Bayesian Persuasion by an Informed Mechanism Designer

Yanlin Chen [*] Jun Zhang [†]

March 10, 2022

Abstract

We investigate how an informed designer maximizes her objective when facing a player whose payoff depends on both the designer's private information and on an unknown state within the classical quasilinear environment. The designer can disclose arbitrary information about the state via Bayesian persuasion and adopt arbitrary mechanisms. We characterize the Rothschild-Stiglitz-Wilson (RSW) mechanism and identify three channels for achieving separation. While disclosing inefficient information is essential, providing a bonus for participation

[*]Institute for Social and Economic Research, Nanjing Audit University, China, Email: yanlinchen@nau.edu.cn.

[†]Economics Discipline Group, School of Business, University of Technology Sydney; Tel: +61 2 95143203, Fax: +61 2 95147722, E-mail: zhangjunqueens@gmail.com. We would like to thank Rabah Amir, Benjamin Balzer, Yeon-Koo Che, Guillaume Frechette, Isa Hafalir, Anton Kolotilin, Jingfeng Lu, Xianwen Shi, Toru Suzuki, Emil Temnyalov, Kentaro Tomoeda, Chengsi Wang, John Wooders, Jingjing Zhang, Xin Zhao, Junjie Zhou and the audience of various seminars in University of Technology Sydney, Shanghai University of Finance and Economics, Nanjing Audit university, Nanjing University, Australasian Economic Theory Workshop 2018, Asian Meeting of the Econometric Society 2018, Econometric Society European Winter Meeting 2018, ISER-NEXL Joint Economic Summer Workshop 2019, and Australasian Economic Theory Workshop 2021 for their valuable comments and suggestions.

and randomization are supplementary. The equilibrium is unique and robust under weak assumptions. Our results can provide rationales for many phenomenons in practice.

Keywords: Bayesian persuasion, information design, mechanism design, informed principal.

JEL Classifications: D11, D82, D83, L12

1 Introduction

In many, if not most, mechanism design situations, designers can disclose information regarding a payoff relevant state of nature to players. In selling problems, for example, film producers can fully determine how much information to include in trailers; video game or software developers can fully design trial versions for their products. This allows consumers to learn more about their valuations of the products. In procurement, procurers can freely decide how much details about their pursued products to put in information sessions, so that suppliers can form expectation about associated production costs. In hiring CEOs, employers can freely determine how to disclose information about the current status of their companies to candidates so that the candidates can judge whether their skills can meet the challenges in the companies. In regulations, governments can disclose information about the features of the industry to regulated firms who then can evaluate whether their existing technologies could fit in the industry. In designing emission quotas, governments can reveal information to polluting firms for estimating the efficiency of their pollution reduction technologies.

In these situations, it is also common that designers may possess private and informative (not necessarily perfect) information about the state at the beginning. In selling problems, sellers know the quality of their products since they are likely to have an unobservable informational advantage over consumers, i.e., the well-known lemon problem. In procurements, procurers know more than the suppliers about the difficulty in producing desired products. In hiring CEOs, companies know more than the candidates about their own profitability. In regulations, governments are likely to be better informed than the monopolies about policies that affect the industry production efficiency. In designing emission quotas, governments are usually better informed than the polluting firms about the average marginal emission abatement efficiency of the industry.

If privately informed designers have total freedom to disclose arbitrary information about the state and to design arbitrary mechanisms, what outcome would we expect? Would it be possible for designers with different types to separate from each other? If so, what would be the most efficient way to do so, given the flexibility of designers' strategies? These are some of the questions we aim to address in this paper. In our model, there is an unknown payoff relevant state of nature and the designer observes a binary information which is informative about the state. The informed designer (she) determines how much information regarding the state to reveal to the player (he), chooses between binary social alternatives, and decides how monetary transfers should be made. In order to allow any disclosure rules under full commitment, we adopt the well-known approach of Bayesian persuasion pioneered by Kamenica and Gentzkow (2011).

Following the literature on informed principal, we allow the designer to design a direct grand mechanism in which she is a participant.[1] This formulation allows the most general strategy space for the designer. We first consider the perfect Bayesian equilibrium (PBE) and show that it is outcome equivalent to consider a simplified game. We then characterize the RSW (Rothschild-Stiglitz-Wilson) mechanism, which was initially introduced by Maskin and Tirole (1992) and has been the main mechanism studied in the literature on informed principal with common values. The RSW mechanism is important and intuitive. It is defined as achieving the maximal payoff for each type of designer among all safe mechanisms in which the player participates regardless of his belief about the designer's type.

[1] See Myerson (1983), Maskin and Tirole (1990, 1992), Severinov (2008), Mylovanov and Troger (2012, 2014), Balkenborg and Makris (2015); Koessler and Skreta (2016, 2019).

In the RSW mechanism of our model, the low-type designer's problem is easy to solve, and she achieves the same payoff as if her type were known by the player. She discloses efficient information and extracts the full surplus. In contrast, the high-type designer's problem is much more challenging. We show that her problem can be reduced to a modified concavification problem with a single variable. We identify three channels for achieving separation: disclosing inefficient information, providing a bonus for participation, and randomization. Disclosing inefficient information always arises and is thus essential. Providing a bonus for participation is necessary when the designer has "distinct types", and randomization may be needed when the designer has "similar types". Under weak assumptions, the equilibrium is unique. Our results can rationalize why monopolists provide bonuses for trials and adopt promotional strategies, why procurers hold free information session accompanied with complementary benefits, why employers treat candidates in a generous manner, why regulators and governments subsidize firms in early stages before establishment, and why governments provide various benefits to participants in foreign investment.

We further conduct robustness checks for the RSW mechanism. First, we show that the RSW mechanism can be supported as a PBE and then provide a sufficient condition such that its outcome equals that from the PBE. Second, we show that the RSW mechanism always survives the intuitive criterion, and that the set of outcomes from the intuitive equilibrium coincides with that from the RSW mechanism when the designer's two types are "similar". It is worth noting that we obtain these results without sorting assumptions commonly imposed in the literature. Therefore, this reinforces the focus on the RSW mechanism in the literature. Finally, with more than binary social alternatives, we show that similar results hold with linear payoffs.

In Chen and Zhang (2020), instead of considering a general mechanism design problem, we restrict to the selling problem where a seller makes a take-it-or-leave-it price after disclosing information to a buyer. It has been shown that a unique equilibrium exists that survives the intuitive criterion. Separating is achieved by disclosing inefficient information. A natural question then is: what if the seller is allowed to choose any selling mechanism? To answer this, one needs to take the informed principal approach, which is dramatically different from and more challenging than the normative approach in Chen and Zhang (2020). It turns out that this question can be treated as a very special case of the current paper, which covers many other applications such as procurements, hiring, regulations and emission quotas. Our results reveal that the high-type seller can improve on the outcome in Chen and Zhang (2020). This is achieved by uncovering two supplementary channels by means of which the low-type seller's mimicking incentive can be deterred: providing a bonus for the buyer's participation, and randomization. Therefore, compared with Chen and Zhang (2020), the current paper consider a much more general problem, adopts a much more general approach, and obtains new results.

This paper is related to the literature on Bayesian persuasion with an informed sender.[2] Perez-Richet (2014) considers a perfectly informed sender and demonstrates that it is without loss of generality to focus on pooling equilibria. Alonso and Câmara (2018) compare the profits of an informed and an uninformed sender and conclude that the sender does not benefit from private information. Hedlund (2017) studies an imperfectly informed sender who signals her type through Bayesian persuasion and selects equilibria with D1 criterion. Koessler and Skreta (2022) allow a general framework with multiple agents and introduce interim-optimal mechanisms. In these studies, the only channel for signalling is information disclosure. In our paper, the designer can also signal through the design of the mechanism, and we adopt the informed principal approach, which requires different techniques.

This paper is related to the literature on mechanism design by an informed principal. Myerson (1983) introduces the inscrutability principle, which states that it is

[2] See Hedlund (2017) for an excellent review of this strand of literature. There is also a large literature on Bayesian persuasion when the sender is not privately informed following the pioneering work of Kamenica and Gentzkow (2011), and its multiple-receiver generalization which is often referred as information design such as Bergemann and Morris (2016, 2019), Mathevet et al. (2020) and Taneva (2019).

without loss of generality to focus on pooling equilibria. He also introduces the concept of safe allocation, which is incentive compatible and individually rational for the agent given any belief about the principal's type. Depending on whether the principal's private type affects the agent's payoff, this literature can be divided into two strands: private values and common values. In this paper, the player's payoff depends on the designer's private information, and thus our model belongs to the latter. For this strand, Maskin and Tirole (1992) introduce the RSW mechanism, which yields the optimal payoff for each type of principal among safe allocations.[3] The RSW mechanism then becomes the main focus of the informed principal with common values literature. Its applications include collusion in Quesada (2005), auction design in Zhao (2018) and bilateral trade in Segal and Whinston (2003) and Nishimura (2022).[4] In these papers, informed principals do not disclose information. Besides being practical relevant, information disclosure is special since it does not involve any explicit cost and is an action with infinite dimensions. Furthermore, introducing information disclosure to the informed principal problem brings new technical challenges as sorting assumptions no longer hold. However, we show that many results remain valid without the sorting assumptions.

In Skreta (2011), an informed seller observes a vector of signals correlated with buyers' valuations before choosing the selling mechanism. She can decide whether to disclose this information to the buyers. Koessler and Skreta (2016, 2019) investigate whether an informed principal can benefit from certification technology with respect

[3] For the strand with private values, Maskin and Tirole (1990) examine whether the principal can benefit from private information. Mylovanov and Troger (2012, 2014) investigate the existence and properties of strongly neologism-proof allocations. Its applications include bilateral trading in Yilankaya (1999), procurement contracting Tan (1996) and Balestrieri (2008), social decision making in Severinov (2008), collusion in Francetich and Troyan (2012) and Hotelling competitions in Balestrieri and Izmalkov (2016).

[4] Bedard (2017) provides conditions under which the principal strictly prefers not being fully informed in the RSW mechanism. Dosis (2019) provides an alternative proof to Theorem 1 of Maskin and Tirole (1992), which states that any allocation that weakly dominates the RSW mechanism can be supported as an equilibrium allocation.

to her own type. They focus on ex ante optimal allocation and show that it can always be supported as an equilibrium. In those papers, the seller can disclose her private information directly. In contrast, our designer directly discloses information about a payoff-relevant state but not about her private information. In addition, we can accommodate a general class of mechanism design problem.

Finally, this paper is related to the literature in which sellers design information disclosure and selling mechanism jointly—but which lacks the signalling issue, the main feature of our model. A buyer's payoff depends on both his private information and on other information that can be disclosed by the seller without cost. Eso and Szentes (2007) develop an orthogonal decomposition technique, and show that the optimal mechanism can be implemented by handicap auctions with full disclosure. Li and Shi (2017) construct a discriminatory information disclosure that improves the revenue in Eso and Szentes (2007). Hoffmann and Inderst (2011) assume costly information disclosure, and identify the conditions such that information is over and under provided.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 considers the full information benchmark. Section 4 considers perfect Bayesian equilibria. Section 5 characterizes the RSW mechanism. Section 6 establishes the robustness of the RSW mechanism. Section 7 considers applications and extensions. Section 8 concludes. All proofs are contained in appendices.

2 The model

We build our general model based on the classical quasi-linear mechanism design framework used in recent studies by Chen et al. (2019) and Gershkov et al. (2013).

Preliminaries

A risk-neutral designer (she) faces a risk-neutral player (he). There is an unknown payoff-relevant state of nature denoted as ω. The designer receives private and informative (but not necessarily perfect) information, denoted as θ, about ω. We assume that θ is a binary random variable on $\{L, H\}$ with commonly known prior probabilities μ_L^0 and $\mu_H^0 = 1 - \mu_L^0$, respectively. When the designer's type is θ, the state ω follows an atomless distribution with *c.d.f* $F_\theta(\omega)$ and *p.d.f* $f_\theta(\omega)$ on the common support $[\underline{\omega}, \overline{\omega}] \subseteq \mathbb{R}_+$. We assume that $f_\theta(\omega)$ satisfies the standard monotone likelihood ratio property: $\frac{f_H(\omega)}{f_L(\omega)} > \frac{f_H(\omega')}{f_L(\omega')}, \forall \omega > \omega'$, which means that when the designer's type is H, the state is more likely to be higher. Bayes' rule implies that the prior belief about the state is $f(\omega) = \mu_H^0 f_H(\omega) + \mu_L^0 f_L(\omega)$.

We assume that the designer can directly disclose information about the state ω via Bayesian persuasion, but not about her private information θ. Note that the situation where the designer could disclose her private information θ directly is a special case of our model by letting ω and θ be perfectly correlated. In this case, it is trivial that the full information outcome can be achieved.[5]

For expositional clarity, we start with a binary set $\mathcal{K} = \{0, 1\}$ of social alternatives, which has natural interpretations in applications. For example, in monopoly pricing with single unit object, the decision is to sell or not to sell. In procurement, a procurer decides whether to proceed with a supplier. In hiring, the decision is whether to hire a candidate. In regulation, the decision is whether to allow the regulated firm to produce. In emission problem, the government decides whether to issue a permission to a polluting firm. In Section 7.3, we show that all results extend beyond binary alternatives if we assume linear payoffs. For alternative 1, given the designer's type θ, the state ω and the monetary transfer from the player to the designer t, the player's ex-post payoff is $P(\theta, \omega) - t$, and the designer's ex-post payoff is $D(\theta, \omega) + t$. We do not impose any sign restrictions on $P(\theta, \omega)$, $D(\theta, \omega)$ and t to allow various applications. For instance, the monetary transfer could be made from the player to the designer (positive), or the other way around (negative). We call $P(\theta, \omega)$ the player's return, and $D(\theta, \omega)$ the designer's return. The ex-post total surplus can be calculated as $T(\theta, \omega) = P(\theta, \omega) + D(\theta, \omega)$, which is independent of t due to the quasi-linearity. We

[5] See more discussion in Section 7.2 of Chen and Zhang (2020).

let $k = 0$ denote the alternative representing the outside option where both the player and the designer receive zero payoffs.[6]

Regular environment

We impose some regularity conditions on the environment.

Definition 1 *The environment is regular if*
(i) $T(\theta, \omega) \geq 0$, *and is strictly increasing in* ω,
(ii) $P(\theta, \omega)$ *is weakly increasing in* θ, *and* $P(\theta, \omega)$ *is strictly increasing in* ω,
(iii) $D(\theta, \omega)$ *is weakly decreasing in* θ, *and* $D(\theta, \omega)$ *is weakly decreasing in* ω,
(iv) $\frac{-D(H,\omega)+D(L,\underline{\omega})}{P(H,w)+D(L,\underline{\omega})}$ *is weakly decreasing in* ω, *and* $\frac{[D(L,\omega)-D(L,\underline{\omega})]f_L(\omega)}{[D(H,\omega)-D(L,\underline{\omega})]f_H(\omega)}$ *is weakly increasing in* ω.

Conditions (i) to (iii) are more about intuitive ways to define the designer's type and the state: a higher designer's type weakly benefits the player, and weakly hurts the designer; and a higher state strictly benefits the player, weakly hurts the designer, and strictly enhances the total surplus. The opposite relationship can be accommodated by redefining variables. Condition (iv) has intuitive explanations. All terms are modified by $D(L, \underline{\omega})$, i.e., low-type designer's return at the lowest state. If $D(L, \underline{\omega}) = 0$, as is common in applications, the condition requires that when the state is higher, the ratio between the returns for the designer and player weakly increases, and the ratio between the low-type and high-type designer's return increases faster than the likelihood ratio. Condition (iv) is always satisfied if the designer's return does not depend on the state. Now we describe some applications of the general framework and explicitly restate the regularity conditions.

Application 1 Monopoly pricing*: Suppose a monopoly sells a single unit product to a buyer. Here* ω *represents the valuation of the product to the buyer, and* θ *represents the quality of the product, which only influences the distribution of* ω*. The seller only cares about the payment from the buyer. The production cost is normalized to be zero.*

[6] If the outside option is nonzero as in Akerlof (1978), then what matters is the payoff difference between alternatives.

In this case, we have $P(\theta, \omega) = \omega$ and $D(\theta, \omega) = 0$. The regularity condition is always satisfied.

Application 2 Procurement*: Suppose a procurer purchases a product from a supplier. Here ω represents the supplier's production efficiency, and θ represents the simplicity in producing the required product. The procurer's valuation for the product is V, and it costs the supplier $c(\theta, \omega)$ to produce the product. Therefore, we have $P(\theta, \omega) = -c(\theta, \omega)$ and $D(\theta, \omega) = V$. The regularity conditions reduce to (i) $-c(L, \underline{\omega}) + V \geq 0$, (ii) $c(\theta, \omega)$ is weakly decreasing in θ, and strictly decreasing in ω.*

Application 3 Hiring: *Suppose a firm wants to hire a CEO. Here ω represents the CEO's productivity in the firm. And θ represents the multiplicative inverse of the firm's profitability. If the CEO is hired, the firm obtains a revenue $\frac{1}{\theta}$, and the CEO incurs a cost $c(\omega)$ in running the firm. Therefore, we have $P(\theta, \omega) = -c(\omega)$ and $D(\theta, \omega) = \frac{1}{\theta}$. The regularity conditions reduce to (i) $c(\underline{\omega}) \leq \frac{1}{H}$, (ii) $c(\omega)$ is strictly decreasing in ω, (iii) $H \geq L$.*

Application 4 Regulating a monopoly: *Suppose a government decides whether to allow a monopoly to produce. Here ω represents the monopoly's production efficiency, and θ represents the government's private information about the industry production efficiency that is informative about ω. If the monopoly produces, it generates a revenue of R by incurring a cost $c(\theta, \omega)$, and produces a consumer surplus CS. The government cares about the consumer surplus and the transfer from the regulated firm. Therefore, $P(\theta, \omega) = R - c(\theta, \omega)$ and $D(\theta, \omega) = CS$. The regularity conditions reduce to (i) $R + CS \geq c(\theta, \omega)$, (ii) $c(\theta, \omega)$ is weakly decreasing in θ, and strictly decreasing in ω.*

Application 5 Emission quotas: *Suppose a government decides whether to issue a permission to a polluting firm. Here ω represents the polluting firm's revenue from production, and θ represents the government's private information about the abatement efficiency of the industry. If the government issues the permission, the polluting firm enjoys a revenue of ω, but generates a public environmental cost $c(\theta)$. The government* *cares about the transfer from the polluting firm and the environmental consequence. Therefore, $P(\theta, \omega) = \omega$ and $D(\theta, \omega) = -c(\theta)$. The regularity conditions reduce to (i) $\underline{\omega} \geq c(H)$, (ii) $c(H) \geq c(L)$.*

The direct grand mechanism

Following the informed principal literature, we allow the designer to propose a grand mechanism in which she herself is a participant. More specifically, the designer and the player engage in the following game. First, the designer proposes a grand mechanism, which is a continuation game that specifies (i) message spaces for both the designer and the player, and (ii) for each possible type of the designer, a rule to implement a mixed package of information disclosure and mechanism. Second, the player decides whether to participate. If he declines, the game ends; otherwise, the game goes to the next stage. Third, if the player participates, the grand mechanism is executed. By the revelation principle, for any on-equilibrium-path strategy, it is without loss of generality to focus on direct grand mechanisms in which the designer's and the player's message spaces are their type spaces.

