x_i^{cr}]$. Then, $\mathring{\pi}_i = (x_i^{cl}, N_i(x_i^{cl})) \neq \hat{\pi}_i$. Now suppose that $\partial\frac{1-N_i(x_i)}{x_i}/\partial x_i \leq 0$ for all $x_i \in [x_i^{cl}, x_i^{cr}]$. Then, $\mathring{\pi}_i = (x_i^{cr}, N_i(x_i^{cr}))$, which implies that $\hat{\pi}_i \succ \mathring{\pi}_i$ as $\hat{x}_i < x_i^{cr}$. However, it is not possible as $\mathring{\pi}_i \succsim \hat{\pi}_i$. Lastly, suppose that there exists $x_i \in (x_i^{cl}, x_i^{cr})$ such that $\partial\frac{1-N_i(x_i)}{x_i}/\partial x_i = 0$. Note that, at x_i which makes the first derivative above zero, the second derivative above is greater than zero. Thus, $\frac{1-N_i(x_i)}{x_i}$ is minimized at π_i such that $-N_i'(x_i)x_i - (1-N_i(x_i)) = 0$. In other words, $\mathring{\pi}_i$ is the experiment such that $-N_i'(\mathring{x}_i) = \frac{1-N_i(\mathring{x}_i)}{\mathring{x}_i}$.

We already know that $\mathring{\pi}_i \succsim \hat{\pi}_i$: $\mathring{x}_i \leq \hat{x}_i$, and, thus, $N_i(\mathring{x}_i) \geq N_i(\hat{x}_i)$. Suppose that $\mathring{\pi}_i = \hat{\pi}_i$ (i.e., $\mathring{x}_i = \hat{x}_i$). Then, we must have

$$N_i'(\mathring{x}_i) = N_i'(\hat{x}_i) \iff -\frac{1-N_i(\mathring{x}_i)}{\mathring{x}_i} = -\frac{1-p}{p},$$

which implies that

$$2p - 1 = pN_i(\mathring{x}_i) - (1-p)(1-\mathring{x}_i).$$

Note that $\mathring{\pi}_i$ is on $NEB(\Pi_i)$. Thus, it is a convincing experiment which guarantees that $pN_i(\mathring{x}_i) - (1-p)(1-\mathring{x}_i) \geq 0$. This implies that $2p-1 \geq 0$, which contradicts to the assumption that $p \in (0, 1/2)$. Thus, $\mathring{\pi}_1 \neq \hat{\pi}_1$.

Similarly, with differentiability of $N_i(x_i)$, we can find $\tilde{\pi}_i$ by studying the first and second derivatives of $\frac{N_i(x_i)}{1-x_i}$:

$$\frac{\partial \frac{N_i(x_i)}{1-x_i}}{\partial x_i} = \frac{N_i'(x_i)(1-x_i) + N_i(x_i)}{(1-x_i)^2},$$
$$\frac{\partial^2 \frac{N_i(x_i)}{1-x_i}}{\partial x_i^2} = \frac{N_i''(x_i)(1-x_i)(1-x_i)^2 + (N_i'(x_i)(1-x_i) + N_i(x_i))2(1-x_i)}{(1-x_i)^4}.$$

Suppose that $\partial\frac{N_i(x_i)}{1-x_i}/\partial x_i \geq 0$ for all $x_i \in [x_i^{cl}, x_i^{cr}]$. Then, $\tilde{\pi} = (x_i^{cr}, N_i(x_i^{cr})) \neq \hat{\pi}_i$. Suppose that $\partial\frac{N_i(x_i)}{1-x_i}/\partial x_i \leq 0$ for all $x_i \in [x_i^{cl}, x_i^{cr}]$. Then, $\tilde{\pi}_i = (x_i^{cl}, N_i(x_i^{cl}))$, which implies that $\tilde{\pi}_i \succ \hat{\pi}_i$ as $x_i^{cl} < \hat{x}_i$. However, it is not possible as $\hat{\pi}_i \succsim \tilde{\pi}_i$. Lastly, suppose that there exists

