Mechanism Design with Endogenous Principal Learning*

Daniel Clark† Yingkai Li‡

Abstract

We study a principal-agent framework where the principal designs a mechanism which consists of an allocation rule and an information structure for learning payoff relevant features. Crucially, the principal cannot commit on how acquired information is used in the mechanism. We show that in binary feature environments or those with independent private values and quasilinear transfers, there exists an optimal mechanism with full revealing information structures. We provide sufficient conditions where mechanisms with full revealing information structures are optimal or strictly suboptimal in environments like interdependent values or correlated types. In general, optimal mechanisms may require information structures with supports strictly larger than the feature set, contrasting results in standard information design.

Keywords— Mechanism design, learning, fully revealing information structures.
JEL— D82, D83

*The authors thank Larry Samuelson, Kai Hao Yang, and the audience at UCLA, CUHK-CUHK(SZ) Joint Theory Workshop and Singapore Joint Economic Theory Workshop for helpful comments and suggestions. Part of this work was completed while Daniel Clark and Yingkai Li were postdocs at Yale University. Yingkai Li thanks Sloan Research Fellowship FG-2019-12378 and NUS Start-up Grant for financial support.

†Department of Economics, UCLA. Email: dclark@econ.ucla.edu.
‡Department of Economics, National University of Singapore. Email: yk.li@nus.edu.sg.

1 Introduction

Information plays a crucial role in the operations of firms, organizations, and governments. Moreover, modern technology has revolutionized the ways in which information is gathered and processed, enabling many organizations to collect vast amounts of data from diverse sources at extremely high speeds and to make use of such information in various applications. For instance, auto insurance companies, such as Progressive, adopt programs like Snapshot to monitor the real-time driving behavior of policyholders and provide personalized insurance rates based on usage;[1] retailers such as Walmart collaborate with manufacturers using the Collaborative Planning, Forecasting, and Replenishment (CPFR) system to share demand forecasts for improved inventory control (Seifert, 2003); online platforms such as Google AdWords adopt sophisticated learning algorithms to forecast the clicks bidders can expect at a given bid, and provide these forecasts as a bidding landscape to guide bidders in AdWords auctions (Deng et al., 2021); and manufacturers such as Ford use automated vision systems for gathering information on the defects in product lines to improve the overall qualities of their products.[2]

The traditional wisdom is that more information should always help for making better decisions. However, when the information is used for contracting with strategic agents, due to privacy concerns, regulation policies on data protection, or simply the inability to credibly disclose information, the principal cannot commit on how they will use their private information when establishing the contracts. This raises the additional concern of trusting the principal to act in a way that is beneficial for the agents, and hence the contracts designed by the principal must also be incentive compatible for themselves. In this paper, we study the problem where the principal can jointly design the mechanisms and the information structures for learning payoff relevant features. We show that the principal may strategically neglect useful information to avoid the loss from satisfying the incentive constraints.

A leading example of our model is forecast sharing in supply chains. In this application, retailers such as Walmart and Target control both the terms of the contracts for replenishment from manufacturers and the technologies used for demand forecasting.[3] The ex-ante unknown payoff relevant feature in this example is the future market demands for various products and the technologies for making the demand forecasts are publicly ob-

[1] See `https://www.progressive.com/auto/discounts/snapshot/`.

[2] See `https://www.vision-systems.com/factory/article/16743188/automated-vision-system-creates-3d-model-of-ford-cars-to-detect-dirt-in-paint-jobs`

[3] Similarly, large manufacturing companies such as Coca-Cola can use the CPFR system to share demand forecasts with upstream suppliers for more efficient production.

served by both the retailer and the manufacturer. For instance, IBM provides solutions for implementing the CPFR system, in which transparency in demand forecasting technologies can be ensured among all parties. This solution has been adopted by various companies. Furthermore, recent advancements in blockchain technology have further enhanced transparency, which IBM has adapted to provide services for companies such as Walmart and Pfizer.[4] Transparency in forecasting technologies ensures that all parties have a shared understanding of the best possible predictions the retailer can make based on its proprietary data, and a better prediction technology corresponds to the retailer's choice of acquiring more precise information.[5] Moreover, as noted by (Deimen and Szalay, 2019), who consider applications where the transparency of information structures is justified in in-house consulting, even if the principal has the option to keep the information structure private, they always benefit from publicly disclosing it.