For the information disclosure, we model it as Bayesian persuasion following Kamenica and Gentzkow (2011). A disclosure policy is a costless statistical experiment π, which is a family of conditional distribution $\pi(s|\omega)$ over a finite set of signal realization space S such that $\sum_{s \in S} \pi(s|\omega) = 1, \forall \omega$. Similar to Kamenica and Gentzkow (2011), the same results apply if π is a measurable function over a compact metric space S. This is because in the end we only need two signals. Let Π denote the set of all possible statistical experiments.

When the designer discloses information through Bayesian persuasion, we assume that the realized signal is only observable to the player. For instance, after users' trials of a game or software, sellers do not know whether users like the product or not; suppliers, CEOs, regulated firms and polluting firms also privately updates their information after information sessions. When the player reports to the designer that his realized signal is s, which need not be true, the designer implements alternative 1 with probability $q(s)$ for a payment $t(s)$ which could be either positive or negative.[7] The designer can also demand a monetary transfer I, from the player for participation. We assume an ex-ante participation constraint for the player. All our results hold with an ex-post participation constraint if the player can only observe the signal but not the state ex-post. This is because the ex-post participation constraint is a stronger requirement than the ex-ante participation constraint from the point view of the designer. As to be shown in the paper, the equilibrium strategies for both types satisfy the ex-post participation constraints. If I is positive (negative), we call it a participation fee (bonus). The mechanism is thus characterized by $\gamma = (I, q(s), t(s))_{s\in S}$. Let Γ denote the set of all possible mechanisms.

A package is a combination of one statistical experiment and one mechanism. A mixed package $\phi \in \Delta(\Pi \times \Gamma)$ is a probability measure on packages. The support of ϕ could be either continuous or discrete: when it is continuous, ϕ is a density function with respect to the Lebesgue measure on the Borel $\sigma-$field of $\Pi \times \Gamma$; when it is discrete, ϕ is a density with respect to the discrete uniform measure, and we adopt the convention that for any function ϑ on $\Pi \times \Gamma$:

$$\int \vartheta(\pi, \gamma) \phi(\pi, \gamma) d\pi d\gamma = \sum\nolimits_{\pi,\gamma \in \mathrm{supp}(\phi)} \vartheta(\pi, \gamma) \phi(\pi, \gamma). \tag{1}$$

We assume that two mixed packages that differ in zero measure will be perceived identically by the player.[8]

A direct grand mechanism is a menu of mixed packages, one for each type of designer, denoted as $\Phi = \{\phi_\theta\}_{\theta=L,H}$. Each type needs to specify the mixed package she would adopt if her type were both L and H. Therefore, when the designer adopts the direct grand mechanism Φ, she states that "If I am type L, I would implement ϕ_L; and if I am type H, I would implement ϕ_H".

Timing

[7] Since all parties are risk neutral, all monetary transfers matter only in expectation.

[8] This is mainly for expositional convenience. Otherwise, for any equilibrium, there exist infinite many other equilibria that differ in zero measure.

The timing of the game is as follows.

• The nature draws a private type for the designer.

• The designer proposes a direct grand mechanism Φ.

• The player observes the designer's choice and decides whether to participate.

• If the player does not participate, the game ends; otherwise, the direct grand mechanism is executed, within three stages.

• In the first stage (interim stage), given the designer's reported type θ, a package (π, γ) is publicly chosen according to the mixed package $\phi_\theta(\pi, \gamma)$.

• In the second stage, a monetary transfer I is made from the player to the designer for participation, and a signal is generated according to π. The player observes the signal privately and reports to the designer, say s.

• In the third stage (posterior stage), the designer implements alternative 1 with probability $q(s)$ for a payment $t(s)$.

Payoffs

First look at the player's payoff. In the third stage, given the designer's type θ, the statistical experiment π, and the realized signal s, the player's posterior return of the alternative 1 is

$$V(\theta, s, \pi) = \frac{\int_{\underline{\omega}}^{\overline{\omega}} P(\theta, \omega)\pi(s|\omega) f_\theta(\omega) d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) f_\theta(\omega) d\omega}. \tag{2}$$

When the player holds an interim belief ξ about the designer's type, with a slight misuse of notation, the player's expected posterior return of the alternative is

$$V(\xi, s, \pi) = \frac{\xi \int_{\underline{\omega}}^{\overline{\omega}} P(H, \omega)\pi(s|\omega) f_H(\omega) d\omega + (1-\xi) \int_{\underline{\omega}}^{\overline{\omega}} P(L, \omega)\pi(s|\omega) f_L(\omega) d\omega}{\xi \int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) f_H(\omega) d\omega + (1-\xi) \int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) f_L(\omega) d\omega}. \tag{3}$$

In the second stage, given the designer's type θ, the probability of generating signal s is

$$g(\theta, s, \pi) = \int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) f_\theta(\omega)\, d\omega. \tag{4}$$

Thus, given the designer's type θ, a truthful player's interim payoff in the first stage is

$$u\left(\theta,\pi,\gamma\right)=\sum_{s\in S}\left[q\left(s\right)\int_{\underline{\omega}}^{\overline{\omega}}P(\theta,\omega)\pi(s|\omega)f_{\theta}(\omega)d\omega-t\left(s\right)g\left(\theta,s,\pi\right)\right]-I. \tag{5}$$

Given ϕ and the designer's type θ, a truthful player's expected payoff is

$$U(\theta,\phi)=\int u\left(\theta,\pi,\gamma\right)\phi\left(\pi,\gamma\right)d\pi d\gamma. \tag{6}$$

Now let us look at the designer's payoff at various stages. Type-θ designer's interim payoff is

$$r_{\theta}(\pi,\gamma)=\sum_{s\in S}\left[t\left(s\right)g\left(\theta,s,\pi\right)+q\left(s\right)\int_{\underline{\omega}}^{\overline{\omega}}D(\theta,\omega)\pi(s|\omega)f_{\theta}(\omega)d\omega\right]+I. \tag{7}$$

Type-θ designer's expected payoff is

$$R_{\theta}(\phi)=\int r_{\theta}(\pi,\gamma)\phi\left(\pi,\gamma\right)d\pi d\gamma. \tag{8}$$

3 The full information benchmark

We first examine the benchmark where the designer's type is commonly known. It is clear that the designer's payoff cannot be higher than what she can achieve from a fully efficient outcome with full surplus extraction. For a fully efficient outcome, she should always implement alternative 1 due to the regularity condition (i) in Definition 1. This results in a total surplus of $\int_{\underline{\omega}}^{\overline{\omega}}T(\theta,\omega)f_{\theta}(\omega)d\omega$. To extract the full surplus, the designer can adopt no disclosure and charge a participation fee $\int_{\underline{\omega}}^{\overline{\omega}}P(\theta,\omega)f_{\theta}(\omega)d\omega$. This is summarized in the following proposition.

Proposition 1 *In the full information benchmark, it is optimal for each type of the designer to disclose no information and charge a participation fee* $\int_{\underline{\omega}}^{\overline{\omega}}P(\theta,\omega)f_{\theta}(\omega)d\omega$*.*

This proposition is important for us to understand the distortion introduced by the privacy of designer's information. Now we go back to the original model. In what follows, we first introduce the perfect Bayesian equilibrium (PBE) and establish an important result to simplify the problem. Second, we solve the RSW mechanism. Third, we establish the robustness of the RSW mechanism.

4 PBE

In a PBE, (1) both the designer and the player update their beliefs according to Bayes' rule wherever possible; (2) given their beliefs, the designer chooses a direct grand mechanism optimally, and both the designer and the player report optimally. By the inscrutability principle in Myerson (1983), it is without loss of generality to assume that both types of designer choose the same direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$. Note that the mixed package to be implemented still depends on the type of designer. The idea is that any information that can be revealed through the choice of direct grand mechanism can be revealed through the mixed package to be implemented following the designer's report. We refer to ϕ_θ as type$-\theta$ designer's mixed package.

We do not aim to characterize every PBE since our focus is the RSW mechanism. Instead, we consider a simplified game by imposing some restrictions on the direct grand mechanism, and show that it is outcome equivalent to the original game.

Definition 2 *In a simplified game, the designer randomizes only on packages where*

(a) the signal realization space for the statistical experiment is binary, i.e., $S = \{s_1, s_2\}$*,*

(b) the choice of the alternative in the mechanism is deterministic with $q(s_1) = 1$ *and* $q(s_2) = 0$*,*

(c) if only one signal $i \in \{1, 2\}$ *is used in the statistical experiment, the mechanism has zero payment conditional on* s_i*,* $t(s_i) = 0$*; if both signals are used, the mechanism has zero payment conditional on* s_2*,* $t(s_2) = 0$*.*

In the simplified game, the direct grand mechanism for the designer is much simpler. For the statistical experiment, the designer needs only to decide how to divide the states into two partitions. If only one signal is used in the statistical experiment, the mechanism conditional on the unused signal is not well defined. We thus have two different situations. If both signals are used, the designer needs to determine I and $t(s_1)$; otherwise, the designer needs only to determine I. Note that the direct grand mechanism in the simplified game is still flexible since I can be none zero and the designer can randomize.[9]

Definition 3 *Two outcomes are equivalent if each type of designer achieves the same expected payoff, implements alternative* 1 *with the same probability, and the player gains the same expected payoff given the designer's type.*

The equivalency is imposed not only on the designer's expected payoff, but also on the choice of alternative and the player's expected payoff, and therefore, is in a strong form. The following proposition states the equivalency result.

Proposition 2 *For any PBE in the original game, there exists an outcome equivalent PBE in the simplified game, and vice versa.*

The equivalency is intuitive. Similar to Myerson (1985), any choice rules of alternatives that are incentive compatible for the player "can be approximated arbitrarily closely (except possible on a countable set) by a convex combination of the deterministic allocation rules without changing expected payoff". We provide a formal proof of this statement in our setup. With deterministic choice rules, we can combine all signals that result in the same choice of alternative into a unique signal. By applying the revelation principle in a similar way to Kamenica and Gentzkow (2011), we can restrict ourselves to straightforward experimental statistics. With binary alternatives, we can restrict to binary signal space, denoted as $S = \{s_1, s_2\}$, i.e., (a) in Definition 2. Since the roles of the two signals can always be exchanged, we refer to them as s_1 and s_2, such that $q(s_1) = 1$ and $q(s_2) = 0$, respectively, i.e., (b) in Definition 2. Furthermore, whether a constant payment is made before or after information is disclosed makes no difference due to risk neutrality, which implies (c) in Definition 2.

In the original game, while more PBE can be identified, the set of outcomes is the same as that of the simplified game; while more off-equilibrium-path deviations have to be considered, it is sufficient to consider those satisfying the restrictions in

[9]For the application to monopoly pricing, the simplified game is different from Chen and Zhang (2020) where $I = 0$ and the seller does not randomize.

the simplified game. As a result, if we care only about the equilibrium outcome, it is without loss of generality to examine the simplified game. This proposition will help us to establish similar equivalency when we consider the RSW mechanism.

We introduce a class of statistical experiments that is of significance: a monotone binary partition with cutoff y, denoted by $\pi(y)$. With this statistical experiment, signal s_1 will be realized if the state is higher than y; otherwise, s_2 will be realized, i.e.,

$$\pi(s_1|\omega) = \begin{cases} 1, & \text{if } \omega \geq y; \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

By Definition 2, a mixed package with monotone binary partition $\pi\left(\underline{\omega}\right)$ always implements alternative 1 and requires zero payment; for a mixed package with monotone binary partition $y > \underline{\omega}$, it implements alternative 1 if and only if $\omega \geq y$, and the payment when implementing alternative 0 is zero. Since the total surplus from alternative 1 is always greater than that from alternative 0, $\pi(\underline{\omega})$ can be called the efficient disclosure. The higher y is, the less efficient the statistical experiment will be. Denote the set of all monotone binary partitions as Π^M.

5 The RSW mechanism

As discussed in the introduction, the most prominent mechanism to examine is the RSW (Rothschild-Stiglitz-Wilson) mechanism introduced by Maskin and Tirole (1992). In our paper, the RSW mechanism can always be supported as a PBE and survives the intuitive criterion as to be shown in Section 6. We first reproduce its definition in our setup below.

Definition 4 *A direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$ is a safe mechanism if*

$$R_\theta(\phi_\theta) \geq R_\theta(\phi_{\theta'}), \forall \theta \neq \theta', \tag{10}$$

$$\left[q\left(s\right) - q\left(\hat{s}\right)\right] V(\theta, s, \pi) \geq t\left(s\right) - t\left(\hat{s}\right), \forall s, \hat{s}, \theta, \tag{11}$$

$$U(\theta, \phi_\theta) \geq 0, \forall \theta. \tag{12}$$

Constraint (10) is the designer's incentive compatibility constraint. Constraint (11) is the player's incentive compatibility constraint when he knows the designer's type.[10] Constraint (12) is the player's individual rationality constraint when he knows the designer's type. Constraints (11) and (12) are often termed as the player's full-information IC and IR constraints.

Definition 5 *A direct grand mechanism $\Phi^W = \{\phi_\theta^W\}_{\theta=L,H}$ is a RSW mechanism if $\forall\theta$,*

$$R_\theta^W = R_\theta(\phi_\theta^W) = \max_{\Phi} R_\theta(\phi_\theta), \text{ s.t. } \Phi \text{ is a safe mechanism.} \tag{13}$$

In the RSW mechanism, the two types of designer simultaneously maximize their own payoffs among the safe mechanisms, given the mixed package adopted by the other type. This is also known as the best safe mechanism.

5.1 Simplify the problem

The RSW mechanism is difficult to solve since we need to find a fixed point. The following proposition shows that the RSW mechanism can be obtained successively.

Proposition 3 *The RSW mechanism can be obtained by successively solving Problem P_L :*

$$R_L^W = R_L(\phi_L^W) = \max_{\phi_L} R_L(\phi_L), \tag{14}$$

$$s.t. \quad [q(s) - q(\hat{s})]\, V(L, s, \pi) \geq t(s) - t(\hat{s}), \forall s, \hat{s}, \tag{15}$$

$$U(L, \phi_L) \geq 0; \tag{16}$$

[10] Strictly speaking, this is $\phi_\theta(\pi,\gamma)[q(s)V(\theta,s,\pi) - t(s)] \geq \phi_\theta(\pi,\gamma)[q(\hat{s})V(\theta,s,\pi) - t(\hat{s})], \forall s, \hat{s}$, since it only needs to be satisfied for the package in the support of the direct grand mechanism. When $\phi_\theta(\pi,\gamma) = 0$, the constraint is always satisfied; when $\phi_\theta(\pi,\gamma) > 0$, it reduces to the expression above. Furthermore, if a statistical experiment uses a single signal, this constraint disappears since the player cannot misreport. Throughout the paper, we follow this interpretation for all player's incentive compatibility constraints to avoid cumbersome notations.

and Problem P_H :

$$R_H^W = R_H(\phi_H^W) = \max_{\phi_H} R_H(\phi_H), \tag{17}$$

$$s.t. \quad R_L^W \geq R_L(\phi_H), \tag{18}$$

$$[q(s) - q(\hat{s})] V(H, s, \pi) \geq t(s) - t(\hat{s}), \forall s, \hat{s}, \tag{19}$$

$$U(H, \phi_H) \geq 0. \tag{20}$$

Compared with the original problem, the successive problem differs in two aspects, both relating to the mimicking constraint (10). First, for Problem P_L, the low-type designer no longer needs to consider the high-type designer's mimicking incentive. As a result, her problem is equivalent to the case where her type is known by the player. Second, in Problem P_H, the low-type designer's on-path payoff is now replaced by the solution from Problem P_L. From this successive form, it is clear that the RSW mechanism corresponds to the least costly separating mechanism—i.e., Riley outcome in the signaling literature. Maskin and Tirole (1992) show that when a sorting assumption holds, the RSW mechanism can be solved successively. Unfortunately, this sorting assumption fails in our setup.[11] We establish the above proposition by constructing a lower bound of the high-type designer's payoff and showing that it is higher than her payoff from mimicking the low-type designer's strategy solved in Problem P_L.

With Proposition 3, the designer's direct grand mechanism is still quite flexible. The following proposition shows that we can restrict ourselves to the simplified game if we care only about outcomes, similarly to Proposition 2 for PBE.

Proposition 4 *For any RSW mechanism in the original game, there exists an outcome-equivalent RSW mechanism in the simplified game which further does not randomize on* I*, and vice versa.*

Thus, if we focus on outcome, it is without loss of generality to consider the simplified game with a further simplification that a mixed package does not randomize on I.

[11] A discussion of the sorting assumption is provided in Appendix D.

For PBE, we cannot impose this further restriction since it results in loss of generality when the designer partially pools on packages. This restriction can be imposed for the RSW mechanism since it corresponds to a separating equilibrium. One may ask whether we can restrict the game even further and establish the equivalency result. For example, in addition we can restrict the game such that the designer sets $I = 0$ or does not randomize. As we will see later, the equivalency fails since the opposite will be the features of the RSW mechanism in general.

Whenever we add "simplified" in front of a terminology, it relates to the simplified game. Let Π^B and Γ^D denote the set of simplified statistical experiments and simplified mechanism. Denote $t(s_1) = \bar{t}$. A simplified mixed package can now be represented by $(\varsigma(\pi, \bar{t}), I) \in \Pi^B \times \Gamma^D$, where ς is a probability measure on binary partitions and the payment when implementing alternative 1. Participation transfer I is outside of ς because of the further restriction in Proposition 4.