$x_i \in (x_i^{cl}, x_i^{cr})$ such that $\partial \frac{N_i(x_i)}{1-x_i} / \partial x_i = 0$. Note that, at x_i which makes the first derivative above zero, the second derivative is less than zero. Thus, at π_i such that $N_i'(x_i)(1-x_i) + N_i(x_i) = 0$, $\frac{N_i(x_i)}{1-x_i}$ is maximized. In other words, $\tilde{\pi}_i = (\tilde{x}_i, N_i(\tilde{x}_i))$ is the experiment such that $-N_i'(\tilde{x}_i) = \frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i}$. We already know that $\hat{\pi}_i \succsim \tilde{\pi}_i$: $\hat{x}_i \leq \tilde{x}_i$, and thus, $N_i(\hat{x}_i) \geq N_i(\tilde{x}_i)$. Suppose that $\hat{\pi}_i = \tilde{\pi}_i$ (i.e., $\hat{x}_i = \tilde{x}_i$). Then, we must have

$$N_i'(\hat{x}_i) = N_i'(\tilde{x}_i) \iff -\frac{N_i(\tilde{x}_i)}{1-\tilde{x}_i} = -\frac{1-p}{p},$$

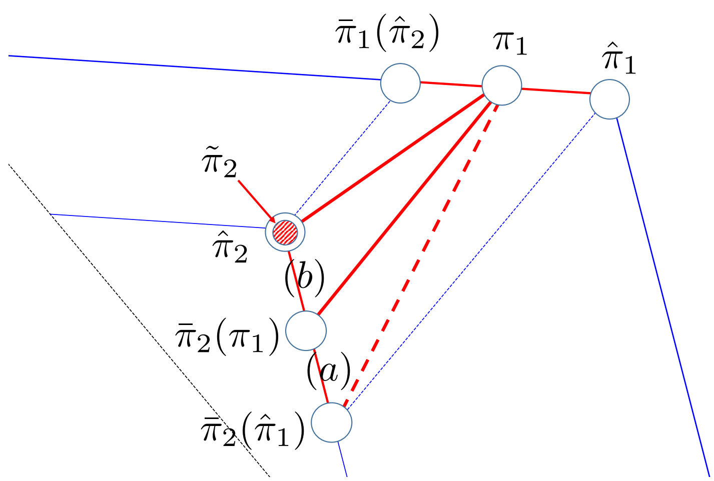
which implies that

$$pN_i(\tilde{x}_i) - (1-p)(1-\tilde{x}_i) = pN_i(\hat{x}_i) - (1-p)(1-\hat{x}_i) = 0.$$

This implies that $pN_i(x_i) - (1-p)(1-x_i) = 0$ for all $\pi_i \in NEB(\Pi_i)$, which is a contradiction to Assumption 1.(c). Hence, $\hat{\pi}_i \neq \tilde{\pi}_i$. $\square$

Proof of Lemma 6

Proof. The proof strategy is exactly opposite to that for Lemma 5. I first choose and fix an arbitrary π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Then I consider all decision rules which assign the chosen π_1 along with different π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$. Among these decision rules, I show that the decision rule which assigns the chosen π_1 and $\hat{\pi}_2$ payoff-dominates any other decision rules which assign the chosen π_1 and any other $\pi_2 \neq \hat{\pi}_2$. The following figure helps to understand the proof strategy again.



Choose and fix any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succ \pi_1 \succsim \hat{\pi}_1$. Note that any π_2 such that $\bar{\pi}_2(\pi_1) \succ \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$ (π_2 on region (a) in the figure above) cannot be optimal because such a π_2 cannot make the IC constraint binding with the optimal action-probability pairs in Remark

2. Thus, we only need to consider π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_1(\pi_2)$ (π_2 on region (b) in the figure above). If $\pi_1 = \bar{\pi}_1(\hat{\pi}_2)$, it is immediate that $\hat{\pi}_2$ is optimal.

The chosen π_1 and π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succ \bar{\pi}_2(\pi_1)$ satisfy $py_1 + (1-p)(1-x_1) < py_2 + (1-p)(1-x_1)$. Thus, by Proposition 3, either $M^1(\pi_2;\pi_1) = \{(\pi_1, 1, A_b^1(\pi_2;\pi_1)), (\pi_2, 1, 0)\}$ or $M^2(\pi_2;\pi_1) = \{(\pi_1, 1, 0), (\pi_2, A_g^2(\pi_2;\pi_1), 0)\}$ is optimal, where

$$\begin{aligned} A_b^1(\pi_2;\pi_1) &= \frac{py_2 + (1-p)(1-x_2) - (py_1 + (1-p)(1-x_1))}{1 - (py_1 + (1-p)(1-x_1))}, \text{ and} \\ A_g^2(\pi_2;\pi_1) &= \frac{py_1 + (1-p)(1-x_1)}{py_2 + (1-p)(1-x_2)}. \end{aligned}$$