Nonetheless, the consumer data used for forecasting is proprietary to the retailers and will not be publicly shared with manufacturers due to business or regulatory concerns. In this case, retailers may provide partial consumer data or even falsify it to generate inaccurate demand forecasts, misleading manufacturers for their own benefit if the mechanism is not designed to be incentive compatible for sharing the demand forecasts. For instance, a retailer may attempt to withhold a portion of the sales data to generate a lower demand forecast, allowing them to acquire replenishment from manufacturers at a reduced price based on the contracts. Indeed, building trust is one of the key concerns regarding credible information sharing in supply chains (Özer et al., 2011; Ebrahim-Khanjari et al., 2012). Therefore, the implementation of the CPFR system and the associated contracts needs to be designed to be incentive compatible for the retailers to truthfully share data for accurate demand forecasts.

Throughout the paper, we make the simplifying assumption that learning is costless for the principal (i.e., the retailer in the above example), as our focus is on issues concerning the value of information in the presence of incentive constraints. We first provide a simple illustration showing that the principal may have a strict incentive not to fully learn the unknown features, even in environments where the information gathered by the principal is payoff-irrelevant for the agent if transfers are not allowed. Note that in the context of supply chains, this includes environments with monetary payments, provided that the prices of the products are fixed and non-negotiable. In essence, the requirement of "no transfers"

[4] `https://medium.com/@ieeecomputersocietyiit/breaking-boundaries-how-ibm-walmart-and-pfizer-lead-the-blockchain-revolution-0eb13d5cbac2`

[5] In the application of online platforms, a common understanding of the best possible forecasting can also be achieved through the platform's publicly disclosed data policy, which specifies that certain aspects of user information will not be collected for business purposes.

means that the principal cannot design contracts with arbitrary contingent transfers based on the reports.

Next, we elaborate on the example where full learning is not optimal for the principal. In this example, there are three possible demands: high, medium, and low, each with equal prior probabilities. The prices of the products are not negotiable, but the quantities are. Specifically, the retailer can choose to acquire nothing, a small quantity, or a large quantity from the manufacturer based on the demand forecasts. For illustration, it suffices to specify the net benefits (considering market price, inventory costs, production costs, etc.) for both the retailer and the manufacturer for each quantity transaction given each demand forecast. The net benefits of acquiring nothing (the outside option) are normalized to 0, while the net benefits of acquiring a small quantity are 10 and -2 for the retailer and the manufacturer, respectively, regardless of demand. The net benefit of acquiring a large quantity is 1 for the manufacturer, regardless of demand, while it is 20 if the demand is high, 9 if medium, and 3 if low for the retailer. In this example, if the retailer can perfectly forecast the demand, they would prefer to maximize the probability of purchasing only a small quantity when demand is medium or low, while the manufacturer's individual rationality constraint implies that these probabilities can be at most $\frac{1}{2}$. This further indicates that the retailer's expected revenue is at most 12. However, if the retailer can only forecast whether the demand is low or not, an incentive compatible and individually rational contract would involve purchasing a small quantity from the manufacturer if the demand is low and a large quantity otherwise. The expected revenue under this contract is 13, which is strictly higher than the optimal profit from full learning.

The strict suboptimality of full learning in the above illustration relies heavily on the assumption that transfers are not allowed. When transfers are permitted, consider a mechanism where the retailer provides a subsidy of 2 to the manufacturer for purchasing a small quantity and requests a price reduction of 1 for purchasing a large quantity. This mechanism is equivalent to vertically integrating the manufacturer with an ex ante transfer of 0. Under this arrangement, by internalizing the profitability of the products from the manufacturer, the retailer's optimal choice, given a perfect demand forecast, is to purchase a small quantity if the demand is low and a large quantity otherwise. The expected utility of the retailer under this mechanism is 13, which is optimal. The intuition is that transfers help align the incentives between the principal and the agent, even when the principal fully learns the unknown payoff-relevant features. This idea extends beyond the illustrated example; it also applies to environments where the agent has additional private information, which may make simple vertical integration for full surplus extraction ineffective.

The first main result of our paper is to show that in any independent private value

environment with quasilinear preference on the transfers, there always exists an optimal mechanism where the unknown feature is fully revealed to the principal (Theorem 1). The high level idea is that given any optimal mechanism, there exists an alternative mechanism with fully revealing information structures where the principal can always rearrange the allocations for each feature to maximize the principal's utility without affecting the marginal probability distribution on the allocations for each agent type. Moreover, a transfer rule can be carefully chosen to ensure that the mechanism with adjusted allocations is incentive compatible for the principal without changing the expected transfers for each agent type. Since this is an independent private value environment, the agent's expected utility will not be affected and hence the incentive constraints for the agent remains intact as well. Therefore, the alternative mechanism is feasible and weakly increases the principal's utility, and hence is optimal for the principal as well. Note that this result can also be easily extended when the principal's utility depends on the agent's private type.