5.2 Problem P_L

By utilizing Proposition 4, we can rewrite Problem P_L in Proposition 3:

$$R_L^W = R_L(\varsigma_L^W, I_L^W) = \max_{\varsigma_L, I_L} R_L(\varsigma_L, I_L), \tag{21}$$

$$s.t. \quad V(L, s_1, \pi) \geq \bar{t} \geq V(L, s_2, \pi), \tag{22}$$

$$\int g(L, s_1, \pi) \left[V(L, s_1, \pi) - \bar{t} \right] \varsigma_L(\pi, \bar{t}) d\pi d\bar{t} - I_L \geq 0. \tag{23}$$

Note that for statistical experiments using a single signal, the player's IC constraint (22) disappears since the player cannot misreport his observed signal. Problem P_L is equivalent to the low-type designer's payoff maximization problem if her type is observable by the player. It is clear that her payoff cannot be higher than what she can achieve from a fully efficient outcome with full surplus extraction. For a fully efficient outcome, she should always implement alternative 1. According to the definition of the statistical experiment, s_1 should always be sent, i.e., $\pi(\underline{\omega})$, implying that the statistical experiment is unique and discloses efficient information. This generates an expected return of $\int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega$ for the player. To extract the full surplus, she sets the participation fee as equal to the player's expected return. This is summarized in the following proposition.

Proposition 5 *In the RSW mechanism, the low-type designer's mixed package is unique and involves a single package: an efficient statistical experiment* $\pi(\underline{\omega})$ *with a participation fee* $I_L^W = \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega$.

In our framework, no disclosure is an efficient statistical experiment. This proposition implies that there is no distortion for the low-type. In contrast to Proposition 1, the uniqueness can be established since we are now restricted to simplified games.

5.3 Problem P_H

Here is the road map to solve Problem P_H. First, we show that for each package, it has to be a monotone binary partition and the payment for alternative 1 equals the player's conditional expected return (Lemma 1). Second, we show that the low-type designer's mimicking constraint must be binding (Lemma 2). These two steps reduce the high-type designer's problem to determining the randomization on cutoffs for monotone binary partitions only, which becomes a modified concavification problem for a function with a single variable (Lemma 3). We then summarize the solution to Problem P_H in Proposition 6.

By utilizing Proposition 4, we can rewrite Problem P_H in Proposition 3:

$$R_H^W = R_H(\varsigma_H^W, I_H^W) = \max_{\varsigma_H, I_H} R_H(\varsigma_H, I_H), \tag{24}$$

$$s.t. \quad R_L^W \geq R_L(\varsigma_H, I_H), \tag{25}$$

$$V(H, s_1, \pi) \geq \bar{t} \geq V(H, s_2, \pi), \tag{26}$$

$$\int g(H, s_1, \pi) \left[V(H, s_1, \pi) - \bar{t}\right] \varsigma_H(\pi, \bar{t}) d\pi d\bar{t} - I_H \geq 0. \tag{27}$$

The following lemma establishes some necessary features of the high-type designer's mixed package implemented in the RSW mechanism.

Lemma 1 *In the RSW mechanism, the high-type designer's mixed package only randomizes on packages where*

(i) *the statistical experiment is a monotone binary partition* $\pi(y)$ *with* $y \in [\underline{\omega}, \overline{\omega}]$*,*

(ii) *for monotone binary partitions except* $\pi(\underline{\omega})$ *or* $\pi(\overline{\omega})$*, the payment for alternative* 1 *is equal to the player's conditional expected return* $V(H, s_1, \pi)$*:* $\bar{t} = \frac{\int_y^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1-F_H(y)}$.

Here is the intuition. (i) Given the high-type designer's probability of implementing alternative 1, a monotone binary partition maximizes the player's return and minimizes the low-type designer's probability of implementing alternative 1 due to likelihood-ratio dominance. This maximizes the player's willingness to participate and minimizes the low-type designer's incentive to mimic. (ii) The payment when implementing alternative 1 is such that the player who obtains alternative 1 is indifferent from outside option. In this way, the designer does not leave any surplus to the player who obtains alternative 1 after participation. Note that when $\pi = \pi(\underline{\omega})$ or $\pi(\overline{\omega})$, only one signal is used in the experimental statistic and we have $\bar{t} = 0$ according to the definition of the simplified game.

Lemma 1 simplifies Problem P_H significantly. The statistical experiment is determined by a single variable, the cutoff y, according to (i). Given the cutoff, $\bar{t}$ is uniquely determined according to (ii). As a result, choosing $\varsigma_H(\pi, \bar{t})$ is reduced to choosing the randomization on the cutoff for the monotone binary partition only. Let $\sigma_H(y)$ denote a probability measure on cutoffs $y \in [\underline{\omega}, \overline{\omega}]$.[12] Then in Problem P_H we only need to determine the optimal $\sigma_H(y)$ and I_H. The following lemma implies that I_H can be fully determined by $\sigma_H(y)$.

Lemma 2 *The low-type designer's mimicking constraint* (25) *must be binding.*

This is intuitive. Otherwise, the high-type designer can always achieving a higher payoff by either increasing the participation fee or choosing alternative 1 more often.

[12] For discrete distribution we adopt the convention as in (1).

By (25) , we have

$$I_H = R_L^W - \int_{y\in(\underline{\omega},\overline{\omega})} \left[\frac{1-F_L(y)}{1-F_H(y)} \int_y^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega + \int_y^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega\right] \sigma_H(y) dy. \tag{28}$$

Therefore, the remaining problem is to determine the optimal $\sigma_H(y)$. Let

$$N(y) = \int_y^{\overline{\omega}} \frac{1-F_L(y)}{1-F_H(y)} P(H,\omega) f_H(\omega) d\omega + \int_y^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega \tag{29}$$

$$M(y) = \int_y^{\overline{\omega}} \left[\frac{F_L(y)-F_H(y)}{1-F_H(y)} P(H,\omega) + D(H,\omega)\right] f_H(\omega) d\omega - \int_y^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega \tag{30}$$

The term $N(y)$ is the low-type designer's mimicking payoff excluding participation transfer, and $M(y)$ is the difference between the high-type designer's payoff and the low-type designer's mimicking payoff. It can be shown that $N(y)$ is strictly decreasing in y. As a result, the inverse function of $N(y)$ exists and we have $M(y) = M(N^{-1}(N(y)))$. Denote $B = M(N^{-1})$ and $x = N(y)$. Since $N(y)$ is strictly decreasing, choosing a randomization on the cutoff y is equivalent to choosing a randomization on x. Let $\kappa_H(x)$ denote a probability measure for x on $[N(\overline{\omega}), N(\underline{\omega})]$. The following lemma shows how to determine the optimal $\kappa_H(x)$.

Lemma 3 *The optimal $\kappa_H(x)$ is determined by solving*

$$\max_{\kappa_H} \int_{N(\overline{\omega})}^{N(\underline{\omega})} B(x)\kappa_H(x) dx \tag{31}$$

$$s.t. \qquad R_L^W \leq \int_{N(\overline{\omega})}^{N(\underline{\omega})} x\kappa_H(x) dx. \tag{32}$$

This formula has a natural interpretation. The high-type designer chooses the lower-type designer's mimicking payoff excluding the participation transfer to maximize the difference between her own payoff and the low-type designer's mimicking payoff, such that the low-designer's on-path payoff is less than the mimicking payoff excluding the participation transfer.

If the constraint (32) binds, the above problem is a standard concavification problem for a function with a single variable. While our problem involves some modification, it can be solved similarly. Let $\widehat{B}(x)$ denote the concave closure of $B(x)$:

$$\widehat{B}(x) = \sup\{z | (x, z) \in co(B)\} \tag{33}$$

where $co(B)$ defines the convex hull of the graph of B. Note that while we allow the state to follow a general distribution, our concavification problem is captured by a single variable and has a graphical representation.[13] With a single variable, the optimal solution can always be achieved by mixing over at most two points, say $x_1 \leq x_2$ with probabilities η and $1 - \eta$ such that $x_1 + (1 - \eta)x_2 = R_L^W$. The solution to our modified concavification problem then depends on the relationship between R_L^W and the peak of $B(x)$ denoted as $x_H^{\#}$, which is always strictly higher than $N(\underline{\omega})$.[14] If $R_L^W < x_H^{\#}$, then the solution is $\kappa_H(x)$ degenerating to $x_H^{\#}$; otherwise, the solution is determined by the concavification. We now can summarize the solution to Problem P_H.

Proposition 6 *In the RSW mechanism, the following mixed package solves the high-type designer's problem.*

(i) When $R_L^W < x_H^{\#}$, it involves a single package: a monotone binary partition $\pi(N^{-1}(x_H^{\#}))$, a participation bonus $I_H^W = R_L^W - x_H^{\#} < 0$, and a payment $\bar{t}_H^W = \frac{\int_{N^{-1}(x_H^{\#})}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1 - F_H(N^{-1}(x_H^{\#}))}$ for implementing alternative 1.

(ii) When $R_L^W \geq x_H^{\#}$, it involves a (possible) randomization on at most two packages: for package $i \in \{1, 2\}$, it features a monotone binary partition with cutoff $N^{-1}(x_i)$, zero participation transfer $I_H^W = 0$, and a payment $\bar{t}_H^W = \frac{\int_{N^{-1}(x_i)}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1 - F_H(N^{-1}(x_i))}$ for implementing alternative 1.

The RSW mechanism is unique if and only if the modified concavification problem in Lemma 3 has a unique solution.

In the RSW mechanism, the high-type designer has three channels by means of which to separate: (C1) disclosing inefficient information, (C2) providing a bonus for

[13] In Kamenica and Gentzkow (2011), the concavification problem beyond binary states usually has higher dimensions and is much more difficult to solve. Significant progresses have been made in Dworczak and Martini (2019), Kolotilin (2018), and Kolotilin and Wolitzky (2020).

[14] If there are multiple peaks, let $x_H^{\#}$ be the one on the most right.

participation, and (C3) randomizing on packages. For channel (C1), although lower efficiency hurts the high-type designer, it hurts the low-type designer more due to the monotone likelihood ratio property. In some sense, it satisfies the property of increasing difference required for separation. For channel (C2), it hurts both types of designer in the same way as it entails merely a transfer to the player, and does not satisfy the increasing difference. As a result, it cannot achieve separation by itself. For Channel (C3), randomization on cutoffs already implies inefficiency, thus, (C1) becomes active. In summary, (C1) is active all the time; (C2) and (C3) cannot achieve separation alone, but can be used as supplements to (C1) since they influence the cost of separation.

To understand the two cases in the proposition, we consider the following interpretation. Let $F_1(\omega)$ and $F_2(\omega)$ be two distributions that satisfy the monotone likelihood ratio property: $\frac{f_1(\omega)}{f_2(\omega)} > \frac{f_1(\omega')}{f_2(\omega')}, \forall \omega > \omega'$. Let $P(1,\omega), D(1,\omega)$, $P(2,\omega)$ and $D(2,\omega)$ be return functions that satisfy the regularity conditions. Suppose for the high type $(F_H(\omega), P(H,\omega), D(H,\omega)) = (F_1(\omega), P(1,\omega), D(1,\omega))$; for the low type, $(F_L(\omega), P(L,\omega), D(L,\omega))$ is $(F_1(\omega), P(1,\omega), D(1,\omega))$ with probability of $\alpha \in [0,1)$, and $(F_2(\omega), P(2,\omega), D(2,\omega))$ with probability of $1-\alpha$. By construction, the high type is fixed, and when α increases, the low type gets closer to the high type.

Lemma 4 *There exists a unique cutoff α^* such that $R_L^W \geq x_H^{\#}$ if and only if $\alpha \geq \alpha^*$.*

Therefore, we can interpret the case of $R_L^W \geq x_H^{\#}$ as similar types and $R_L^W < x_H^{\#}$ as distinct types.[15] With similar types, the low-type designer's equilibrium payoff R_L^W is already high, making the unconstrained optimal solution $x_H^{\#}$ easier to violate Constraint (32), implying zero participation bonus. In contrast, with distinct types, the low-type designer's equilibrium payoff R_L^W is low, making the unconstrained optimal solution $x_H^{\#}$ satisfy Constraint (32) automatically, implying strictly positive participation bonus and no need for randomization. Now we can develop more intuitions.

[15] When $\alpha^* \leq 0$, only the case of similar types applies; when $\alpha^* > 1$, only the case of distinct types applies.

When qualities become more distinct, the low-type designer's equilibrium payoff becomes lower and she has a stronger incentive to mimic. This requires the high-type designer to introduce more distortion. Note that the marginal cost of a higher cutoff is increasing and the marginal cost of a higher participation bonus is a constant. As a result, when the low-type becomes too distinct, it is optimal to use a strictly participation bonus. The reason why randomization may be necessary is consistent with the optimality of partial disclosure in Bayesian persuasion literature. Randomization maintains the low-type designer's mimicking payoff, but can potentially increase the high-type designer's payoff. The roles of these three channels are summarized in the following corollary.

Corollary 1 *In the high-type designer's mixed package above, disclosing inefficient information always arises, and is thus essential. Providing a participation bonus is a necessary supplement when the designer has distinct types, and randomization may serve as a supplement when the designer has similar types. The two supplements do not appear at the same time.*

If general disclosure policies are allowed, a monotone binary partition can be replaced by an upper censorship in Kolotilin et al. (2017), or upper-censoring in Alonso and Câmara (2016). With this interpretation, monotone binary partitions can be ranked in terms of Blackwell informativeness: a higher cutoff means more Blackwell informative. The modified concavification problem is on a function with a single variable and it is easy to verify the uniqueness of the solution if we know the primitives.

The following example illustrate the effect of the closeness of types.

Example 1 *Suppose $P(\theta, \omega) = \omega$ and $D(\theta, \omega) = 0$ as in the monopoly pricing problem. Let $F_H(v) = v^2$, $F_L(v) = \alpha v^2 + (1-\alpha) v^{0.02}$, on the same support $[0, 1]$, and $\mu_H = 0.5$. We fix the high type, and increase the low type in terms of likelihood-ratio dominance. The two types become closer when α increases. The two different cases are divided according to $\alpha^* = 0.45$. As we can see in Figure 1, the cutoff for the monotone binary partition remains constant until α^* and then decreases, meaning that the information disclosure becomes less informative and trade occurs more often. Meanwhile,*

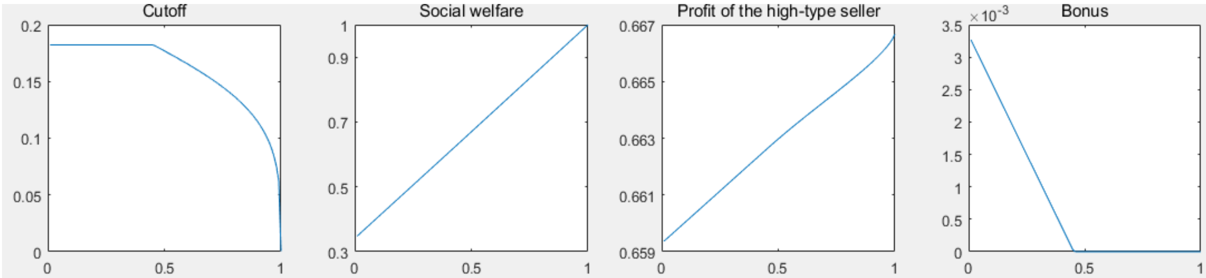

Figure 1: The impact of closeness of qualities

the high-type monopoly's expected payoff increases, along with the total surplus. However, the buyer's expected payoff decreases. A higher α can also be interpreted as a minimum quality standard. This implies that while a higher minimum quality standard increases the monopoly's payoff and total surplus, it hurts buyers.[16]

6 Robustness

While the RSW mechanism is the main focus in the literature of informed principal with common values, in this section we carry out some robustness checks by comparing it with other equilibrium concepts. Maskin and Tirole (1992) show that the RSW outcome is a PBE outcome and is the unique one under a sufficient condition. They also show that the RSW mechanism always satisfies the intuitive criterion and establishes its uniqueness under certain condition. One crucial assumption for achieving these results in their paper is the sorting assumption. In our setup, the sorting assumption fails. However, all of these desirable properties are preserved even without the sorting assumption.[17]

[16] The comparative statics results hold more generally. Since this is not the main focus of the paper, formal results are available upon request from the authors.

[17] Readers who regard RSW outcomes as the correct outcomes to examine can skip this section; readers interested in technical details can find the proofs for Proposition 7 and Proposition 8 in Appendix B.

6.1 RSW outcome as PBE outcome

Proposition 7 *The outcome of the RSW mechanism can be supported as a PBE outcome. Moreover, if $R_L^W > x_H^\#$ and $\mu_L^0 > 1 + \widehat{B}'\left(R_L^W\right)$, the set of outcomes from PBE equals that from the RSW mechanism.*

As is common in general informed principal problems, the challenge of demonstrating the first part of the proposition is that often the belief that makes a deviation undesirable must be tailored to the particular deviation. In our model, when $x_H^\# \geq R_L^W$, this issue does not arise and a universal belief of certainty of low-type can make all deviations undesirable. When $x_H^\# < R_L^W$, while the issue arises, it turns out that the RSW mechanism is interim efficient relative to a non-degenerate belief, and we can apply the well-known result of Theorem 1 in Maskin and Tirole (1992) directly. For the second part of the proposition, Maskin and Tirole (1992) also show that if the prior is such that the RSW mechanism is interim efficient, then the set of outcomes from PBE equals that from the RSW mechanism, which is met when the conditions in the proposition hold.[18]

6.2 Intuitive equilibrium

When the sufficient condition in Proposition 7 fails, multiple PBE outcomes may exist. To select the equilibrium outcome, we impose the intuitive criterion introduced by Cho and Kreps (1987). It requires that a reasonable off-equilibrium-path belief assigns zero probability to those types who are strictly worse off than their equilibrium payoff. The outcomes that survive the intuitive criterion are called intuitive outcomes. The following proposition shows that the outcome of the RSW mechanism always survives the intuitive criterion and is the unique outcome under certain condition.

Proposition 8 *The outcome of the RSW mechanism survives the intuitive criterion. Furthermore, with similar types, i.e., $R_L^W \geq x_H^\#$, the set of intuitive outcome coincides with that from the RSW mechanism.*

[18] We need to modify their proofs slightly to allow multiple dimensional action space.

7 Applications and extensions

7.1 An application to Chen and Zhang (2020)

Chen and Zhang (2020) restrict to monopoly pricing. Furthermore, instead of allowing arbitrary mechanisms, it is assumed that the seller makes a take-it-or-leave-it offer to the buyer. A natural question is whether the seller can be better off when she can adopt any mechanism. This question can be answered as a very special case of our general model with $P(\theta, \omega) = \omega$ and $D(\theta, \omega) = 0$. Proposition 6 implies that this often causes loss of generality. The following corollary provides a necessary and sufficient condition under which the RSW mechanism coincides with the unique intuitive equilibrium identified in that paper.