Note that, if $\pi_2 = \bar{\pi}_2(\pi_1)$, $M^1(\pi_2;\pi_1) = M^2(\pi_2;\pi_1) = \bar{M}(\pi_2;\pi_1) = \{(\pi_1, 1, 0), (\pi_2, 1, 0)\}$. Then, the principal's *ex ante* payoff under each decision rule for any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$ is

$$\begin{aligned} EU^P(M^1(\pi_2;\pi_1)) &= t\big(1 - p + py_1 - (1-p)(1-x_1) - A_b^1(\pi_2;\pi_1)(1 - 2p + py_1 - (1-p)(1-x_1)\big) \\ &\quad + \big(1-t\big)\big(1 - p + py_2 - (1-p)(1-x_2)\big), \\ EU^P(M^2(\pi_2;\pi_1)) &= t(1 - p + py_1 - (1-p)(1-x_1)) \\ &\quad + (1-t)\big(1 - p + A_g^2(\pi_2;\pi_1)(py_2 - (1-p)(1-x_2))\big), \end{aligned}$$

where $EU^P(M^1(\pi_2;\pi_1)) = EU^P(M^2(\pi_2;\pi_1)) = EU^P(\bar{M}(\pi_2;\pi_1))$ if $\pi_2 = \bar{\pi}_2(\pi_1)$.

Consider $M^2(\pi_2;\pi_1)$. With simple algebra, we can write $EU^P(M^2(\pi_2;\pi_1))$ as a function of the positive likelihood ratio of π_2, $\phi := \frac{y_2}{1-x_2}$. Denote $t\big(1 - p + py_1 - (1-p)(1-x_1)\big)$ by $\bar{c}_1$ as this part would not change as π_2 changes. Then we have

$$\begin{aligned} EU^P(M^2(\pi_2;\pi_1)) &= \bar{c}_1 + (1-t)\big[1 - p + \frac{py_1 + (1-p)(1-x_1)}{py_2 + (1-p)(1-x_2)}(py_2 - (1-p)(1-x_2))\big], \\ &= \bar{c}_1 + (1-t)\left[1 - p + (py_1 + (1-p)(1-x_1))\left(\frac{py_2 - (1-p)(1-x_2)}{py_2 + (1-p)(1-x_2)}\right)\right], \\ &= \bar{c}_1 + (1-t)\left[1 - p + (py_1 + (1-p)(1-x_1))\left(\frac{p\left(\frac{y_2}{1-x_2}\right) - (1-p)}{p\left(\frac{y_2}{1-x_2}\right) + (1-p)}\right)\right], \\ &= \bar{c}_1 + (1-t)\left[1 - p + (py_1 + (1-p)(1-x_1))\left(\frac{p\phi - (1-p)}{p\phi + (1-p)}\right)\right]. \end{aligned}$$

It is easy to see that $\partial EU^P(M^2(\pi_2;\pi_1))/\partial\phi > 0$. Thus, $EU^P(M^2(\pi_2;\pi_1))$ is maximized at $\pi_2 = \tilde{\pi}_2$ which has the highest $\frac{y_2}{1-x_2}$. Then, as $\tilde{\pi}_2 = \hat{\pi}_2$, we have

$$EU^P(M^2(\tilde{\pi}_2 = \hat{\pi}_2;\pi_1)) \geq EU^P(M^2(\pi_2;\pi_1)) \;\text{ for any }\; \pi_2 \;\; s.t. \;\; \hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1).$$

Now consider $M^1(\pi_2;\pi_1)$. Note that it is not possible to write $EU^P(M^1(\pi_2;\pi_1))$ as a function of the positive likelihood ratio as for $EU^P(M^2(\pi_2;\pi_1))$. As in Lemma 5, I directly compare $M^1(\hat{\pi}_2;\pi_1)$ and $M^1(\pi_2;\pi_1)$ with any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$. By subtracting $EU^P(M^1(\pi_2;\pi_1))$ from $EU^P(M^1(\hat{\pi}_2;\pi_1))$, we have

$$\begin{aligned}
& EU^P(M^1(\hat{\pi}_2;\pi_1)) - EU^P(M^1(\pi_2;\pi_1)) \\
= \; & t(A_b^1(\pi_2;\pi_1) - A_b^1(\hat{\pi}_2;\pi_1))(1-2p+py_1-(1-p)(1-x_1)) \\
& \qquad +(1-t)(p\hat{y}_2-(1-p)(1-\hat{x}_2)-(py_2-(1-p)(1-x_2))).
\end{aligned}$$