The private value and independence assumptions are also crucial for the optimality of full learning. Specifically, if the agent's utility depends on unknown features, as in the market for lemons (e.g., Akerlof, 1970), or if the agent's private type is correlated with the unknown features (e.g., Crémer and McLean, 1988), the principal may benefit from not fully learning these features, even with transfers. In these environments, we provide sufficient conditions under which an optimal mechanism with fully revealing information structures always exists if transfers are allowed (see Proposition 2 for interdependent values and Proposition 3 for correlated environments). We also illustrate how the principal's utility can be strictly improved using partially revealing information structures when these conditions are violated. A notable aspect of correlated environments is that, unlike in Crémer and McLean (1988), full surplus extraction is generally not possible due to the additional principal's incentive constraints.

In our previous illustration where mechanisms with fully revealing information structures are strictly suboptimal, the result in fact also relies on the richness of the unknown feature space. That is, in the illustration there exist multiple distinct features (i.e., demands) such that the principal favors acquiring a small quantity. We show that when the feature space is restricted, in particular when it is binary, mechanisms with fully revealing information structures are optimal for the principal (Theorem 2). This observation holds broadly when there are correlation or interdependent values, regardless of whether transfers are allowed. In contrast, when there is a rich feature space, mechanisms with fully revealing information structures in general are strictly suboptimal when transfers are not allowed even in independent private value environments (Proposition 6).

As we illustrated above, the principal may benefit from pooling features in general

environments, with or without transfers. However, we show that simply combining different features may not be optimal for the principal, and more complex information structure are required in order to alleviate the concerns of incentive constraints. In particular, when transfers are not allowed or the agent has non-degenerate type space, there exist instances in which the number of signals in the optimal information structure is strictly larger than the number of possible features (Examples 1 and 2). This highlights the need to create a rich signal space to partially and randomly pool unknown features in order to satisfy the incentive constraints without too much loss on efficiency. Intuitively, with a richer signal space, the principal can design mechanisms that suffer from a smaller efficiency loss while maintaining incentive compatibility constraints, which ultimately leads to a higher expected payoff for the principal. This leads to a sharp contrast to the classical information design literature where Carathéodory's theorem implies that the number of signals is at most the number of states (Kamenica and Gentzkow, 2011). We complement this observation by showing that Carathéodory's theorem applies in our model when transfers are allowed and the agent has degenerate type space (Proposition 5).

Finally, this paper focuses for the most part on environments where the principal's interim individual rationality constraints are ignored. This is plausible if the principal is not protected by limited liability. However, there also exist environments in which the principal will not forgo their outside options after the agent agrees to participate in the mechanism. In such environments, with independent private values and quasilinear transfers, fully revealing information structure is still optimal for the principal (Proposition 7). However, the individual rationality constraints could impact the principal's incentives for fully learning the features outside such canonical environments. For example, with binary feature space, if individual rationality constraints are imposed, there exist instances in which the optimal information structure may require at least three signals, as we show in Section 5.

1.1 Related Work

Several papers studying joint mechanism-information design problems have recently emerged. Papers that share our focus on information received by the principal include Bergemann et al. (2015), Haghpanah and Siegel (2023), and Kartik and Zhong (2023). Both Bergemann et al. (2015) and Haghpanah and Siegel (2023) focus on characterizing implementable principal-agent utility pairs, while our work emphasizes the principal's ex-ante optimization problem. Kartik and Zhong (2023) examines interdependent values with design of information structures for both parties. Unlike these studies where the principal can commit to truthful reporting, our focus on principal incentive compatibility highlights that fully revealing information structures can be suboptimal for the principal. Other relevant works on

joint design emphasize information disclosures made to agents to influence their willingness to pay, such as Bergemann and Pesendorfer (2007), Daskalakis et al. (2016), Bergemann et al. (2022), and Wei and Green (2024). Additionally, papers like Bergemann et al. (2018), Li (2022), and Yang (2022) investigate the pricing of information. Eső and Szentes (2007) and Li and Shi (2017) study models involving both of these considerations.