Corollary 2 *If and only if the designer's types are similar and no randomization is needed from the concavification problem, the unique intuitive equilibrium identified in Chen and Zhang (2020) is robust to more general mechanisms.*

7.2 Interpretation of the channels in applications

Recall the three channels for separating we have identified. Here we would connect them to phenomenons in practice. Channel (C1) implies that high-type designers reveal more than efficient information, which coincides with no disclosure in our setup. In practice, it is rather common for designers to allow players to learn more information: trial versions of software and video games, trailers for movies, information sessions for procurement, campus visits for hiring assistant professors in economics, etc. (C2) implies that high-type designers provide bonus to players for their learning. In practice, it is common to see a cocktail party with a lottery following an information session about new apartments presented by real estate agents, or small gifts following trials in a tutorial market. Governments also provide various benefits to participants in foreign investment information session. (C3) implies that high-type designers randomize on packages, which can be interpreted as promotional strategies as in Gal-Or (1982), Narasimhan (1988), Sobel (1984) and Varian (1980). The package with $N^{-1}(x_1)$ discloses more information and charges a higher price than the package with $N^{-1}(x_2)$. This is consistent with the practice in which products on sale are often limited-time offers and consumers can learn only limited information about the products.

When the designer has the full power to disclose information and choose mechanisms, it may seem irrational to provide a bonus for trials and to adopt promotional strategies. Our results provide a potential explanation for these scenarios: in order for the high-type designer to separate from the low-type, bonuses for trials are necessary when the designer's types are distinct, and promotional strategies may be needed when the designer's types are similar.

7.3 Beyond binary alternatives: linear payoffs

When the alternatives are more than two, let $\mathcal{K} = \{0, 1, \cdots, K\}$. Given alternative k, let $P^k(\theta, \omega)$, $D^k(\theta, \omega)$, and $T^k(\theta, \omega)$ be the player's return, the designer's return, and the total surplus, respectively. We assume that, $\forall k \in \mathcal{K}$, $P^k(\theta, \omega) = P(\theta, \omega) k$ and $D^k(\theta, \omega) = D(\theta, \omega)k$, which are both linear in k. The total surplus is then $T^k(\theta, \omega) = T(\theta, \omega) k$ with $T(\theta, \omega) = P(\theta, \omega) + D(\theta, \omega)$. The regularity conditions, i.e., Definition 1, are now imposed on the highest alternative K. Let $q(k, s)$ denote the probability of choosing alternative k when the player reports s.

Proposition 9 *With linear payoffs, in terms of the designer's and the player's expected payoffs given the designer's type, it is without loss of generality to assume that the designer only chooses between alternative* 0 *and* K*. The problem then reduces to binary alternatives.*

Note that the outcome equivalence here is weaker than that in Definition 3 since two equivalent outcomes can potentially have different probabilities for implementing a certain alternative. By letting $q(s) = \sum_{k=1}^{K} q(k, s)k$, the proof is similar to Proposition 2.

8 Conclusion and discussion

Providing incentives and revealing information are two common practices through which designers achieve their goals. However, when a designer also holds unverifiable private information, her strategy itself could signal her private information. In this paper, we place no restriction on what designers can do and provide a prediction for the outcome by examining the RSW mechanism. We show that the low type achieves her full-information outcome by disclosing efficient information and implementing an efficient mechanism that leaves the player with zero surplus. In contrast, the high type is forced to inefficiently disclose more information and sometimes to leave strictly positive surplus to the player, or to randomize on packages in order to separate from the low type. This outcome is robust.

Our results can be extended to allow for continuous type space of the designer ordered by the likelihood-ratio dominance. The proof is tedious but essentially follows a similar logic. We conjecture that a condition can be found to guarantee that local incentive constraints are necessary and sufficient. As a result, the problem for the lowest type is the same as Problem P_L; and the problem for all other types is similar to Problem P_H. In the RSW mechanism, the lowest type of designer achieves her full information outcome. The other types (possibly) randomize on packages with a monotone binary partition, a payment equal to the conditional expected return, and a (possible) participation bonus.

One direction of future research is to allow multiple players. Bayesian persuasion with multiple receivers (players) is also known as information design. In general, as pointed out by Kamenica and Gentzkow (2011), '... the key simplifying step in our analysis reducing the problem of finding an optimal signal to one of maximizing over distributions of posterior beliefs does not apply'. While some progress has been made, such as by Bergemann and Morris (2016,2019), Mathevet et al. (2020) and Taneva (2019), it is generally difficult to characterize the optimal information disclosure explicitly, even when the sender has no private information.

Another extension is to allow the player to have private information to begin with. In cases where the designer does not hold private information, the question of how to design optimal information disclosure and optimal mechanism jointly has received significant attention recently for the selling problem. Krähmer (2020) offers an excellent overview of the results. He shows that the conclusion relies on conditions. The first condition is whether the designer has full or partial information control. Full information control (FIC) means that the information that can be disclosed by the designer fully identifies the player's valuation, and the player's initial private information only affects his belief of the valuation. Otherwise, there is partial information control (PIC). The second condition relates to whether the information controlled by the designer is orthogonal to the player's private information. Eso and Szentes (2007) show that with PIC and an orthogonal structure, full information is optimal. Li and Shi (2017) construct a discriminatory disclosure that improves on the full disclosure with FIC and nonorthogonal structure. Krähmer (2020) further demonstrates that, in fact, the first-best surplus can be extracted within Li and Shi's framework. With PIC, full surplus extraction can be achieved with "full rank" condition. The optimal mechanism and information disclosure remain unknown when full surplus extraction is not achievable. Adding the private information of the designer results in another layer of complication, since now each type of designer needs to take into consideration other types' incentive to mimic.

Finally, for each specific application beyond binary alternatives, while our linear payoff structures can serve as a benchmark to provide simple answers, it is of particular interest to consider more details motivated by practical observations leading to nonlinear payoff structures. These issues will be left for future investigation.

Appendix A: Omitted Proofs

Proof for Proposition 2

We first introduce the concept of feasible direct grand mechanism. Suppose the player's belief is μ_H before seeing the designer's package (π, γ), his interim belief after will be updated according to Bayes' rule: $\xi(\pi, \gamma, \mu_H) = \frac{\mu_H \phi_H(\pi,\gamma)}{\sum_\theta \mu_\theta \phi_\theta(\pi,\gamma)}$.

Definition 6 *A direct grand mechanism* $\Phi = \{\phi_\theta\}_{\theta=L,H}$ *is* μ_H- *feasible if and only if*

$$R_\theta(\phi_\theta) \geq R_\theta(\phi_{\theta'}), \forall \theta \neq \theta', \tag{34}$$

$$[q(s) - q(\widehat{s})] \, V(\xi(\pi, \gamma, \mu_H), s, \pi) \geq t(s) - t(\widehat{s}), \forall s, \widehat{s}, \tag{35}$$

$$\sum_\theta \mu_\theta U_\theta(\Phi) \geq 0. \tag{36}$$

By the inscrutability principle of **Myerson (1983)** and the revelation principle, any PBE outcome can be implemented by a $\mu_H^0 - feasible$ direct grand mechanism. However, the reverse does not necessarily hold. Therefore, our proof is in two steps. In step 1, we show that for any feasible direct grand mechanism in the original game, there exists an outcome-equivalent feasible direct grand mechanism in the simplified game, and vice versa. In step 2, we show that for any feasible direct grand mechanism in the original game that can be supported as a PBE, its outcome-equivalent feasible direct grand mechanism in the simplified game can be supported as a PBE, and vice versa.

Step 1: The direction from the simplified game to the original game is straightforward, since any $\mu_H^0 - feasible$ direct grand mechanism in the simplified game is also $\mu_H^0 - feasible$ in the original game. We show the opposite direction by construction. Suppose we have a $\mu_H^0 - feasible$ direct grand mechanism in the original game $\{\phi_\theta(\pi, \gamma)\}_{\theta=L,H}$. Pick up an arbitrary package in the direct grand mechanism (π, γ), we will replace it with a mixture among simplified packages constructed as follows. The signal realization space for π is arbitrary and we can categorize the signals according to their probabilities of choosing alternative 1. Suppose there are Q different possible probabilities that can arise in the package, we can order them from low to high according to $q_1 < q_2 < \cdots < q_Q$. By (35), if two signal realizations result in the same probabilities, say q_i, they also result in the same payment for alternative 1, denoted as t_i. Now we can construct $Q + 1$ simplified packages recursively as follows. The 1^{st}

simplified package $\{\pi^1, \gamma^1\}$ is defined as $\pi^1(\overline{s}|\omega) = 1$, $q^1(\overline{s}) = 1$, $t^1(\overline{s}) = 0, I^1 = t_1 + I$. $\forall i \in \{2, \cdots, Q\}$, the i^{th} simplified package $\{\pi^i, \gamma^i\}$ is defined as:

$$\begin{aligned} Statistical\ experiment \ &: \ \pi^i(\overline{s}|\omega) = \sum_{q(s)\geq q_i} \pi(s|\omega), \pi^i(\underline{s}|\omega) = 1 - \pi^i(\overline{s}|\omega) && (37) \\ Implemantation\ rule \ &: \ q^i(\underline{s}) = 0, q^i(\overline{s}) = 1 && (38) \\ Payment\ rule \ &: \ t^i(\overline{s}) = \frac{t_i - t_{i-1}}{q_i - q_{i-1}}, t^i(\underline{s}) = 0, I^i = t_1 + I && (39) \end{aligned}$$

Finally, define $Q+1$th simplified package $\{\pi^{Q+1}, \gamma^{Q+1}\}$ as $\pi^{Q+1}(\underline{s}|\omega) = 1$, $q^{Q+1}(\underline{s}) = 0$, $t^{Q+1}(\underline{s}) = 0$, $I^{Q+1} = t_1 + I$. Note that the 1st simplified package uses signal $\overline{s}$ only and the $Q+1$th one uses signal $\underline{s}$ only. Assign probability ι_i to the simplified package i, where

$$\iota_i = \begin{cases} q_1 & if\ i = 1 \\ q_i - q_{i-1} & if\ 2 \leq i \leq Q \\ 1 - q_Q & if\ i = Q+1 \end{cases} . \tag{40}$$

Replace each package (π, γ) with a mixture $\phi^{\pi,\gamma}(\widetilde{\pi}, \widetilde{\gamma})$ over simplified package $(\widetilde{\pi}, \widetilde{\gamma})$ such that

$$\phi^{\pi,\gamma}(\widetilde{\pi}, \widetilde{\gamma}) = \begin{cases} \iota_i & \text{if } (\widetilde{\pi}, \widetilde{\gamma}) = (\pi^i, \gamma^i) \\ 0 & \text{otherwise} \end{cases} . \tag{41}$$

We thus can construct $\phi'_\theta(\widetilde{\pi}, \widetilde{\gamma}) = \int \phi^{\pi,\gamma}(\widetilde{\pi}, \widetilde{\gamma})\, \phi_\theta(\pi, \gamma)\, d\pi d\gamma$. It remains to show that it is $\mu_H^0 - feasible$ and outcome equivalent to $\{\phi_\theta(\pi, \gamma)\}_{\theta = L,H}$.

Since (π, γ) is $\mu_H^0 - feasible$,

$$(q_i - q_{i-1})V(\xi(\pi, \gamma, \mu_H^0), s_i, \pi) \geq \quad t_i - t_{i-1}, \tag{42}$$

$$(q_{i-1} - q_i)V(\xi(\pi, \gamma, \mu_H^0), s_{i-1}, \pi) \geq \quad t_{i-1} - t_i. \tag{43}$$

These together imply that

$$\begin{aligned} & V(\xi(\pi, \gamma, \mu_H^0), s_{i-1}, \pi) \leq \frac{t_i - t_{i-1}}{q_i - q_{i-1}} \leq V(\xi(\pi, \gamma, \mu_H^0), s_i, \pi) \\ \Leftrightarrow \quad & V(\xi(\pi, \gamma, \mu_H^0), s_{i-1}, \pi) \leq t(\overline{s}) \leq V(\xi(\pi, \gamma, \mu_H^0), s_i, \pi). \end{aligned} \tag{44}$$

We also have

$$
\begin{aligned}
& V(\xi(\pi,\gamma,\mu_H^0),\overline{s},\pi^i) \\
= \; & \frac{\int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0) P\left(H,\omega\right)\pi^i(\overline{s}|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right] P\left(L,\omega\right)\pi^i(\overline{s}|\omega) f_L(\omega)\right\} d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0)\pi^i(\overline{s}|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right]\pi^i(\overline{s}|\omega) f_L(\omega)\right\} d\omega} \\
= \; & \sum_{q(s)\geq q_i} \frac{V(\xi(\pi,\gamma,\mu_H^0),s,\pi)}{\sum_{q(s)\geq q_i}\int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0)\pi(s|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right]\pi(s|\omega) f_L(\omega)\right\} d\omega} \\
& * \int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0)\pi(s|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right]\pi(s|\omega) f_L(\omega)\right\} d\omega \\
\geq \; & \sum_{q(s)\geq q_i} \frac{V(\xi(\pi,\gamma,\mu_H^0),s_i,\pi)}{\sum_{q(s)\geq q_i}\int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0)\pi(s|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right]\pi(s|\omega) f_L(\omega)\right\} d\omega} \\
& * \int_{\underline{\omega}}^{\overline{\omega}} \left\{\xi(\pi,\gamma,\mu_H^0)\pi(s|\omega) f_H(\omega) + \left[1-\xi(\pi,\gamma,\mu_H^0)\right]\pi(s|\omega) f_L(\omega)\right\} d\omega \\
= \; & V(\xi(\pi,\gamma,\mu_H^0),s_i,\pi).
\end{aligned}
\tag{45}
$$

Similar to (45), $V(\xi(\widetilde{\pi},\widetilde{\gamma},\mu_H^0),s,\widetilde{\pi})$ is a linear combination of $V(\xi(\pi,\gamma,\mu_H^0),s,\widetilde{\pi})$. By the construction of $(\widetilde{\pi},\widetilde{\gamma})$, (35) holds for $\xi(\pi,\gamma,\mu_H^0)$, and therefore, (35) holds for $\xi(\widetilde{\pi},\widetilde{\gamma},\mu_H^0)$.

For the expected selling probability, we have:

$$
\begin{aligned}
& \int_{\underline{\omega}}^{\overline{\omega}} \sum_{i=1,\cdots,Q+1} \iota_i \left[\pi^i(\underline{s}|\omega) q^i(\underline{s}) + \pi^i(\overline{s}|\omega) q^i(\overline{s})\right] f_\theta\left(\omega\right) d\omega \\
= \; & \int_{\underline{\omega}}^{\overline{\omega}} \sum_{i=1,\cdots,Q+1} \iota_i \pi^i(\overline{s}|\omega) f_\theta\left(\omega\right) d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \sum_{i=1,\cdots,Q} \sum_{q(s)\geq q_i} \iota_i \pi\left(s|\omega\right) f_\theta\left(\omega\right) d\omega \\
= \; & \int_{\underline{\omega}}^{\overline{\omega}} \sum_{s} \left(\sum_{q(s)\geq q_i} \iota_i\right) \pi\left(s|\omega\right) f_\theta\left(\omega\right) d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \sum_{s} \pi(s|\omega) q(s) f_\theta\left(\omega\right) d\omega.
\end{aligned}
\tag{46}
$$

For the expected payment, we have:

$$
\sum_{i=1,\cdots,Q+1} \iota_i \left\{\int_{\underline{\omega}}^{\overline{\omega}} \pi^i(\overline{s}|\omega) P\left(\theta,\omega\right) q^i(\overline{s}) f_\theta\left(\omega\right) d\omega - \left[\int_{\underline{\omega}}^{\overline{\omega}} \pi^i(\overline{s}|\omega) f_\theta\left(\omega\right) d\omega\right] t^i\left(\overline{s}\right) - I^i\right\}
$$

$$
\begin{aligned}
&= q_1 \int_{\underline{\omega}}^{\overline{\omega}} P(\theta,\omega) f_\theta(\omega)\, d\omega - t_1 - I \\
&\quad + \sum_{i=2,\cdots,Q} \iota_i \left\{ \int_{\underline{\omega}}^{\overline{\omega}} \pi^i(\overline{s}|\omega) \left[P(\theta,\omega) - t^i(\overline{s}) \right] f_\theta(\omega)\, d\omega \right\} \\
&= q_1 \int_{\underline{\omega}}^{\overline{\omega}} P(\theta,\omega) f_\theta(\omega)\, d\omega - I \\
&\quad + \sum_s \int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) \left\{ [q(s) - q_1] P(\theta,\omega) - t(s) \right\} f_\theta(\omega)\, d\omega \\
&= \sum_s \int_{\underline{\omega}}^{\overline{\omega}} \pi(s|\omega) [q(s) P(\theta,\omega) - t(s)] f_\theta(\omega)\, d\omega - I. \qquad (47)
\end{aligned}
$$

Similarly the designer's expected payoff is the same. Therefore, for any θ, $U(\theta, \phi'_\theta) = U(\theta, \phi_\theta)$, $R_{\theta'}(\phi'_\theta) = R_{\theta'}(\phi_\theta)$. As a consequence, (34) and (36) hold, and the constructed simplified direct grand mechanism is $\mu_H^0 - feasible$.

Step 2: Consider a $\mu_H^0 - feasible$ direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$ in the original game and a $\mu_H^0 - feasible$ direct grand mechanism $\Phi' = \{\phi'_\theta\}_{\theta=L,H}$ in the simplified game such that Φ and Φ' are outcome equivalent. Since any direct grand mechanism in the simplified game is also a direct grand mechanism in the original game, the set of off-equilibrium-path strategy in the simplified game is a subset of union of off-equilibrium-path strategy in the original game and the outcome-equivalent grand mechanisms to Φ'. Therefore, if $\Phi = \{\phi_\theta\}_{\theta=L,H}$ can be supported as a PBE in the original game, $\Phi' = \{\phi'_\theta\}_{\theta=L,H}$ can also be supported as a PBE in the simplified game. It remains to show the direction from the simplified game to the original game. We prove it by contradiction. Suppose in contrary, Φ' can be supported as a PBE in the simplified game, but Φ cannot be supported as a PBE in the original game. Then there exists an off-equilibrium-path direct grand mechanism Ψ in the original game such that a certain type of designer, say θ, has incentive to deviate. For any belief μ_H, by the revelation principle, Ψ is outcome-equivalent to a $\mu_H - feasible$ direct grand mechanism denoted as $\Phi^1(\Psi,\mu) = \left\{\phi_\theta^1(\Psi,\mu)\right\}_{\theta=L,H}$. Thus $R_\theta\left(\phi_\theta^1(\Psi,\mu)\right) > R_\theta(\phi_\theta)$. By arguments similar to Step 1, there exists an outcome-equivalent $\mu_H - feasible$ direct grand mechanism denoted as $\Phi^2(\Psi,\mu)$ in the simplified game. This implies

$R_\theta\left(\phi_\theta^2(\Psi,\mu)\right) = R_\theta\left(\phi_\theta^1(\Psi,\mu)\right) > R_\theta(\phi_\theta) = R_\theta(\phi'_\theta)$. As a result, the type-$\theta$ designer has incentive to deviate in the simplified game, a contradiction. **Q.E.D.**

Proof for Proposition 3

Let $\{\phi_\theta^W\}_{\theta=L,H}$ be an RSW mechanism and $\{\phi_\theta^*\}_{\theta=L,H}$ be a solution to the problem P_θ.