Note that $EU^P(M^1(\hat{\pi}_2;\pi_1)) - EU^P(M^1(\pi_2;\pi_1)) \geq 0$ if and only if

$$\frac{1-t}{t} \geq \frac{(A_b^1(\hat{\pi}_2;\pi_1) - A_b^1(\pi_2;\pi_1))(1-2p+py_1-(1-p)(1-x_1))}{p\hat{y}_2-(1-p)(1-\hat{x}_2)-(py_2-(1-p)(1-x_2))},$$

where $A_b^1(\hat{\pi}_2;\pi_1)_b - A_b^1(\pi_2;\pi_1) = \frac{p\hat{y}_2+(1-p)(1-\hat{x}_2)-(py_2+(1-p)(1-x_2))}{1-(py_1+(1-p)(1-x_1))} \geq 0$ as $\hat{\pi}_2 \succsim \pi_2$. Denote the right-hand side of the inequality above by $\rho_2(\hat{\pi}_2,\pi_2;\pi_1)$. By rearraging terms, we have

$$\rho_2(\hat{\pi}_2,\pi_2;\pi_1) = \left(\frac{p\hat{y}_2+(1-p)(1-\hat{x}_2)-(py_2+(1-p)(1-x_2))}{p\hat{y}_2-(1-p)(1-\hat{x}_2)-(py_2-(1-p)(1-x_2))}\right)\left(\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}\right).$$

Note that $\tau(\pi_1,\pi_2) = \left(\frac{py_2+(1-p)(1-x_2)}{py_2-(1-p)(1-x_2)}\right)\left(\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}\right)$ and $\rho_2(\hat{\pi}_2,\pi_2;\pi_1)$ share the same fraction, $\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}$. By subtracting $\rho_2(\hat{\pi}_2,\pi_2;\pi_1)$ from $\tau(\pi_1,\pi_2)$, we have

$$\tau(\pi_1,\pi_2) - \rho_2(\hat{\pi}_2,\pi_2;\pi_1) = \left(\frac{2p(1-p)}{D_1'D_2'}\right)(\hat{y}_2(1-x_2)-y_2(1-\hat{x}_2))\left(\frac{1-2p+py_1-(1-p)(1-x_1)}{1-(py_1+(1-p)(1-x_1))}\right),$$

where $D_1' = py_2-(1-p)(1-x_2)$ and $D_2' = p\hat{y}_2-(1-p)(1-\hat{x}_2)-(py_2-(1-p)(1-x_2))$. Then, note that

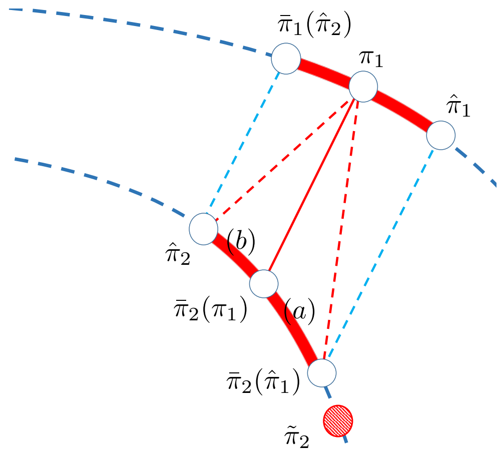
$$\tau(\pi_1,\pi_2) - \rho_2(\hat{\pi}_2,\pi_2;\pi_1) \geq 0 \iff \frac{\hat{y}_2}{1-\hat{x}_2} \geq \frac{y_2}{1-x_2},$$

which is true under the assumption that $\hat{\pi}_2 = \tilde{\pi}_2$. Hence, if $\frac{1-t}{t} \geq \tau(\pi_1,\pi_2)$, we must have $\frac{1-t}{t} \geq \tau(\pi_1,\pi_2) \geq \rho_2(\hat{\pi}_2,\pi_2;\pi_1)$, which implies $EU^P(M^1(\hat{\pi}_2;\pi_1)) \geq EU^P(M^1(\pi_2;\pi_1))$ for any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$.