Due to our focus on learning by principals in principal-agent environments, our paper is closely related to the literature on information acquisition by a sender in cheap-talk games (e.g., Ivanov (2010), Kreutzkamp and Lou (2024), and Lyu and Suen (2022)). Outside of cheap-talk games, Li and Xu (2024) studies a principal-agent environment in which the principal learns some underlying state before playing a coordination game with the agent. Like our paper, these papers assume that the information structure used in the principal's learning is public knowledge, and they address issues of incentive compatibility related to how the relevant party uses the acquired information; specifically, the principal/sender cannot commit to how they will utilize the information they obtain.[6] Some other papers that focus on environments in which a principal/sender learns through covert means before interacting with their counterparty include Pavan and Tirole (2023). However, in all of these papers, the principal/sender has no control over the rules governing their interaction with their counterparty after their initial learning has concluded. In contrast, in this paper, we take a mechanism design perspective and thus afford the principal flexibility in designing the rules that govern their interaction with the agent.

Our paper is conceptually related to the literature on strategic ignorance (e.g., Kessler, 1998; Creane, 1998; Taneva and Wiseman, 2024) where the designer benefits from strategically ignoring payoff-relevant information when they cannot commit to how that information will be used. The benefit of strategic ignorance has also been identified in lemon markets (Akerlof, 1970), Stackelberg games for maintaining first-mover advantages (Gal-Or, 1987), risk sharing markets (Hirshleifer, 1971), buyer optimal learning (Roesler and Szentes, 2017), communication and delegation (Deimen and Szalay, 2019), and etc. In contrast, our paper provides two general environments where fully revealing information structures are optimal for the principal despite the additional concerns for incentive constraints.

Our paper is also related to the literature studying informed principal problems (e.g., Myerson (1981), Maskin and Tirole (1990), Maskin and Tirole (1992), Mylovanov and Tröger (2012, 2014), Koessler and Skreta (2023), Clark (2024a,b), and Clark and Yang (2024)). However, in the environments studied by these papers, the principal is privately informed before contracting occurs, while, crucially in our model, the principal can become privately

[6]Our requirement of incentive compatibility for the principal also relates to the literature on credible auctions (e.g. Akbarpour and Li, 2020; Ferreira and Weinberg, 2020).

informed only after they have committed to and implemented a mechanism.

2 Model

2.1 Preliminaries

There is a principal and an agent. (For convenience, throughout the paper, we study settings with only one agent, but all of our results extend naturally to settings with multiple agents.) There is a non-empty set of possible features Ω as well as a non-empty set of possible agent types Θ. We assume that both Ω and Θ are finite. Moreover, the realized feature-agent type pair $(\omega, \theta) \in \Omega \times \Theta$ is distributed ex-ante according to distribution $F \in \Delta(\Omega \times \Theta)$, which is commonly known by the principal and agent. We let $F_\Omega \in \Delta(\Omega)$ denote the marginal distribution over realized features obtained from F and $F_\Theta \in \Delta(\Theta)$ denote the marginal distribution over realized agent types obtained from F, and we assume that both F_Ω and F_Θ are full-support. Additionally, we let $\mathbf{F}_\Theta : \Omega \to \Delta(\Theta)$ be the mapping that gives the conditional probability distribution of Θ given Ω under F and $\mathbf{F}_\Omega : \Theta \to \Delta(\Omega)$ be the mapping that gives the conditional probability distribution of Ω given Θ under F.

The agent's type is directly observed by the agent before they interact with the principal; however, there is no information gained about the feature by either party until their interaction commences. Instead, the principal commits to a mechanism which, should it be accepted by the agent, among its other purposes, dictates the manner in which the principal learns about the underlying feature. (We will formulate the other purposes of such mechanisms shortly.) Specifically, the mechanism commits to an information structure (S, σ), which is a tuple consisting of a signal space S, which is a non-empty compact metric space, and a measurable mapping $\sigma : \Omega \to \Delta(S)$ from underlying features to probability distributions over signals. The ultimate signal realization of the principal's information structure can be viewed as the principal's endogenously acquired "type." While, in principle, we could allow the principal to choose from all possible information structures, familiar arguments show that it is without loss for our purposes to restrict the principal to choosing from "canonical" information structures in which $S = \Delta(\Omega)$ and σ gives a regular conditional probability distribution over Ω given $\Delta(\Omega)$ under the probability distribution over $\Omega \times \Delta(\Omega)$ that would be generated by F_Ω and σ. (Ignoring technical qualifications, what this captures intuitively is that the realized signal $s \in \Delta(\Omega)$ almost surely coincides with the Bayesian posterior that would be held by a principal with initial prior F_Ω upon observing s under information structure $(\Delta(\Omega), \sigma)$.) Motivated by this, we will identify an arbitrary canonical information structure $(\Delta(\Omega), \sigma)$ with its underlying mapping σ. We let $\mathcal{I}$ denote the set of

canonical information structures.