Step 1: we show that $\{\phi_\theta^*\}_{\theta=L,H}$ is a safe mechanism. Since $\{\phi_\theta^*\}_{\theta=L,H}$ satisfies the constraints in the successive problems, we only need to show that $R_H(\phi_H^*) \geq R_H(\phi_L^*)$. Since $R_H(\phi_H^*)$ solves Problem P_H, it is sufficient to show that ϕ_L^* satisfies the constraints in Problem P_H. First, it is obvious that (18) is satisfied since it holds with equality. Second, Problem P_L is equivalent to the full information benchmark in Proposition 1. Therefore, $\phi_L^*(\pi,\gamma) > 0$ only if $q(s) = 1, \forall s$, with payment $t(s) + I = \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega, \forall s$. Thus, (19) becomes $0 \geq 0$, which is satisfied. Finally, (20) is satisfied since

$$
\begin{aligned}
& U(H,\phi_L^*) \\
= & \int \left[\int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega + I - I\right] \phi_L^*(\pi,\gamma)\, d\pi d\gamma \\
= & \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega \geq 0. \qquad (48)
\end{aligned}
$$

Step 2: we show that ϕ_θ^W satisfies constraints of Problem I_θ. First, by Definition 4 and 5, ϕ_L^W satisfies all constraints in Problem P_L. Second, for any safe mechanism $\{\phi_L, \phi_H^W\}$, by Definition 5, $R_L(\phi_H^W) \leq R_L(\phi_L) \leq R_L(\phi_L^W) \leq R_L(\phi_L^*)$. Thus, ϕ_H^W satisfies constraint (18). By Definition 5, ϕ_H^W satisfies (19), (20) and therefore all constraints in Problem P_H.

Step 1 implies $R_\theta(\phi_\theta^W) \geq R_\theta(\phi_\theta^*)$, and **Step 2** implies $R_\theta(\phi_\theta^W) \leq R_\theta(\phi_\theta^*)$. Therefore $R_\theta(\phi_\theta^W) = R_\theta(\phi_\theta^*)$ and the RSW mechanism can be solved by Problem P_L and P_H. **Q.E.D.**

Proof for Proposition 4

We first show that, for any RSW mechanism in the original game, there exists an outcome-equivalent RSW mechanism in the simplified game, and vice versa. It is sufficient to show that for any safe mechanism in the original game, there exists an outcome-equivalent safe mechanism in the simplified game, and vice versa. It is similar to Step 1 in the proof for Proposition 2.[19] The direction from the simplified game to the original game is straightforward since any safe mechanism in the simplified game is also safe in the original game. The opposite direction is shown by construction. By definition, a RSW mechanism is $\mu_H - feasible$ for any μ_H. With any μ_H, we can construct an outcome-equivalent $\mu_H - feasible$ mechanism in the simplified game as in Step 1 in the proof for Proposition 2. Since this construction is independent of μ_H, the constructed mechanism is the same and is $\mu_H - feasible$ for any μ_H. Then by definition, the constructed mechanism is safe.

Now we show that we can further restrict no randomization on I. Let $\{\phi_\theta^*\}_{\theta=L,H}$ be a solution to the successive problem P_θ. We have

$$\begin{aligned}
& U(\theta, \phi_\theta^*) = \int u(\theta, \pi, \gamma)\, \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma \text{ (By (6))} \\
= & \int \left[\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\, P(\theta, \omega)\, f_\theta(\omega)\, d\omega - g(\theta, s_1, \pi)\, t(s_1) - I \right] \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma \\
= & \int \left[\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\, P(\theta, \omega)\, f_\theta(\omega)\, d\omega - g(\theta, s_1, \pi)\, t(s_1) \right] \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma \\
& \quad - \int I \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma. \qquad (49)
\end{aligned}$$

Similarly, $R_{\theta'}(\phi_\theta^*) = \int g(\theta', s_1, \pi)\, t(s_1)\, \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma + \int I \phi_\theta^*(\pi, \gamma)\, d\pi d\gamma$. That is, I matters only in expectation for the player's and the designer's expected payoff. Obviously, I does not affect the allocation probability and the player's IC constraint. Thus, if we replace every I in the mixed package with its expectation, it is outcome-equivalent. **Q.E.D.**

Proof for Lemma 1

We first establish a useful lemma which will be used several times throughout the proofs.[20]

[19] We do not need Step 2 since for RSW mechanism we do not need to consider off-equilibrium-path beliefs.

[20] The proof is similar to Lemma 7 of Chen and Zhang (2020). For those who are interested in the

Lemma 5 *Consider a simplified package* (π, γ) *that satisfies the player's incentive constraint under belief* $\xi \in (0,1]$, *i.e.,* $V(\xi, s_2, \pi) \leq \bar{t} \leq V(\xi, s_1, \pi)$.

(a) the statistical experiment is not a monotone binary partition, i.e., $\pi \notin \Pi^M$.

(b) for statistical experiment not equal to $\pi(\underline{\omega})$ *or* $\pi(\overline{\omega})$, *the payment when implementing alternative* 1 *is strictly less than the conditional expected return, i.e.,* $\bar{t} < V(H, s_1, \pi)$.

If at least one of the above two conditions holds, we can construct the following simplified package (π', γ') *with* $\pi' = \pi(y'), \bar{t}' = \frac{\int_{y'}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1-F_H(y')}, I' = -u(H, \pi, \gamma)$, *where* $y' \in (\underline{\omega}, \overline{\omega})$ *is the unique solution to*

$$\int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) T(H,\omega) f_H(\omega) d\omega. \tag{50}$$

Then $r_H(\pi', \gamma') = r_H(\pi, \gamma)$, $u(H, \pi', \gamma') = u(H, \pi, \gamma)$, $r_L(\pi', \gamma') < r_L(\pi, \gamma)$.

Suppose the high-type designer's mixed package in RSW mechanism ϕ_H^W randomizes on packages where at least one of (i) and (ii) does not hold. It is equivalent to the condition that at least one of (a) and (b) hold, and therefore Lemma 5 applies. By setting $\xi = 1$ in Lemma 5, we can construct (π', γ') such that $r_L(\pi', \gamma') < r_L(\pi, \gamma)$, $r_H(\pi', \gamma') = r_H(\pi, \gamma)$, $u(H, \pi', \gamma') = u(H, \pi, \gamma)$. Consider (π'', γ'') with $\pi'' = y' - \epsilon$, $\bar{t}'' = \frac{\int_{y'-\epsilon}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1-F_H(y-\epsilon)}$, $I'' = I'$. Since both $r_L(\pi', \gamma')$ and $r_H(\pi', \gamma')$ are continuous and strictly decreasing in y', there exists a $\epsilon > 0$ such that $r_L(\pi(y'-\epsilon), \gamma'') < r_L(\pi(y'-\epsilon), \gamma'')$, $r_H(\pi(y'-\epsilon), \gamma'') > r_H(\pi, \gamma)$, and $u(H, \pi(y'-\epsilon), \gamma'') = u(H, \pi, \gamma)$. Obviously (π'', γ'') satisfies the player's IC. If we replace (π, γ) with (π'', γ'') in ϕ_H^W, the new mixed package satisfies all the constraints in Problem P_H and yields a strictly higher payoff for the high-type designer, a contradiction to ϕ_H^W being optimal. The new mixed package could have different participation fees across different packages. However, similar to Proposition 3, we can replace all the participation fees with its expectation. **Q.E.D.**

Proof for Lemma 2

We prove by contradiction. Suppose $(\sigma_H(y), I_H)$ is a solution to Problem P_H and $U(H, \phi_H)$ is the player's payoff. Suppose $R_L(\sigma_H, I_H) < R_L^W$. There are two cases. In

proofs, please refer to Appendix B.

case 1, $\sigma_H(\underline{\omega}) = 1$. Therefore, $R_L(\sigma_H, I_H) = I_H + \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega < R_L^W$. Thus

$$\begin{aligned} I_H &< R_L^W - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega = \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega \\ &< \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega \qquad (51) \\ &\Rightarrow U(H,\phi_H) = \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega - I_H > 0 \qquad (52) \end{aligned}$$

Thus, we can increase $I'_H = I_H + \epsilon$ for some $\epsilon > 0$ such that $R_L(\sigma_H, I'_H) \leq R_L^W$ and $R_H(\sigma_H, I'_H) > R_H(\sigma_H, I'_H)$, which is a contradiction. In case 2, $\sigma_H(\underline{\omega}) < 1$. For $\pi(y)$ with $y > \underline{\omega}$ and some $\epsilon > 0$, we can find some $\pi' = \pi(y-\epsilon)$, $\overline{t}' = \frac{\int_{y-\epsilon}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1-F_H(y-\epsilon)}$ such that $r_L\left(\pi', \overline{t}'\right) < r_L\left(\pi, \overline{t}\right) + R_L^W - R_L(\sigma_H, I_H)$. Define $\rho(y) = y - \epsilon$ for $y > \underline{\omega}$ and $\rho(\underline{\omega}) = \underline{\omega}$. Construct: $\sigma'_H(y') = \int_{\rho(y)=y'} \sigma_H(y) dy$ and

$$I'_H = \int_{y=\underline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega \sigma_H(y) dy - U(H,\phi_H) \qquad (53)$$

Therefore, the player's payoff is $U(H,\phi_H)$. The high-type designer's payoff:

$$\begin{aligned} R_H(\sigma'_H, I'_H) &= \int_{\underline{\omega}}^{\overline{\omega}} \int_{y}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H(y) dy - U(H,\phi_H) \\ &= \int_{\underline{\omega}}^{\overline{\omega}} \int_{y}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \int_{\rho(y')=y}^{\overline{\omega}} \sigma_H(y') dy' dy - U(H,\phi_H) \\ &> \int_{\underline{\omega}}^{\overline{\omega}} \int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \int_{\rho(y')=y}^{\overline{\omega}} \sigma_H(y') dy' dy - U(H,\phi_H) \\ &= \int_{\underline{\omega}}^{\overline{\omega}} \int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma_H(y') dy' - U(H,\phi_H) = R_H(\sigma_H, I_H) \qquad (54) \end{aligned}$$

Similarly,

$$\begin{aligned} R_L(\sigma'_H, I'_H) &< R_L(\sigma_H, I_H) + \int_{\underline{\omega}}^{\overline{\omega}} \left[R_L^W - R_L(\sigma_H, I_H)\right] \sigma'_H(y) dy \\ &< R_L(\sigma_H, I_H) + R_L^W - R_L(\sigma_H, I_H) = R_L^W. \qquad (55) \end{aligned}$$

Contradiction. **Q.E.D.**

Proof for Lemma 3

It remains to show that $N(y)$ is strictly decreasing. Take derivative of $N(y)$ with respect to y, we have $N'(y) = -T(H, y)\, f_H(y)\, n(y)$, where

$$\begin{aligned} n(y) &= \left[\frac{f_L(y)}{f_H(y)} - \frac{1-F_L(y)}{1-F_H(y)}\right] \frac{\int_y^{\overline{\omega}} [P(H,\omega) + D(H,y)]\, f_H(\omega) d\omega}{T(H,y)\,[1-F_H(y)]} \\ &\quad + \frac{D(L,y) - D(H,y)}{T(H,y)} \frac{f_L(y)}{f_H(y)} + \frac{1-F_L(y)}{1-F_H(y)}. \end{aligned} \tag{56}$$

Since $\frac{f_L(y)}{f_H(y)} \geq \frac{1-F_L(y)}{1-F_H(y)}$, $P(H,\omega) + D(H,y) \geq T(H,y) \geq 0$ for any $\omega \geq y$, $D(H,y) \leq D(L,y)$, we have $n(y) > 0$. Therefore, $N(y)$ is strictly decreasing in y. **Q.E.D.**

Proof for Proposition 6

It remains to show that there is efficiency loss. First, take derivative of $M(y)$ at $y = \underline{\omega}$:

$$M'(\underline{\omega}) = [f_L(\underline{\omega}) - f_H(\underline{\omega})] \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega - D(H,\underline{\omega}) f_H(\underline{\omega}) + D(L,\underline{\omega}) f_L(\underline{\omega})$$

$$\begin{aligned} &= [f_L(\underline{\omega}) - f_H(\underline{\omega})] \int_{\underline{\omega}}^{\overline{\omega}} [P(H,\omega) + D(H,\underline{\omega})]\, f_H(\omega) d\omega + f_L(\underline{\omega})\,[D(L,\underline{\omega}) - D(H,\underline{\omega})] \\ &> [f_L(\underline{\omega}) - f_H(\underline{\omega})] \int_{\underline{\omega}}^{\overline{\omega}} [P(H,\omega) + D(H,\omega)]\, f_H(\omega) d\omega > 0 \end{aligned} \tag{57}$$

Therefore, $x_H^{\#} < N(\underline{\omega})$. Second,

$$\begin{aligned} N(\underline{\omega}) &= \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega \\ &> \int_{\underline{\omega}}^{\overline{\omega}} P(L,\omega) f_L(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega = R_L^W. \end{aligned} \tag{58}$$

Q.E.D.

Proof for Lemma 4

It is equivalent to show that there exists a unique cutoff $\alpha^* \in [0,1]$ such that $y_H^{\#} \geq y_H^*$ if and only if $\alpha \geq \alpha^*$.

$y_H^{\#}$ is independent of α since $M(y)$ equals

$$(1-\alpha)\left\{\int_y^{\overline{\omega}} \left[\frac{F_2(y) - F_1(y)}{1-F_1(y)} P(1,\omega) + D(1,\omega)\right] f_1(\omega) d\omega - \int_y^{\overline{\omega}} D(2,\omega)\, f_2(\omega) d\omega\right\} \tag{59}$$

Now we will show that y_H^* is strictly decreasing in $\alpha \in [0,1)$. Recall that y_H^* is defined by

$$\begin{aligned}
& \int_{y_H^*}^{\overline{\omega}} \left[\frac{1-F_\alpha\left(y_H^*\right)}{1-F_1\left(y_H^*\right)} P\left(1,\omega\right) + \alpha D\left(1,\omega\right)\right] f_1(\omega) d\omega + (1-\alpha) \int_{y_H^*}^{\overline{\omega}} D\left(2,\omega\right) f_2(\omega) d\omega \\
= \; & \alpha \int_{\underline{\omega}}^{\overline{\omega}} T\left(1,\omega\right) f_1(\omega) d\omega + (1-\alpha) \int_{\underline{\omega}}^{\overline{\omega}} T\left(2,\omega\right) f_2(\omega) d\omega, \qquad (60)
\end{aligned}$$

Applying the Implicit Function Theorem,

$$\begin{aligned}
\frac{dy_H^*}{d\alpha} &= -\frac{1}{N'\left(y_H^*\right)}\Big\{\int_{y_H^*}^{\overline{\omega}} \left[\frac{F_2(y_H^*) - F_1(y_H^*)}{1-F_1(y_H^*)} P(1,\omega) + D(1,\omega)\right] f_1(\omega) d\omega \\
& \quad - \int_{y_H^*}^{\overline{\omega}} D(2,\omega) f_2(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} \left[T(1,\omega) f_1(\omega) - T\left(2,\omega\right) f_2(\omega)\right] d\omega\Big\} \\
&= -\frac{1}{N'\left(y_H^*\right)}\Big\{\int_{y_H^*}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega \\
& \quad + \frac{R_L^W - \alpha \int_{\underline{\omega}}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega}{1-\alpha} - \frac{N\left(y_H^*\right) - \alpha \int_{y_H^*}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega}{1-\alpha}\Big\} \\
&= -\frac{\int_{y_H^*}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} T(1,\omega) f_1(\omega) d\omega}{N'\left(y_H^*\right)(1-\alpha)} \text{ (By } R_L^W = N\left(y_H^*\right)\text{)} \\
&> 0 \text{ (By } N'\left(y_H^*\right) < 0, y_H^* > \underline{\omega}\text{)} \qquad (61)
\end{aligned}$$

Therefore, $\frac{dy_H^*}{d\alpha}$ is strictly negative.

As a result, with y_H^* being constant and $y_H^{\#}$ being strictly decreasing, they can cross with each other at most once. **Q.E.D.**

Appendix B: Proofs for Robustness

Proof for Proposition 7

By Proposition 1, it is sufficient to consider the simplified game. The proposition has two parts. **Part 1**: The outcome of the RSW mechanism can be supported as a PBE outcome in the simplified game. **Part 2**: Under the conditions, the set of outcomes from PBE in the simplified game equals that from RSW mechanism.

Part 1:

There are two cases. In case 1, $x_H^{\#} \geq R_L^W$ and we will construct a belief to support both types proposing the RSW mechanism as a PBE. In case 2, $x_H^{\#} < R_L^W$, and we will apply Theorem 1 in Maskin and Tirole Maskin and Tirole (1992) to show it.