Choose any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$. If $\tau(\pi_1,\pi_2) \geq \frac{1-t}{t}$, then we must have $EU^P(M^2(\hat{\pi}_2;\pi_1)) \geq EU^P(M^2(\pi_2;\pi_1)) \geq EU^P(M^1(\pi_2;\pi_1))$. If $\tau(\pi_1,\pi_2) \leq \frac{1-t}{t}$, then we must have $EU^P(M^1(\hat{\pi}_2;\pi_1)) \geq EU^P(M^1(\pi_2;\pi_1)) \geq EU^P(M^2(\pi_2;\pi_1))$. This completes the proof that a decision rule which assigns $\hat{\pi}_2$ payoff-dominates any decision rule which assigns $\pi_2 \neq \hat{\pi}_2$ such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. $\square$

Proof for Lemma 8

Proof. The proof strategy is similar to that in Lemma 7. Choose and fix a π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Then I show that any decision rule with $A_g^2 < 1$ cannot be optimal since such a decision rule is payoff-dominated by the decision rule which assigns $\bar{\pi}_2(\pi_1)$ along with a chosen π_1 without sacrificing *ex post* optimality: $EU^P(\bar{M}(\bar{\pi}_2(\pi_1); \pi_1)) \geq EU^P(M^2(\pi_2; \pi_1))$ for any π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$. The figure below is helpful to visualize the proof strategy here.



Choose and fix any π_1 such that $\bar{\pi}_1(\hat{\pi}_2) \succsim \pi_1 \succsim \hat{\pi}_1$. Then, $\bar{\pi}_2(\pi_1)$ is defined accordingly. Note that any π_2 such that $\bar{\pi}_2(\pi_1) \succ \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$ (π_2 on region (a) in the figure above) cannot be optimal since such a π_2 and the chosen π_1 cannot make the IC constraint binding with the optimal action-probability pairs in Remark 2. Thus, we only need to consider π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$ (π_2 on region (b) in the figure above).

Now consider $M^2(\pi_2; \pi_1)$. We know that, for π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$,

$$EU^P(M^2(\pi_2; \pi_1)) = \bar{c}_1 + (1-t)\left[1 - p + (py_1 + (1-p)(1-x_1))\left(\frac{p\phi - (1-p)}{p\phi + (1-p)}\right)\right],$$

where $EU^P(M^2(\pi_2; \pi_1)) = EU^P(\bar{M}(\pi_2; \pi_1))$ at $\pi_2 = \bar{\pi}_2(\pi_1)$, $\phi = \frac{y_2}{1-x_2}$, and $\bar{c}_1 = t(1 - p + py_1 - (1-p)(1-x_1))$. As $EU^P(M^2(\pi_2; \pi_1))$ increases in ϕ, it is optimal to choose π_2 which has the highest $\frac{y_2}{1-x_2}$. It is easy to see that $\frac{y_2}{1-x_2}$ increases as $\pi_2 \to \tilde{\pi}_2$ from the left of it and is maximized at $\tilde{\pi}_2$ since we have a decreasing and concave $NEB(\Pi_2)$. As $\bar{\pi}_2(\hat{\pi}_1) \succsim \tilde{\pi}_2$, among π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1)$, $\bar{\pi}_2(\pi_1)$ has the highest ϕ. Then, we must have

$$EU^P(\bar{M}(\bar{\pi}_2(\pi_1); \pi_1)) \geq EU^P(M^2(\pi_2; \pi_1)) \ \text{ for any } \ \pi_2 \ \ s.t. \ \ \hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\pi_1).$$

Note that $A_g^2(\pi_2; \pi_1)$ associated with $M^2(\pi_2; \pi_1)$ is strictly less than 1 if $\pi_2 \succ \bar{\pi}_2(\pi_1)$. This completes the proof: $\bar{M}(\bar{\pi}_2(\pi_1); \pi_1)$ always payoff-dominates $M^2(\pi_2; \pi_1)$ with π_2 such that $\hat{\pi}_2 \succsim \pi_2 \succsim \bar{\pi}_2(\hat{\pi}_1)$ if $\bar{\pi}_2(\hat{\pi}_1) \succsim \tilde{\pi}_2$. □