Aside from informational aspects of the environment such as the underlying feature or agent type, the payoffs of the parties are affected by the ultimate allocation. The space of possible allocations is a non-empty metric space denoted by X. The principal's utility function is $U : \Omega \times \Theta \times X \to \mathbb{R}$ and the agent's utility function is $V : \Omega \times \Theta \times X \to \mathbb{R}$. We assume that both U and V are continuous. Additionally, both parties have an outside option, and each party's outside option gives them a payoff of 0 regardless of $(\omega, \theta) \in \Omega \times \Theta$.

The mechanism that the principal chooses controls not only how they learn about the underlying feature but also how allocations are determined. We will impose familiar restrictions on the mechanisms we consider. In particular, it can be shown, using a version of the revelation principle appropriate for our setting, that it is without loss of generality for our purposes to restrict attention to direct mechanisms that are incentive compatible and satisfy various forms of individual rationality. Specifically, a direct mechanism $M = (\sigma, \mathbf{x})$ is a tuple consisting of a canonical information structure σ and an allocation rule of the form $\mathbf{x} : \Delta(\Omega) \times \Theta \to \Delta(X \cup \{o\})$. Throughout the paper, o denotes the outside options being realized. We will abuse notation by having $U(\omega, \theta, o) = V(\omega, \theta, o) = 0$ for all $(\omega, \theta) \in \Omega \times \Theta$. We interpret the principal proposing a mechanism $M = (\sigma, \mathbf{x})$ in which $\mathbf{x}(s, \theta) = \delta_o$ for all $(s, \theta) \in \Delta(\Omega) \times \Theta$ as the principal choosing to not form a relationship with the agent and instead have the outside options be realized. We let $\mathcal{M}$ denote the set of direct mechanisms.

The timing of the interaction we consider is as follows. First, nature draws a feature-agent type pair $(\omega, \theta) \in \Omega \times \Theta$ according to the probability distribution $F \in \Delta(\Omega \times \Theta)$, the agent observes their realized type $\theta \in \Theta$, and the principal proposes a direct mechanism $(\sigma, \mathbf{x}) \in \mathcal{M}$. The agent then observes the proposed mechanism $(\sigma, \mathbf{x})$ and either rejects, in which case both parties receive their outside options, or accepts, in which case their interaction proceeds governed by the mechanism M. In particular, if the agent accepts a proposal of $(\sigma, \mathbf{x})$, then nature draws a signal $s \in \Delta(\Omega)$ according to the information structure $\sigma(\omega) \in \Delta(\Delta(\Omega))$.[7] The principal observes the realized signal s, and then the principal and agent simultaneously submit type reports. Regardless of their true signal $s \in \Delta(\Omega)$, every $s' \in \Delta(\Omega)$ is a type report that the principal could choose; likewise, regardless of their true type $\theta \in \Theta$, every $\theta' \in \Theta$ is a type report that the agent could choose. If the principal reports type $s' \in \Delta(\Omega)$ and the agent reports type $\theta' \in \Theta$, then the resulting $x \in X \cup \{o\}$ is drawn according to $\mathbf{x}(s', \theta') \in \Delta(X \cup \{o\})$. Then payoffs are realized and the interaction concludes.

[7] Observe that, conditional on the underlying feature, the draw of the signal is statistically independent of the agent's type.

2.2 The Principal's Problems and Key Information Structures

The mechanism offered by the principal needs to be incentive compatible for both the principal and the agent as well as individually rational for the agent. (Regarding individual rationality for the principal, we will study environments in which no interim individual rationality constraints for the principal are imposed as well as environments in which the mechanism that the principal selects must be interim individually rational for all "types" of the principal.) For all $s \in \Delta(\Omega)$, let $G(s) \in \Delta(\Omega \times \Theta)$ be the probability distribution over $\Omega \times \Theta$ given by first drawing $\omega \in \Omega$ according to s and then drawing $\theta \in \Theta$ according to $\mathbf{F}_\Theta(\omega)$. Note that $G(s)$ gives the belief that the principal would hold over $\Omega \times \Theta$ after updating their prior F upon observing a signal that would lead them to hold s as their marginal conditional probability distribution over Ω. For all $\theta \in \Theta$ and $\sigma \in \mathcal{I}$, let $H(\theta, \sigma) \in \Delta(\Omega \times \Delta(\Omega))$ denote the probability distribution over $\Omega \times \Delta(\Omega)$ generated by first drawing $\omega \in \Omega$ according to $\mathbf{F}_\Omega(\theta)$ and then drawing $s \in \Delta(\Omega)$ according to $\sigma(\omega) \in \Delta(\Delta(\Omega))$. Note that $H(\theta, \sigma)$ gives the belief that the type θ agent would hold over the ultimate $(\omega, s) \in \Omega \times \Delta(\Omega)$ after observing the principal pick information structure σ. Our incentive compatibility constraints can be expressed as follows:

$$s \in \arg\max_{s' \in \Delta(\Omega)} \mathbf{E}_{(\omega,\theta)\sim G(s)} \left[\mathbf{E}_{x \sim \mathbf{x}(s',\theta)} \left[U(\omega, \theta, x) \right] \right] \quad \forall s \in \Delta(\Omega), \tag{PIC}$$

$$\theta \in \arg\max_{\theta' \in \Theta} \mathbf{E}_{(\omega,s)\sim H(\theta,\sigma)} \left[\mathbf{E}_{x \sim \mathbf{x}(s,\theta)} \left[V(\omega, \theta, x) \right] \right] \quad \forall \theta \in \Theta, \tag{AIC}$$

where (PIC) marks the principal incentive compatibility constraints and (AIC) marks the agent incentive compatibility constraints. The agent individual rationality constraints can be expressed as

$$\mathbf{E}_{(\omega,s)\sim H(\theta,\sigma)} \left[\mathbf{E}_{x \sim \mathbf{x}(s,\theta)} \left[V(\omega, \theta, x) \right] \right] \geq 0 \quad \forall \theta \in \Theta. \tag{AIR}$$

Additionally, each mechanism of interest must be such that, for every agent type $\theta \in \Theta$ and all principal types $s, s' \in \Delta(\Omega)$, the probability of o occurring given (s, θ) equals the probability of o occurring given (s', θ):

$$\mathbf{x}(s, \theta)[o] = \mathbf{x}(s', \theta')[o] \quad \forall s, s' \in \Delta(\Omega), \theta \in \Theta. \tag{Consistency}$$

This is because the outside options are realized if and only if the principal chooses to not form a relationship with the agent or the agent rejects the principal's proposal and, in both of these cases, neither party would have acquired information about the underlying feature before the final decisions were made. We say that a mechanism $M \in \mathcal{M}$ is *feasible* if and only if it satisfies the constraints given in (PIC), (AIC), (AIR), and (Consistency), and we

use $\mathcal{M}^{\mathrm{F}}$ to denote the set of feasible mechanisms.

For most of the paper, we will focus on the problem of finding feasible mechanisms that maximize the principal's ex-ante expected payoff across all feasible mechanisms:

$$\max_{(\sigma,\mathbf{x})\in\mathcal{M}^{\mathrm{F}}} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]. \tag{OPT}$$

However, we will also consider the problem of the principal maximizing their ex-ante expected utility with mechanisms that are feasible and also interim individually rational for the various principal types. More specifically, we will consider interim principal individual rationality constraints of the form

$$\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\geq 0 \quad \forall s\in\Delta(\Omega). \tag{PIR}$$

We will say that a mechanism M is *feasible and individually rational* if and only if it is feasible and satisfies the principal individual rationality constraints given by (PIR). We use $\mathcal{M}^{\mathrm{F,IR}}$ to denote the set of feasible and individually rational mechanisms, and we will consider the associated principal problem:

$$\max_{(\sigma,\mathbf{x})\in\mathcal{M}^{\mathrm{F,IR}}} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]. \tag{OPT-IR}$$

For all of the environments we study in this paper, both of these principal problems are well-defined and have solutions. We note here that existence can be established in several classes of environments of usual interest. For instance, standard arguments, which are presented in OA (1) of the Online Appendix, show that, for all primitive environments in which X is compact and U and V are continuous, both of these principal problems are well-defined and have solutions.

Proposition 1. *For all environments in which X is compact and U and V are continuous, the problem given by (OPT) and the problem given by (OPT-IR) have solutions.*

We will focus much of our study on the nature of information structures used in optimal mechanisms. Some benchmark canonical information structures are the *fully revealing* information structure, which is the information structure $\sigma^{FR}\in\mathcal{I}$ given by $\sigma^{FR}(\omega)=\delta_\omega$ for all $\omega\in\Omega$, and the *fully uninformative* information structure, which is the information structure $\sigma^{FU}\in\mathcal{I}$ satisfying $\sigma^{FU}(\omega)=F_\omega$ for all $\omega\in\Omega$. Often it is the case that every optimal mechanism has an information structure that is not fully uninformative since it can be beneficial to condition the allocation on informative signals of the underlying feature. Moreover, despite the principal's choice of information structure not directly impacting their

payoff, it is often the case that every optimal mechanism has an information structure that is not fully revealing. As we will see, the requirements of principal incentive compatibility can lead to garbled information structures facilitating mechanisms with strictly better payoffs for the principal than all feasible mechanisms with fully revealing information structures.