In case 1, $x_H^{\#} \geq R_L^W$. We assign off-equilibrium-path beliefs $\mu_H = 0$ such that the designer's type is believed to be low for sure. It is sufficient to show that, for any $0 - feasible$ direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$, $R_\theta(\phi_\theta) \leq R_\theta^W$. Since the sum of the designer's payoff and the player's payoff cannot exceed the efficient welfare:

$$
\begin{aligned}
& R_L(\phi_L) + U(L,\phi_L) \leq \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega \\
\Rightarrow \quad & R_L(\phi_L) \leq \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega - U(L,\phi_L) \\
\leq \quad & \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega = R_L^W . \qquad (62)
\end{aligned}
$$

Thus, we have $R_L(\phi_L) \leq R_L^W$. If

$$
R_H(\phi_H) < R_L(\phi_L) - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(H,\omega) f_H(\omega) d\omega, \qquad (63)
$$

then

$$
\begin{aligned}
R_H(\phi_H) \quad < \quad & R_L^W - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(H,\omega) f_H(\omega) d\omega \\
= \quad & R_L^W + B(N(\underline{\omega})) < R_L^W + B\left(x_H^{\#}\right) = R_H^W \qquad (64)
\end{aligned}
$$

and the proof is finished. Otherwise, when (63) fails, that is,

$$
R_H(\phi_H) \geq R_L(\phi_L) - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(H,\omega) f_H(\omega) d\omega. \qquad (65)
$$

we need to apply the following lemma:

Lemma 6 *Suppose a $\mu_H - feasible$ mechanism Φ satisfies* (65)*. Then there exists a $\mu_H - feasible$ mechanism Φ' such that:*

(i) *the low-type designer's mixed package only randomizes on packages where the statistical experiment is no disclosure $\pi(\underline{\omega})$;*

(ii) the high-type designer's mixed package only randomizes on packages where the statistical experiment is a monotone binary partition $\pi(y)$ *with* $y \in [\underline{\omega}, \overline{\omega}]$ *and for statistical experiment not equal to* $\pi(\underline{\omega})$ *or* $\pi(\overline{\omega})$*, the payment for selling equals the conditional expected valuation, i.e.,* $\bar{t} = V(H, s_1, \pi)$*;*

(iii) $R_L(\phi'_L) = R_L(\phi_L)$, $R_H(\phi'_H) = R_H(\phi_H)$, $U(L, \phi'_L) \geq U(L, \phi_L)$, $U(H, \phi'_H) = U(H, \phi_H)$.

Proof. The proof is by construction. First, construct the low-type designer's mixed package $\phi'_L\left(\pi\left(\underline{\omega}\right), \gamma^L\right) = 1$, where $t^L\left(s_1\right) = 0, I^L = R_L(\phi_L)$. That is, (i) in the lemma holds. Therefore, the low-type designer's IC constraint holds. Moreover,

$$R_L(\phi'_L) = R_L(\phi_L) \tag{66}$$

$$R_H(\phi'_L) = R_L(\phi_L). \tag{67}$$

and

$$\begin{aligned} U\left(L, \phi'_L\right) &= \int_{\underline{\omega}}^{\overline{\omega}} T\left(L, \omega\right) f_L(\omega) d\omega - R_L(\phi'_L) \\ \geq \ & \left[U\left(L, \phi_L\right) + R_L(\phi_L)\right] - R_L(\phi_L) = U\left(L, \phi_L\right) \end{aligned} \tag{68}$$

Second, we construct the high-type designer's mixed package in the following two steps. In step 1, consider (π, γ) with interim belief $\widehat{\xi} \in (0, 1]$. Denote the function $\rho : \Pi \times \Gamma \to \Pi \times \Gamma$ as follows. There are two cases. In case 1, $\pi = \pi(\underline{\omega})$ *or* $\pi(\overline{\omega})$. Then let $\rho\left(\pi, \gamma\right) = \left(\pi, \gamma\right)$. In case 2, $\pi \neq \pi(\underline{\omega})$ or $\pi(\overline{\omega})$. We apply Lemma 5 to construct (π', γ') where for some y', $\pi' = \pi\left(y'\right)$,

$$t'\left(s_1\right) = \frac{\int_{y'}^{\overline{\omega}} P\left(H, \omega\right) f_H\left(\omega\right) d\omega}{1 - F_H\left(y'\right)}, \ I' = -u\left(H, \pi, \gamma\right), \tag{69}$$

such that

$$r_L(\pi', \gamma') \leq r_L(\pi, \gamma), \tag{70}$$

$$r_H(\pi', \gamma') = r_H(\pi, \gamma), \tag{71}$$

$$u\left(H, \pi', \gamma'\right) = u\left(H, \pi, \gamma\right). \tag{72}$$

Let $\rho(\pi,\gamma) = (\pi',\gamma')$. In step 2, construct

$$\phi'_H(\pi',\gamma') = \int_{\rho(\pi,\gamma)=(\pi',\gamma')} \phi_H(\pi,\gamma)d\pi d\gamma. \tag{73}$$

Therefore, ϕ'_H satisfies (ii) in the lemma. Moreover,

$$\begin{aligned}
R_H(\phi'_H) &= \int r_H(\pi',\gamma')\,\phi'_H(\pi',\gamma')d\pi' d\gamma' \\
&= \int r_H(\pi',\gamma') \int_{\rho(\pi,\gamma)=(\pi',\gamma')} \phi_H(\pi,\gamma)d\pi d\gamma d\pi' d\gamma' = \int r_H(\rho(\pi,\gamma))\,\phi_H(\pi,\gamma)d\pi d\gamma \\
&= \int_{\pi=\pi(\underline{\omega})\ \ or\ \ \pi(\overline{\omega})} r_H(\pi,\gamma)\,\phi_H(\pi,\gamma)d\pi d\gamma + \int_{\pi\neq\pi(\underline{\omega})\ \ or\ \ \pi(\overline{\omega})} r_H(\rho(\pi,\gamma))\,\phi_H(\pi,\gamma)d\pi d\gamma \\
&= \int_{\pi=\pi(\underline{\omega})\ \ or\ \ \pi(\overline{\omega})} r_H(\pi,\gamma)\,\phi_H(\pi,\gamma)d\pi d\gamma + \int_{\pi\neq\pi(\underline{\omega})\ \ or\ \ \pi(\overline{\omega})} r_H(\pi,\gamma)\,\phi_H(\pi,\gamma)d\pi d\gamma \\
&= R_H(\phi_H).
\end{aligned} \tag{74}$$

where second last equality follows (71). Similarly,

$$R_L(\phi'_H) \leq R_L(\phi_H), \tag{75}$$

$$U(H,\phi'_H) = U(H,\phi_H). \tag{76}$$

(68) and (76) show that the player's participation constraint holds. By (67) and (74),

$$R_H(\phi'_H) = R_H(\phi_H) \geq R_L(\phi_L) = R_H(\phi'_L), \tag{77}$$

and therefore the high-type designer's IC constraint holds. (66) and (75) show that the low-type designer's IC constraint holds. ■

This lemma shows that there exists a $0-$ *feasible* mechanism $\Phi' = \{\phi'_\theta\}_{\theta=L,H}$ and a probability measure on $x \in [N(\overline{\omega}), N(\underline{\omega})]$, $\kappa'_H(x)$ such that $R_L(\phi'_L) = R_L(\phi_L) \leq R_L^W$,

$$\begin{aligned}
R_H(\phi_H) = R_H(\phi'_H) &= R_L(\phi'_H) + \int_{N(\overline{\omega})}^{N(\underline{\omega})} B(x)\,\kappa'_H(x)dx \\
&\leq R_L(\phi'_H) + \int_{N(\overline{\omega})}^{N(\underline{\omega})} B\left(x_H^{\#}\right)\kappa'_H(x)dx \quad \text{(By definition of } x_H^{\#}\text{)} \\
&= R_L(\phi'_H) + B\left(x_H^{\#}\right) \\
&\leq R_L(\phi'_L) + B\left(x_H^{\#}\right) \quad \text{(By the low-type designer's IC constraint)} \\
&\leq R_L^W + B\left(x_H^{\#}\right) = R_H^W,
\end{aligned} \tag{78}$$

We are done.

In case 2, $x_H^{\#} < R_L^W$. Suppose $\Phi^W = \{\phi_\theta\}_{\theta=L,H}$ is the RSW mechanism. To apply Theorem 1 of Maskin and Tirole Maskin and Tirole (1992), it is sufficient to show that Φ^W is interim efficient relative to some belief $\mu_H \in (0,1)$ such that the designer's type is high with probability of μ_H. That is, there exists no other $\mu_H - feasible$ grand mechanism Φ that satisfies (1) $R_\theta(\phi_\theta) \geq R_\theta^W$ with inequality strictly satisfied for at least one type, and (2) $\mu_H U(H,\phi_H) + (1-\mu_H) U(L,\phi_L) \geq \mu_H U\left(H,\phi_H^W\right) + (1-\mu_H) U\left(L,\phi_L^W\right)$. We will show that such $\mu_H > 0$ exists if

$$\mu_H + \widehat{B}'(x)\,|_{x=R_L^W} < 0. \tag{79}$$

Since $x_H^{\#} < R_L^W$, $\widehat{B}'(x)\,|_{x=R_L^W} < 0$ and therefore (79) is well specified for $\mu_H > 0$. Since $x_H^{\#} < R_L^W$, $U\left(\theta,\phi_\theta^W\right) = 0$, and therefore condition (2) is redundant for any $\mu_H - feasible$ mechanism Φ. It is sufficient to show that, for any $\mu_H > 0$ that satisfies (79) and any $\mu_H - feasible$ mechanism Φ, $R_\theta(\phi_\theta) \geq R_\theta^W$ implies that Φ is the RSW mechanism. There are three steps.

Step 1: We prove by contradiction that (65) holds. Suppose in contrary, (63) holds. Since the sum of the designer's payoff cannot exceed the efficient welfare, i.e.,

$$\begin{aligned} & \mu_H R_H(\phi_H) + (1-\mu_H) R_L(\phi_L) \\ \leq\ & \mu_H \int_{\underline{\omega}}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega + (1-\mu_H) \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega. \end{aligned} \tag{80}$$

Therefore,

$$\begin{aligned} & R_H(\phi_H) \\ \leq\ & \mu_H \int_{\underline{\omega}}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega + (1-\mu_H) \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega \\ & -(1-\mu_H)\left[\int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} D(H,\omega) f_H(\omega) d\omega\right] \\ =\ & \mu_H\left[N(\underline{\omega})\right] + \widehat{B}(N(\underline{\omega})) + (1-\mu_H) R_L^W && (81) \\ <\ & \mu_H R_L^W + \widehat{B}\left(R_L^W\right) + (1-\mu_H) R_L^W && (82) \\ =\ & R_L^W + \widehat{B}\left(R_L^W\right) = R_H^W, && (83) \end{aligned}$$

which contradicts that $R_H(\phi_H) \geq R_H^W$. (81) follows

$$\widehat{B}\left(N\left(\underline{\omega}\right)\right) = B\left(N\left(\underline{\omega}\right)\right) = \int_{\underline{\omega}}^{\overline{\omega}} D\left(H,\omega\right) f_H(\omega) d\omega - \int_{\underline{\omega}}^{\overline{\omega}} D\left(L,\omega\right) f_L(\omega) d\omega, \tag{84}$$

and $R_L^W = \int_{\underline{\omega}}^{\overline{\omega}} T\left(L,\omega\right) f_L(\omega) d\omega$. (82) follows the following claim:

Claim 1 *If $x_H^{\#} < R_L^W$ and (79) holds, $\mu_H x + \widehat{B}\left(x\right)$ is strictly decreasing in $x \geq R_L^W$.*

Proof. (79) implies that $\frac{\partial\left[\mu_H x + \widehat{B}(x)\right]}{\partial x}|_{x=R_L^W} < 0$. By definition of $\widehat{B}\left(\cdot\right)$, $\frac{\partial^2\left[\mu_H x + \widehat{B}(x)\right]}{\partial x^2} = \frac{\partial^2 \widehat{B}(x)}{\partial x^2} \leq 0$. Thus, for any $x > R_L^W$,

$$\frac{\partial\left[\mu_H x + \widehat{B}\left(x\right)\right]}{\partial x} \leq \frac{\partial\left[\mu_H x + \widehat{B}\left(x\right)\right]}{\partial x}|_{x=R_L^W} < 0. \tag{85}$$

■

Step 2: We will show by contradiction that $U\left(H,\phi_H\right) = 0$. First, since the sum of the designer's payoff and the player's payoff cannot exceed the efficient welfare,

$$U\left(L,\phi_L\right) + R_L(\phi_L) \leq \int_{\underline{\omega}}^{\overline{\omega}} T\left(L,\omega\right) f_L(\omega) d\omega = R_L^W \Rightarrow U\left(L,\phi_L\right) \leq R_L^W - R_L(\phi_L). \tag{86}$$

By the player's participation constraint,

$$\begin{aligned} & \mu_H U\left(H,\phi_H\right) + \left(1-\mu_H\right) U\left(L,\phi_L\right) \geq 0 \\ \Rightarrow \;& U\left(H,\phi_H\right) \geq \frac{-\left(1-\mu_H\right) U\left(L,\phi_L\right)}{\mu_H} \geq \frac{\left(1-\mu_H\right)\left[R_L(\phi_L) - R_L^W\right]}{\mu_H} \\ \geq \;& \frac{\left(1-\mu_H\right)\left[R_L(\phi_L) - R_L^W\right]}{\mu_H} \geq 0. \end{aligned} \tag{87}$$

Second, suppose in contrary, $U\left(H,\phi_H\right) > 0$. Given that $R_L(\phi_L) \geq R_L^W$, and (65), we will show that the high-type designer's equilibrium payoff is lower than R_H^W, which contradicts that $R_H(\phi_H) \geq R_H^W$. With (65), we can apply Lemma 6 to construct a $\mu_H -$ *feasible* mechanism $\Phi' = \{\phi'_\theta\}_{\theta=L,H}$ such that $R_L(\phi'_L) = R_L(\phi_L) \geq R_L^W$, $R_H(\phi'_H) = R_H(\phi_H)$, and there exists a probability measure on cutoffs $y \in [0,1]$, $\sigma'_H\left(y\right)$, with

$$R_H\left(\phi'_H\right) + U\left(H,\phi'_H\right) = \int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H\left(y\right) dy, \tag{88}$$

$$R_L\left(\phi'_H\right) + U\left(H,\phi'_H\right) = \int_{\underline{\omega}}^{\overline{\omega}} N\left(y\right) \sigma'_H\left(y\right) dy. \tag{89}$$

We will show that $R_H(\phi'_H) < R_H^W$. There are two cases. In case 1, $\int_0^1 N(y)\,\sigma'_H(y)\,dy \leq R_L^W$. Therefore, $\sigma'_H(y)$ and $I'_H = 0$ is a candidate solution to Problem P_H. By the maximization of R_H^W,

$$\begin{aligned}
&\int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H(y)\,dy \leq R_H^W \\
\Rightarrow\quad & R_H(\phi'_H) = \int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H(y)\,dy - U(H,\phi'_H) \\
< \quad & \int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H(y)\,dy \text{ (By } U(H,\phi'_H) > 0) \\
\leq \quad & R_H^W .
\end{aligned} \tag{90}$$

In case 2, $\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy > R_L^W$. (88) implies that

$$R_H(\phi'_H) = \int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma'_H(y)\,dy - U(H,\phi'_H)$$

$$\begin{aligned}
= \quad & \int_{\underline{\omega}}^{\overline{\omega}} \left[N(y) + B(N(y))\right] \sigma'_H(y)\,dy - U(H,\phi'_H) \text{ (By definition of } N(y) \text{ and } B(x)) \\
= \quad & \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy + \int_{\underline{\omega}}^{\overline{\omega}} B(N(y))\,\sigma'_H(y)\,dy - U(H,\phi'_H) \\
\leq \quad & \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy + \int_{\underline{\omega}}^{\overline{\omega}} \widehat{B}(N(y))\,\sigma'_H(y)\,dy - U(H,\phi'_H) \;\; (B(x) \leq \widehat{B}(x)) \\
\leq \quad & \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy\right) - U(H,\phi'_H) .
\end{aligned} \tag{91}$$

The last inequality follows that $\widehat{B}(x)$ is concave by definition. By (87),

$$\begin{aligned}
& U(H,\phi'_H) \geq \frac{(1-\mu_H)\left[R_L(\phi'_L) - R_L^W\right]}{\mu_H} \geq \frac{(1-\mu_H)\left[R_L(\phi'_H) - R_L^W\right]}{\mu_H} \\
= \quad & \frac{(1-\mu_H)\left[\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy - U(H,\phi'_H) - R_L^W\right]}{\mu_H} \text{ (By (89))} \\
\Leftrightarrow \quad & U(H,\phi'_H) \geq (1-\mu_H)\left[\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy - R_L^W\right].
\end{aligned} \tag{92}$$

Combine (91) and (92),

$$\begin{aligned}
& R_H(\phi'_H) \leq \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy\right) \\
& -(1-\mu_H)\left[\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma'_H(y)\,dy - R_L^W\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mu_H \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy\right) + (1-\mu_H)\,R_L^W \\
&< \mu_H R_L^W + \widehat{B}\left(R_L^W\right) + (1-\mu_H)\,R_L^W && (93) \\
&= R_L^W + \widehat{B}\left(R_L^W\right) = R_H^W. && (94)
\end{aligned}
$$

(93) follows Claim 1 and that $\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy > R_L^W$. To conclude, in both cases, $R_H(\phi_H') < R_H^W$, which contradicts $R_H(\phi_H') \geq R_H^W$.

Step 3: The following Lemma 7 shows that Φ is the RSW mechanism since $R_\theta(\phi_\theta) \geq R_\theta^W$ and $U(H, \phi_H) = 0$. Therefore, Φ^W is interim efficient relative to some belief $\mu_H \in (0,1)$. Then by the Theorem 1 of Maskin and Tirole (1992), the outcome of the RSW mechanism can be supported as a PBE outcome.

Lemma 7 *Suppose $\Phi = \{\phi_\theta\}_{\theta=L,H}$ is a $\mu_H - feasible$ mechanism for some $\mu_H > 0$. If $R_\theta(\phi_\theta) \geq R_\theta^W$, and $U(H, \phi_H) = 0$, then Φ is the RSW mechanism.*

Proof. Since $R_\theta(\phi_\theta) \geq R_\theta^W$, it is sufficient to show that Φ is a safe mechanism. By the player's participation constraint, $U(H, \phi_H) = 0$ implies that $U(L, \phi_L) \geq 0$. Therefore,

$$
U(L, \phi_L) + R_L(\phi_L) \geq R_L^W, \tag{95}
$$

$$
\begin{aligned}
U(L, \phi_L) + R_L(\phi_L) &= \int \left[\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\,T(L,\omega) f_L(\omega) d\omega\right] \phi_L(\pi,\gamma) d\pi d\gamma \\
&\leq \int \left[\int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega\right] \phi_L(\pi,\gamma) d\pi d\gamma \text{ (By } \pi(s_1|\omega) \leq 1, \forall\omega) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega) f_L(\omega) d\omega \text{ (By} \int \phi_L(\pi,\gamma) d\pi d\gamma = 1) && (96)
\end{aligned}
$$

with equality holds only if $\int \phi_L(\pi(\underline{\omega}), \gamma) d\gamma = 1$. Combined with (95), $\int \phi_L(\pi(\underline{\omega}), \gamma) d\gamma = 1$. Thus for any (π, γ) that ϕ_L assigns positive weight, only signal s_1 is sent, and therefore, the player's IC constraint holds. Therefore Φ is a safe mechanism. ■

Part 2:

Consider any $\mu_H^0 - feasible$ direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$ that implements a PBE outcome. First, the following Lemma 8 shows that the equilibrium payoff $R_\theta(\phi_\theta) \geq R_\theta^W$.