2.3 Quasilinear Environments

For much of the paper, we will specialize to the canonical class of quasilinear environments with transfers. In such environments, the allocation space X has a product structure of the form $X = Y \times \mathbb{R}$, where Y is a non-empty compact metric space. Additionally, for such environments, there are continuous functions $u : \Omega \times \Theta \times Y \to \mathbb{R}$ and $v : \Omega \times \Theta \times Y \to \mathbb{R}$ such that, for each $(\omega, \theta, (y, t)) \in \Omega \times \Theta \times X$, the corresponding principal utility satisfies

$$U(\omega, \theta, (y, t)) = u(\omega, \theta, y) + t$$

and the corresponding agent utility satisfies

$$V(\omega, \theta, (y, t)) = v(\omega, \theta, y) - t.$$

3 Optimal Mechanisms for Quasilinear Payoffs

In this section, we study our mechanism design problem in the canonical class of quasilinear environments with transfers formalized in Section 2.3. We first focus on a subclass of environments that includes the classic "independent private values" environments as well as, more broadly, all environments in which the underlying feature and agent type are statistically independent and the agent does not directly care about the underlying feature. We will see that, for all environments in this class of "IAPV" environments, there is an optimal mechanism that utilizes a fully revealing information structure. Outside of this class of environments, it can be that every optimal mechanism involves a non-fully-revealing information structure, which we demonstrate both in various interdependent environments that formalize some of the illustrative examples discussed earlier in the introduction, as well as in environments that are not interdependent but feature correlation between the underlying feature and agent type. We conclude the section by analyzing properties of the numbers of on-path signals used in the information structures of the optimal mechanisms in various quasilinear environments.

3.1 Independent and Agent Private Values

We begin with the natural benchmark in which the agent type and underlying feature are independent and the agent does not directly care about the feature. Formally, we say that a quasilinear environment with transfers exhibits ***independent and agent private values*** if and only if $F = F_\Omega \times F_\Theta$ and $v(\omega, \theta, x) = v(\omega', \theta, x)$ for all $\omega, \omega' \in \Omega, \theta \in \Theta, x \in X$. For brevity, we will sometimes refer to such environments as being ***IAPV***. We note that, in the terminology of Mylovanov and Tröger (2014), such environments would be said to have "generalized private values."

In all IAPV environments, fully learning the underlying feature is consistent with principal ex-ante optimality.

Theorem 1. *In all independent and agent private value environments, there is an optimal mechanism with a fully revealing information structure.*

We saw in the introduction an example of an IAPV environment and a (somewhat informal) description of an optimal mechanism with a fully revealing information structure. Key to that mechanism is that the presence of transfers enables the construction of a transfer scheme that perfectly aligns the incentives of a fully informed principal with efficient trade. Intuitively, the ability of the transfers to align the incentives of a fully informed principal with efficient allocation distributions (under appropriate notions of efficiency) extends generally across quasilinear environments.

We show in Appendix A that, for an arbitrary feasible mechanism, holding fixed an arbitrary agent type, there is an allocation-transfer scheme that would (1) result in that agent type obtaining the same expected transfer and marginal distribution over allocations as the original mechanism, (2) maximize the ex-ante principal surplus, conditional on the agent type, across all allocation-transfer schemes that give the agent type the same expected transfer and marginal distribution over allocations as the original mechanism, and (3) be incentive compatible for the principal. This is due to the presence of transfers and their quasilinearity in the principal's utility function, and can be seen as a consequence of the fact, which we do not explicitly develop here, that in all environments with quasilinear linear transfers and a single player with potentially multiple types, for an arbitrary marginal distribution over allocations, all allocation rules that achieve this marginal distribution and maximize the total expected surplus across all allocation rules that achieve this marginal distribution can be made incentive compatible with an appropriate transfer rule.[8]

[8]Other papers that have developed similar results for other settings include Bei and Huang (2011); Hartline et al. (2015).