Lemma 8 *In any PBE, the type-θ designer's payoff is at least R_θ^W.*

Proof. Since an RSW mechanism Φ^W is a safe mechanism, type-θ designer's payoff is R_θ^W regardless of the player's belief. Suppose on-equilibrium-path, type-θ designer proposes Φ^W. Then type-θ designer's equilibrium payoff is R_θ^W. And type-θ' designer's payoff is at least $R_{\theta'}^W$. If Φ^W is off-equilibrium-path, type-θ designer's payoff is at least R_θ^W. ■

With this property, by the same arguments in case 2 of Part 1, it follows that the outcome of the RSW mechanism is the unique PBE outcome. **Q.E.D.**

Proof for Proposition 8

There are two parts. Part 1: the outcome of the RSW mechanism survives the intuitive criterion. Part 2: when $x_H^\# \leq R_L^W$, the set of intuitive outcome equals that of the RSW mechanism.[21]

Part 1 :

It is sufficient to show that there exists an off-equilibrium-path belief that satisfies the intuitive criterion and supports both types of the designer proposing the RSW mechanism as a PBE. By revelation principle, given belief μ_H such that the designer's type is high with probability μ_H, a grand mechanism Ψ is outcome-equivalent to a $\mu_H - feasible$ direct grand mechanism $\Phi(\Psi, \mu_H) = \{\phi_\theta(\Psi, \mu_H)\}_{\theta=L,H}$. There are three cases.

In case 1, $\max_{\mu_H} R_L(\Phi(\Psi, \mu_H)) < R_L^W$, and $\max_{\mu_H} R_H(\Phi(\Psi, \mu_H)) \geq R_H^W$. Then by the intuitive criterion, construct $\mu_H(\Psi) = 1$. We show $R_H(\Phi(\Psi, 1)) \leq R_H^W$ by contradiction. Suppose $R_H(\Phi(\Psi, 1)) > R_H^W$. By the player's participation constraint, $U(H, \phi_H(\Psi, 1)) \geq 0$. Since

$$\begin{aligned} R_L(\phi_L(\Psi, 1)) &< R_L^W = R_H^W - B\left(R_L^W\right) < R_H^W - B\left(N\left(\underline{\omega}\right)\right) \\ &\leq R_H(\Phi(\Psi, 1)) - B\left(N\left(\underline{\omega}\right)\right), \end{aligned} \tag{97}$$

[21] The proposition can be established by following either Maskin and Tirole (1992) or Nishimura (2022). Here we follow the former by extending the action space to be infinite dimensional.

we can apply Lemma 6 to construct a $1-feasible$ mechanism, Φ', such that $R_\theta(\phi'_\theta) = R_\theta\left(\phi_\theta\left(\Psi,\mu^{**}\right)\right), U(H,\phi'_H) = U\left(H,\phi_H\left(\Psi,1\right)\right) \geq 0$, and (i), (ii) of the Lemma 6 holds. (i), (ii) of the Lemma 6 implies that full-information IC constraints of the player hold. (i) of the Lemma 6 also implies that

$$\begin{aligned}
& U\left(L,\phi'_L\right) + R_L\left(\phi'_L\right) = \int\left[\int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega)f_L(\omega)d\omega\right]\phi'_L(\pi,\gamma)d\pi d\gamma \\
= & \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega)f_L(\omega)d\omega = R_L^W \\
\Rightarrow & U(L,\phi'_L) = R_L^W - R_L\left(\phi'_L\right) = R_L^W - R_L\left(\phi_L\left(\Psi,1\right)\right) > 0 \qquad (98)
\end{aligned}$$

Therefore, Φ' is a safe mechanism. By maximization, $R_\theta(\phi'_\theta) \leq R_\theta^W$. Therefore, $R_\theta\left(\phi_\theta\left(\Psi,1\right)\right) \leq R_\theta^W$, a contradiction. To conclude, both types of designer have no incentive to deviate.

In case 2, $\max_\mu R_L\left(\Phi\left(\Psi,\mu\right)\right) \geq R_L^W$, and $\max_\mu R_H\left(\Phi\left(\Psi,\mu\right)\right) < R_H^W$. Then by the intuitive criterion, construct $\mu_H\left(\Psi\right) = 0$. It remains to show that $R_L\left(\phi_L\left(\Psi,0\right)\right) \leq R_L^W$. By the player's participation constraint, $U(L,\phi_L\left(\Psi,0\right)) \geq 0$. Since the sum of the designer's payoff and the player's payoff cannot exceed the efficient welfare,

$$\begin{aligned}
& U\left(L,\phi_L\left(\Psi,0\right)\right) + R_L\left(\phi_L\left(\Psi,0\right)\right) \leq \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega)f_L(\omega)d\omega \\
\Rightarrow & R_L\left(\phi_L\left(\Psi,0\right)\right) \leq \int_{\underline{\omega}}^{\overline{\omega}} T(L,\omega)f_L(\omega)d\omega - U(L,\phi_L\left(\Psi,0\right)) \\
= & R_L^W - U(L,\phi_L\left(\Psi,0\right)) < R_L^W. \qquad (99)
\end{aligned}$$

Both types of designer have no incentive to deviate.

In case 3, the intuitive criterion puts no restriction on the belief. The claim is verified with Proposition 6.

Part 2 :

Consider any $\mu_H^0 - feasible$ direct grand mechanism $\Phi = \{\phi_\theta\}_{\theta=L,H}$ that implements an intuitive outcome. It is sufficient to show that Φ is the RSW mechanism.

First, Lemma 8 shows that $R_\theta(\phi_\theta) \geq R_\theta^W$. Moreover, since the sum of the designer's payoff and the player's payoff cannot exceed the efficient welfare,

$$U\left(L,\phi_L\right)+R_L(\phi_L)\leq\int_{\underline{\omega}}^{\overline{\omega}}T\left(L,\omega\right)f_L(\omega)d\omega=R_L^W\Rightarrow U\left(L,\phi_L\right)\leq R_L^W-R_L(\phi_L). \tag{100}$$

By the player's participation constraint,

$$\begin{aligned}
&\mu_H^0U\left(H,\phi_H\right)+\mu_L^0U\left(L,\phi_L\right)\geq 0\\
\Rightarrow\quad &U\left(H,\phi_H\right)\geq\frac{-\mu_L^0U\left(L,\phi_L\right)}{\mu_H^0}\geq\frac{\mu_L^0\left[R_L(\phi_L)-R_L^W\right]}{\mu_H^0}\geq 0.
\end{aligned} \tag{101}$$

Second, we will show that if $U\left(H,\phi_H\right)>0$, there exists a $\Phi'=\{\phi'\}_{\theta=L,H}$ such that $R_H\left(\phi'_H\right)>R_H(\phi_H)$, $R_L(\phi'_L)<R_L(\phi_L)$ and $U\left(H,\phi'_H\right)\geq 0$, i.e., only the high-type designer has incentive to deviate.

Suppose the following fails:

$$R_H\left(\phi_H\right)\geq R_L(\phi_L)-\int_{\underline{\omega}}^{\overline{\omega}}D\left(L,\omega\right)f_L(\omega)d\omega+\int_{\underline{\omega}}^{\overline{\omega}}D\left(H,\omega\right)f_H(\omega)d\omega. \tag{102}$$

Since the sum of the designer's payoff cannot exceed the efficient welfare, similar to (83), $R_H(\phi_H)<R_H^W$, which contradicts that $R_H(\phi_H)\geq R_H^W$. Therefore, (102) holds and we can apply Lemma 6 to construct a $\mu_H^0-feasible$ mechanism $\Phi^1=\{\phi_\theta^1\}_{\theta=L,H}$ such that $R_L(\phi_L^1)=R_L(\phi_L)$, $R_H(\phi_H^1)=R_H(\phi_H)\geq R_H^W$, $U\left(H,\phi_H^1\right)=U\left(H,\phi_H\right)>0$ and there exists a probability measure on cutoffs $y\in[0,1]$, $\sigma_H^1\left(y\right)$ with

$$R_H\left(\phi_H^1\right)+U\left(H,\phi_H^1\right)=\int_{\underline{\omega}}^{\overline{\omega}}\int_y^{\overline{\omega}}T(H,\omega)f_H(\omega)d\omega\sigma_H^1\left(y\right)dy, \tag{103}$$

$$R_L\left(\phi_H^1\right)+U\left(H,\phi_H^1\right)=\int_{\underline{\omega}}^{\overline{\omega}}N\left(y\right)\sigma_H^1\left(y\right)dy. \tag{104}$$

There exists $\sigma'_H\left(y\right)$ such that

$$\int_{\underline{\omega}}^{\overline{\omega}}N\left(y\right)\sigma'_H\left(y\right)dy=R_L\left(\phi_H^1\right), \tag{105}$$

$$\widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}}N\left(y\right)\sigma'_H\left(y\right)dy\right)=\int_{\underline{\omega}}^{\overline{\omega}}B\left(N\left(y\right)\right)\sigma'_H\left(y\right)dy. \tag{106}$$

First, we prove by contradiction that

$$R_L\left(\phi_H^1\right)\geq R_L^W. \tag{107}$$

Suppose in contrary, $R_L\left(\phi_H^1\right) < R_L^W$. Then similar to the proof of the Lemma 2, by modifying ϕ_H^1 slightly, we can construct a σ_H'' with

$$\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H''(y)\,dy - U\left(H,\phi_H^1\right) < R_L^W, \tag{108}$$

$$\int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma_H''(y)\,dy - U\left(H,\phi_H^1\right) > R_H\left(\phi_H^1\right). \tag{109}$$

(108) implies that σ_H'' and $I_H'' = \int_{y=\underline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega \sigma_H''(y)\,dy - U\left(H,\phi_H^1\right)$ is a candidate solution to Problem P_H and therefore

$$\int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma_H''(y)\,dy - U\left(H,\phi_H^1\right) \leq R_H^W \leq R_H\left(\phi_H^1\right),$$

which contradicts (109). Second, since $U\left(H,\phi_H^1\right) > 0$,

$$\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy = R_L\left(\phi_H^1\right) < \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H^1(y)\,dy. \tag{110}$$

Thus, by arguments similar to (91) in Proposition 6,

$$\begin{aligned}
R_H\left(\phi_H^1\right) &\leq \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H^1(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H^1(y)\,dy\right) - U\left(H,\phi_H^1\right) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H^1(y)\,dy\right) \\
&< \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy\right) \\
&= \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma_H'(y)\,dy + \int_{\underline{\omega}}^{\overline{\omega}} B(N(y))\,\sigma_H'(y)\,dy \\
&= \int_{\underline{\omega}}^{\overline{\omega}} \int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma_H'(y)\,dy. \qquad (111)
\end{aligned}$$

The last inequality follows (110), (107) and the following claim:

Claim 2 $\widehat{B}(x)$ *is strictly decreasing in* $x \geq x_H^{\#}$.

Proof. Suppose in contrary, there exists $x' > x'' \geq x_H^{\#}$ such that $\widehat{B}(x') \geq \widehat{B}(x'')$. First, since $B(x) < B\left(x_H^{\#}\right)$ for any $x > x_H^{\#}$, $\widehat{B}(x'') < B\left(x_H^{\#}\right)$. Second, there exists a η with $\eta x' + (1-\eta)\,x_H^{\#} = x''$ such that

$$\begin{aligned}
\eta\widehat{B}(x') + (1-\eta)\,B\left(x_H^{\#}\right) &> \eta\widehat{B}(x'') + (1-\eta)\,\widehat{B}(x'') \text{ (By } \widehat{B}(x'') < B\left(x_H^{\#}\right)\text{)} \\
&= \widehat{B}(x'') \qquad (112)
\end{aligned}$$

There contradicts the definition of $\widehat{B}(x'')$. $\blacksquare$

Since $N(y)$ is continuous in y, similar to the proof of the Lemma 2, by modifying σ'_H a bit, there exists a σ^2_H such that

$$\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy < R_L\left(\phi^1_H\right), \tag{113}$$

$$\int_{\underline{\omega}}^{\overline{\omega}} \int_{y}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma^2_H(y)\,dy > R_H\left(\phi^1_H\right), \tag{114}$$

$$\widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy\right) = \int_{\underline{\omega}}^{\overline{\omega}} B(N(y))\,\sigma^2_H(y)\,dy \tag{115}$$

Construct degenerate probability $\phi'_L(\pi(\underline{\omega}), \gamma^3) = 1$, where

$$I^3 = \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega, \tag{116}$$

$$\phi'_H(\pi(y), \left(\frac{\int_y^{\overline{\omega}} P(H,\omega)\, f_H(\omega)\, d\omega}{1 - F_H(y)}, 0\right)) = \sigma^2_H(y). \tag{117}$$

Therefore,

$$\begin{aligned} R_L(\phi'_L) &= R_L(\phi'_H) = \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy, & (118)\\ R_H(\phi'_H) &= \int_{\underline{\omega}}^{\overline{\omega}} \int_{y}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega \sigma^2_H(y)\,dy. & (119) \end{aligned}$$

Since

$$\begin{aligned} R_H(\phi'_L) &= \int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy - \int_{\underline{\omega}}^{\overline{\omega}} D(L,\omega) f_L(\omega) d\omega + \int_{\underline{\omega}}^{\overline{\omega}} D(H,\omega) f_H(\omega) d\omega \\ &= R_L(\phi'_H) + B(N(\underline{\omega})) = R_L(\phi'_H) + \widehat{B}(N(\underline{\omega})) \\ &\leq R_L(\phi'_H) + \widehat{B}\left(\int_{\underline{\omega}}^{\overline{\omega}} N(y)\,\sigma^2_H(y)\,dy\right) \text{ (By Claim A2)} \\ &= R_L(\phi'_H) + \int_{\underline{\omega}}^{\overline{\omega}} B(N(y))\,\sigma^2_H(y)\,dy = R_H(\phi'_H), \end{aligned} \tag{120}$$

the high-type designer's IC constraint holds. Since $R_L(\phi'_L) < R_L\left(\phi^1_H\right) \leq R_L\left(\phi^1_L\right) = R_L(\phi_L)$, by the intuitive criterion, $\mu_H(\Phi') = 1$. Since $U(H, \phi'_H) = 0$, the player participates. Since $R_H(\phi'_H) > R_H(\phi^1_H) = R_H(\phi_H)$, the high-type designer deviates.

To conclude, in both cases, an equilibrium survives the intuitive criterion only if $U(H, \phi_H) = 0$.

Third, by Lemma 7, $R_\theta(\phi_\theta) \geq R_\theta^W$ and $U(H, \phi_H) = 0$ imply that Φ is the RSW mechanism. **Q.E.D.**

Appendix C: Proof for Lemma 5

Both conditions (a) and (b) imply $\pi \neq \pi(\underline{\omega})$ or $\pi(\overline{\omega})$. Given that we do not distinguish statistical experiments with zero measure difference, we have

$$0 < \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) T(H,\omega) f_H(\omega) d\omega < \int_{\underline{\omega}}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega. \tag{121}$$

Since $\int_y^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega$ is strictly decreasing in y, there exists a unique $y' \in (\underline{\omega}, \overline{\omega})$ with

$$\int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) T(H,\omega) f_H(\omega) d\omega. \tag{122}$$

Thus, we can construct the simplified package as in the Lemma. Obviously, (π', γ') satisfies the player's IC constraint under belief $\xi' = 1$ since $\bar{t}' = V(H, s_1, \pi)$. We also have

$$u(H, \pi', \gamma') = [1 - F_H(y')] [V(H, s_1, \pi') - \bar{t}'] - I' = u(H, \pi, \gamma), \tag{123}$$

and

$$\begin{aligned} r_H(\pi', \gamma') &= \int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega) d\omega - u(H, \pi, \gamma) \\ &= \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) T(H,\omega) f_H(\omega) d\omega - u(H, \pi, \gamma) = r_H(\pi, \gamma). \end{aligned} \tag{124}$$

It thus remains to show that $r_L(\pi', \gamma') < r_L(\pi, \gamma)$. When condition (a) is satisfied, we will use the following lemma:

Lemma 9 *Suppose $h(\omega) > 0$ for $\omega \in (\underline{\omega}, \overline{\omega}]$ and π is a binary partition that is not monotone. $0 < \int_y^{\overline{\omega}} h(\omega) d\omega \leq \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) h(\omega) d\omega$ implies that for an increasing function ϑ,*

$$\frac{\int_y^{\overline{\omega}} \vartheta(\omega) h(\omega) d\omega}{\int_y^{\overline{\omega}} h(\omega) d\omega} \geq \frac{\int_{\underline{\omega}}^{\overline{\omega}} \vartheta(\omega) \pi(s_1|\omega) h(\omega) d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) h(\omega) d\omega} \tag{125}$$

with inequality strictly holds if ϑ is strictly increasing.

Proof. First, since $\pi(\cdot|\omega) \neq \pi(\underline{\omega})$, there exists $y' \leq y$ such that

$$\begin{aligned} & \int_{y'}^{\overline{\omega}} h(\omega)\, d\omega = \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\, h(\omega)\, d\omega \\ \Leftrightarrow \quad & \int_{\underline{\omega}}^{y'} \pi(s_1|\omega)\, h(\omega)\, d\omega = \int_{y'}^{\overline{\omega}} [1-\pi(s_1|\omega)]\, h(\omega)\, d\omega \end{aligned} \tag{126}$$

Since π is not monotone, $\int_{\underline{\omega}}^{y'} \pi(s_1|\omega)\, h(\omega)\, d\omega > 0$ and $\int_{y'}^{\overline{\omega}} [1-\pi(s_1|\omega)]\, h(\omega)\, d\omega > 0$. Since ϑ is increasing,

$$\begin{aligned} & \int_{y'}^{\overline{\omega}} \vartheta(\omega)\, h(\omega)\, d\omega \\ = \quad & \int_{y'}^{\overline{\omega}} \pi(s_1|\omega)\, \vartheta(\omega)\, h(\omega)\, d\omega + \int_{y'}^{\overline{\omega}} [1-\pi(s_1|\omega)]\, \vartheta(\omega)\, h(\omega)\, d\omega \\ \geq \quad & \int_{y'}^{\overline{\omega}} \pi(s_1|\omega)\, \vartheta(\omega)\, h(\omega)\, d\omega + \vartheta(y') \int_{y'}^{\overline{\omega}} [1-\pi(s_1|\omega)]\, h(\omega)\, d\omega & (127) \\ = \quad & \int_{y'}^{\overline{\omega}} \pi(s_1|\omega)\, \vartheta(\omega)\, h(\omega)\, d\omega + \vartheta(y') \int_{\underline{\omega}}^{y'} \pi(s_1|\omega)\, h(\omega)\, d\omega \text{ (By (126))} \\ \geq \quad & \int_{y'}^{\overline{\omega}} \pi(s_1|\omega)\, \vartheta(\omega)\, h(\omega)\, d\omega + \int_{\underline{\omega}}^{y'} \pi(s_1|\omega)\, \vartheta(\omega)\, h(\omega)\, d\omega & (128) \\ = \quad & \int_{\underline{\omega}}^{\overline{\omega}} \vartheta(\omega)\, \pi(s_1|\omega)\, h(\omega)\, d\omega \\ \Leftrightarrow \quad & \frac{\int_{y'}^{\overline{\omega}} \vartheta(\omega)\, h(\omega)\, d\omega}{\int_{y'}^{\overline{\omega}} h(\omega)\, d\omega} \geq \frac{\int_{\underline{\omega}}^{\overline{\omega}} \vartheta(\omega)\, \pi(s_1|\omega)\, h(\omega)\, d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\, h(\omega)\, d\omega} & (129) \end{aligned}$$

If ϑ is strictly increasing, inequalities in (127) and (128) are strict and therefore the inequality in (129) is strict.

Second, take derivative of $\frac{\int_y^{\overline{\omega}} \vartheta(\omega) h(\omega) d\omega}{\int_y^{\overline{\omega}} h(\omega) d\omega}$ with respect to $y > \underline{\omega}$,

$$\frac{h(y) \int_y^{\overline{\omega}} [\vartheta(\omega) - \vartheta(y)]\, h(\omega)\, d\omega}{\left[\int_y^{\overline{\omega}} h(\omega)\, d\omega\right]^2} \geq 0. \tag{130}$$

Therefore,

$$\frac{\int_y^{\overline{\omega}} \vartheta(\omega)\, h(\omega)\, d\omega}{\int_y^{\overline{\omega}} h(\omega)\, d\omega} \geq \frac{\int_{y'}^{\overline{\omega}} \vartheta(\omega)\, h(\omega)\, d\omega}{\int_{y'}^{\overline{\omega}} h(\omega)\, d\omega} \geq \frac{\int_{\underline{\omega}}^{\overline{\omega}} \vartheta(\omega)\, \pi(s_1|\omega)\, h(\omega)\, d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega)\, h(\omega)\, d\omega}, \tag{131}$$

with inequality strictly holds if ϑ is strictly increasing. ■

First, $T(H,\omega)f_H(\omega) > 0$ for $\omega \in (\underline{\omega}, \overline{\omega}]$, $-\frac{1}{T(H,\omega)}$ is strictly increasing in ω. Thus,

$$\begin{aligned} & \frac{1-F_H(y')}{\int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega)d\omega} < \frac{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) f_H(\omega)d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) T(H,\omega) f_H(\omega)d\omega} \\ \Leftrightarrow \quad & 1-F_H(y') < \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) f_H(\omega)d\omega. \end{aligned} \tag{132}$$

Applying Lemma 9 again, since $-\frac{f_L(\omega)}{f_H(\omega)}$ is strictly increasing in ω,

$$\frac{1-F_L(y')}{1-F_H(y')} < \frac{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) f_L(\omega)d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) f_H(\omega)d\omega}. \tag{133}$$

Similarly, we have:

$$\int_{y'}^{\overline{\omega}} [D(L,\underline{\omega}) - D(H,\omega)] f_H(\omega)d\omega \le \int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) [D(L,\underline{\omega}) - D(H,\omega)] f_H(\omega)d\omega \tag{134}$$

$$\frac{\int_{y'}^{\overline{\omega}} [D(L,\omega) - D(L,\underline{\omega})] f_L(\omega)d\omega}{\int_{y'}^{\overline{\omega}} [D(H,\omega) - D(L,\underline{\omega})] f_H(\omega)d\omega} \ge \frac{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) [D(L,\omega) - D(L,\underline{\omega})] f_L(\omega)d\omega}{\int_{\underline{\omega}}^{\overline{\omega}} \pi(s_1|\omega) [D(H,\omega) - D(L,\underline{\omega})] f_H(\omega)d\omega} \tag{135}$$

Condition (b) implies

$$\bar{t} < V(\xi, s_1, \pi) \le V(H, s_1, \pi). \tag{136}$$

Moreover, $\bar{t} + D(L,\underline{\omega}) \ge V(\xi, s_2, \pi) + D(L,\underline{\omega}) > P(L,\underline{\omega}) + D(L,\underline{\omega}) \ge 0$, and

$$\frac{1-F_L(y')}{1-F_H(y')} > \frac{[D(L,\omega) - D(L,\underline{\omega})] f_L(\overline{\omega})}{[D(H,\omega) - D(L,\underline{\omega})] f_H(\overline{\omega})} \ge \frac{\int_{y'}^{\overline{\omega}} [D(L,\omega) - D(L,\underline{\omega})] f_L(\omega)d\omega}{\int_{y'}^{\overline{\omega}} [D(H,\omega) - D(L,\underline{\omega})] f_H(\omega)d\omega}. \tag{137}$$

Therefore,

$$\begin{aligned} & r_L(\pi', \gamma') - r_H(\pi', \gamma') \\ = \quad & g(L, s_1, \pi') \bar{t}' - g(H, s_1, \pi') \bar{t}' + \int_{y'}^{\overline{\omega}} D(L,\omega) f_L(\omega)d\omega - \int_{y'}^{\overline{\omega}} D(H,\omega) f_H(\omega)d\omega \\ = \quad & \left[\frac{1-F_L(y')}{1-F_H(y')} - 1\right] \int_{y'}^{\overline{\omega}} T(H,\omega) f_H(\omega)d\omega + \int_{y'}^{\overline{\omega}} D(L,\omega) f_L(\omega)d\omega \\ & - \frac{[1-F_L(y')] \int_{y'}^{\overline{\omega}} D(H,\omega) f_H(\omega)d\omega}{1-F_H(y')} \end{aligned}$$

$$
\begin{aligned}
&= \left[\frac{1-F_L(y')}{1-F_H(y')}-1\right]\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)T(H,\omega)f_H(\omega)d\omega+\int_{y'}^{\overline{\omega}}D\left(L,\omega\right)f_L(\omega)d\omega \\
&\quad -\frac{[1-F_L(y')]\int_{y'}^{\overline{\omega}}D\left(H,\omega\right)f_H(\omega)d\omega}{1-F_H(y')} \\
&= \left[\frac{1-F_L(y')}{1-F_H(y')}-1\right]\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[P\left(H,\omega\right)+D(L,\underline{\omega})\right]f_H(\omega)d\omega \\
&\quad +\frac{[F_H(y')-F_L(y')]\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[D\left(H,\omega\right)-D(L,\underline{\omega})\right]f_H(\omega)d\omega}{1-F_H(y')} \\
&\quad +\int_{y'}^{\overline{\omega}}\left[D(L,\underline{\omega})-D\left(H,\omega\right)\right]f_H(\omega)d\omega\,* \\
&\quad \left\{\frac{1-F_L(y')}{1-F_H(y')}-\frac{\int_{y'}^{\overline{\omega}}\left[D\left(L,\omega\right)-D(L,\underline{\omega})\right]f_L(\omega)d\omega}{\int_{y'}^{\overline{\omega}}\left[D\left(H,\omega\right)-D(L,\underline{\omega})\right]f_H(\omega)d\omega}\right\} \\
&< \left[\frac{\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)f_L(\omega)d\omega}{\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)f_H(\omega)d\omega}-1\right]\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)f_H(\omega)d\omega\left[\bar{t}+D(L,\underline{\omega})\right] \\
&\quad +\frac{[F_H(y')-F_L(y')]\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[D\left(H,\omega\right)-D(L,\underline{\omega})\right]f_H(\omega)d\omega}{1-F_H(y')} \\
&\quad +\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[D(L,\underline{\omega})-D\left(H,\omega\right)\right]f_H(\omega)d\omega\,* \\
&\quad \left\{\frac{1-F_L(y')}{1-F_H(y')}-\frac{\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[D\left(L,\omega\right)-D(L,\underline{\omega})\right]f_L(\omega)d\omega}{\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[D\left(H,\omega\right)-D(L,\underline{\omega})\right]f_H(\omega)d\omega}\right\} \\
&= \int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[\bar{t}+D\left(L,\omega\right)\right]f_L(\omega)d\omega-\int_{\underline{\omega}}^{\overline{\omega}}\pi(s_1|\omega)\left[\bar{t}+D\left(H,\omega\right)\right]f_H(\omega)d\omega \\
&= r_L\left(\pi,\gamma\right)-r_H\left(\pi,\gamma\right) \qquad (138)
\end{aligned}
$$

The inequality follows (134), (135) and that either (133) or (136) holds. We know $r_H(\pi',\gamma')=r_H(\pi,\gamma)$, and this completes the proof. **Q.E.D.**

Appendix D: The sorting assumption

Maskin and Tirole (1992) consider a general setup in which a designer and a player have payoffs $R_\theta\left(y,a\right), U\left(\theta,y,a\right)$, respectively, where y is a vector of actions; a is a payment to the designer from the player; and θ denotes the designer's private type. They impose the following sorting assumption: $\left(-\frac{\partial R_\theta(y,a)}{\partial y}/\frac{\partial R_\theta(y,a)}{\partial a}\right) > (-\frac{\partial R_{\theta'}(y,a)}{\partial y}/\frac{\partial R_{\theta'}(y,a)}{\partial y})$

for $\theta < \theta'$. This is a discrete state analogue of the "Spence–Mirrlees" single crossing condition (see Fudenberg and Tirole (1991), p. 506). With this assumption, the RSW mechanism has a simple structure and can be solved successively with only adjacent upward incentive constraints. Moreover, the deterministic RSW mechanism is interim efficient relative to some non-degenerate belief, which is used to prove the main theorem of their paper: The RSW mechanism and any feasible mechanism that weakly dominates it can be supported as a PBE. Furthermore, the RSW mechanism is the unique mechanism that survives the intuitive criterion. The literature on the informed principal problem with common values (e.g., Balkenborg and Makris (2015) and Bedard (2017)) imposes different versions of sorting assumption to guarantee desirable properties.

In this paper, since we allow the designer to choose any information design and mechanism design, action space cannot be represented by two variables. Therefore there is no straightforward version of the sorting assumption. In the model with binary alternatives, suppose we restrict the package to a monotone binary partition with a cutoff y, a participation fee, and a payment difference equal to the expected return to the player conditional on that the designer's type is high and that the state of nature is above the chosen cutoff. There are two variables: cutoff y and participation fee I.

$$R_\theta(y,a) = [1 - F_\theta(y)] \frac{\int_y^{\overline{\omega}} P(H,\omega) f_H(\omega)\, d\omega}{1 - F_H(y)} + \int_y^{\overline{\omega}} D(\theta,\omega) f_\theta(\omega)\, d\omega + I, \tag{139}$$

$$U(\theta,y,a) = \int_y^{\overline{\omega}} P(\theta,\omega) f_\theta(\omega)\, d\omega - [1 - F_\theta(y)] \frac{\int_y^{\overline{\omega}} P(H,\omega) f_H(\omega)\, d\omega}{1 - F_H(y)} - I \tag{140}$$

The sorting assumption requires that

$$\begin{aligned}
& \frac{\partial [1 - F_L(y)] \frac{\int_y^{\overline{\omega}} P(H,\omega) f_H(\omega) d\omega}{1 - F_H(y)} + \int_y^{\overline{\omega}} D(L,\omega) f_L(\omega)\, d\omega}{\partial y} \\
> \quad & \frac{\partial \int_y^{\overline{\omega}} P(H,\omega) f_H(\omega)\, d\omega + \int_y^{\overline{\omega}} D(H,\omega) f_H(\omega)\, d\omega}{\partial y} \\
\Leftrightarrow \quad & \frac{dM(y)}{dy} < 0. \qquad (141)
\end{aligned}$$

However, in our model, $M\left(\underline{\omega}\right) < M\left(N^{-1}\left(x_H^{\#}\right)\right)$, which violates (141). This paper provides an environment in which the results of Maskin and Tirole Maskin and Tirole (1992) still apply without the sorting assumption.

References

Akerlof, G. A. (1978). The market for "lemons": Quality uncertainty and the market mechanism. In *Uncertainty in economics*, pp. 235–251. Elsevier.

Alonso, R. and O. Câmara (2016). Political disagreement and information in elections. *Games and Economic Behavior 100*, 390–412.

Alonso, R. and O. Câmara (2018). On the value of persuasion by experts. *Journal of Economic Theory 174*, 103–123.

Balestrieri, F. (2008). A modified english auction for an informed buyer. In *Game Theory and Management.*, pp. 14.

Balestrieri, F. and S. Izmalkov (2016). Informed seller in a hotelling market. *Available at SSRN 2398258*.

Balkenborg, D. and M. Makris (2015). An undominated mechanism for a class of informed principal problems with common values. *Journal of Economic Theory 157*, 918–958.

Bedard, N. C. (2017). The strategically ignorant principal. *Games and Economic Behavior 102*, 548–561.

Bergemann, D. and S. Morris (2016). Information design, bayesian persuasion, and bayes correlated equilibrium. *American Economic Review 106*(5), 586–91.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Chen, Y. and J. Zhang (2020). Signalling by bayesian persuasion and pricing strategy. *The Economic Journal 130*(628), 976–1007.

Chen, Y.-C., W. He, J. Li, and Y. Sun (2019). Equivalence of stochastic and deterministic mechanisms. *Econometrica 87*(4), 1367–1390.

Cho, I.-K. and D. M. Kreps (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics 102*, 179–221.

Dosis, A. (2019). On the informed principal model with common values. *Available at SSRN 2844942*.

Dworczak, P. and G. Martini (2019). The simple economics of optimal persuasion. *Journal of Political Economy 127*(5), 1993–2048.

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

Francetich, A. and P. Troyan (2012). Collusive agreements in auctions: Design and execution by an informed principal. *Working paper*.

Fudenberg, D. and J. Tirole (1991). Game theory, 1991. *Cambridge, Massachusetts 393*(12), 80.

Gal-Or, E. (1982). Hotelling's spatial competition as a model of sales. *Economics Letters 9*(1), 1–6.

Gershkov, A., J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi (2013). On the equivalence of bayesian and dominant strategy implementation. *Econometrica 81*(1), 197–220.

Hedlund, J. (2017). Bayesian persuasion by a privately informed sender. *Journal of Economic Theory 167*, 229–268.

Hoffmann, F. and R. Inderst (2011). Pre-sale information. *Journal of Economic Theory 146*(6), 2333–2355.

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

Koessler, F. and V. Skreta (2016). Informed seller with taste heterogeneity. *Journal of Economic Theory 165*, 456–471.

Koessler, F. and V. Skreta (2019). Selling with evidence. *Theoretical Economics 14*(2), 345–371.

Koessler, F. and V. Skreta (2022). Informed information design.

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics 13*(2), 607–635.

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

Kolotilin, A. and A. Wolitzky (2020). Assortative information disclosure.

Krähmer, D. (2020). Information disclosure and full surplus extraction in mechanism design. *Journal of Economic Theory*, 105020.

Li, H. and X. Shi (2017). Discriminatory information disclosure. *American Economic Review 107*(11), 3363–85.

Maskin, E. and J. Tirole (1990). The principal-agent relationship with an informed principal: The case of private values. *Econometrica*, 379–409.

Maskin, E. and J. Tirole (1992, January). The principal-agent relationship with an informed principal, ii: Common values. *Econometrica 60*(1), 1–42.

Mathevet, L., J. Perego, and I. Taneva (2020). On information design in games. *Journal of Political Economy 128*(4), 1370–1404.

Myerson, R. B. (1983, November). Mechanism design by an informed principal. *Econometrica 51*(6), 1767–97.

Myerson, R. B. (1985). *"Analysis of two bargaining problems with incomplete information" in Game Theoretic Models of Bargaining, A. Roth (ed.).* Cambridge University Press, London.

Mylovanov, T. and T. Tröger (2012). Informed-principal problems in environments with generalized private values. *Theoretical Economics 7*(3), 465–488.

Mylovanov, T. and T. Tröger (2014). Mechanism design by an informed principal: Private values with transferable utility. *The Review of Economic Studies 81*(4), 1668–1707.

Narasimhan, C. (1988). Competitive promotional strategies. *Journal of business*, 427–449.

Nishimura, T. (2022). Informed principal problems in bilateral trading. *arXiv preprint arXiv:1906.10311*.

Perez-Richet, E. (2014). Interim bayesian persuasion: First steps. *The American Economic Review 104*(5), 469–474.

Quesada, L. (2005). Collusion as an informed principal problem. *Available at SSRN 802925*.

Segal, I. and M. D. Whinston (2003). Robust predictions for bilateral contracting with externalities. *Econometrica 71*(3), 757–791.

Severinov, S. (2008, July). An efficient solution to the informed principal problem. *Journal of Economic Theory 141*(1), 114–133.

Skreta, V. (2011). On the informed seller problem: optimal information disclosure. *Review of Economic Design 15*(1), 1–36.

Sobel, J. (1984). The timing of sales. *The Review of Economic Studies 51*(3), 353–368.

Tan, G. (1996). Optimal procurement mechanisms for an informed buyer. *Canadian Journal of Economics*, 699–716.

Taneva, I. (2019). Information design. *American Economic Journal: Microeconomics 11*(4), 151–85.

Varian, H. R. (1980). A model of sales. *The American economic review 70*(4), 651–659.

Yilankaya, O. (1999). A note on the seller's optimal mechanism in bilateral trade with two-sided incomplete information. *Journal of Economic Theory 87*(1), 267–271.

Zhao, X. (2018). Auction design by an informed seller: The optimality of reserve price signaling. *Working Paper, University of Technology Sydney*.