Using this, it follows that, for an arbitrary feasible mechanism, we can obtain a mechanism with a fully revealing information structure that is incentive compatible for the principal, gives the principal a weakly higher ex-ante payoff than the original mechanism, and, for each agent type report, would result in the same expected transfer and marginal distribution over allocations as the original mechanism when the principal reports their observations truthfully with probability 1 regardless of the true agent type. Due to the agent's utility being quasilinear in transfer and not directly varying with the underlying feature, it follows that all such mechanisms must be incentive compatible and individually rational for the agent and thus feasible.

The justification that the fully revealing mechanisms used in the argument are incentive compatible for the agent relies on (1) the agent not directly caring about the underlying feature and (2) the agent's true type being uninformative about the distribution of the feature. In the next subsection, we will relax the first of these assumptions and in the following subsection, we will relax the second. In both subsections, we will see environments in which there is no optimal mechanism with a fully revealing information structure.

3.2 Agent Interdependent Values

To emphasize the effect on agent's interdependent values, in this section, we focus on settings where the types are independent, and the principal has private values. As illustrated in the introduction, when the agent has interdependent values, mechanisms with fully revealing information structures can be strictly suboptimal for the principal. To fully characterize the optimal mechanisms and identify conditions under which mechanisms with fully revealing information structures are optimal, we restrict our attention to a classic lemon's problem with linear utilities. Specifically, we assume that allocation space is $Y = \{0, 1\}$, type spaces $\Omega, \Theta \subset [0, 1]$, and there exists function $c : \Omega \to \mathbb{R}$ such that

$$U(\omega, \theta, y, t) = \omega(1 - y) + t \quad \text{and} \quad V(\omega, \theta, y, t) = (c(\omega) + \theta)y - t.$$

Proposition 2. *In the lemon's problem, fixing the utility function of both the principal and the agent and the distribution over unknown features,*

(1) $c(\omega) - \omega$ *is non-increasing in* ω *for all* ω *in the support of* F_Ω*; or*

(2) $c(\omega) - \omega$ *is linearly increasing in* ω *for all* ω *in the support of* F_Ω*.*

if and only if, for every agent type set and corresponding type distribution, there exists a mechanism with fully revealing information structure.

An immediate observation is that the condition in Proposition 2 is always satisfied when the distribution over features has binary support, and hence Proposition 2 also implies that mechanisms with fully revealing information structures are optimal in the lemon's problem, which is consistent with Theorem 2.

The first condition in Proposition 2 implies that the total surplus from selling the item to the agent is weakly decreasing in the principal's value for the item. This implies that there won't be any conflict of interests for selling the item to the agent. In particular, to maximize the total surplus from selling the item, items with lower values for the principal will be sold, and such allocation rule can be implemented in an incentive compatible way for the principal as the principal has stronger incentives to sell low value items. Therefore, in this case, mechanisms with fully revealing information structures are optimal since it provides the maximum amount information in order for the trade to occur more efficiently, and thereby extracting higher revenue from the agent.

The second condition in Proposition 2 implies that there is a linear relationship between the total surplus and the principal's value for the item. The benefits of the linear structure is that given any information structure for learning the unknown features, the pair of expected total surplus and principal's value given the posterior belief always lies on this straight line. In this case, regardless of the information structure, the optimal mechanism is to ignore the principal's report by pooling all types together. Therefore, mechanisms with any information structure, including fully revealing information structure, can be optimal in this case.

3.3 Correlated Types

In this section, we focus on settings where the feature is correlated with the agent's private type, and we further assume that both the principal and the agent have private values throughout the section.

We first consider the case where the feature is payoff irrelevant for the principal, i.e., $u(\omega, \theta, y) = u(\omega', \theta', y)$ for all $\omega, \omega' \in \Omega, \theta, \theta' \in \Theta, y \in Y$. This special environment includes applications where features represent personal data that are informative about the agents values for the item, which naturally are payoff irrelevant for the principal.

Proposition 3. *In private value environments, if the agent type set is binary and the unknown features are payoff irrelevant for the principal, there exists an optimal mechanism with a fully revealing information structure.*

We first show that if the agent type set is binary, it is without loss of optimality to only consider mechanisms that set principal's utilities that are independent of their own types,

which is formalized in Lemma 1. Proposition 3 then follows by the observation that for any mechanism with principal's utility independent from their reported type, there exists another mechanism with fully revealing information structure that simulates the original mechanism by randomly pooling the principal's report. This generates the same expected utility to the principal without violating the incentive constraints.

Lemma 1. *In private value environments, if the agent has binary type and the unknown features are payoff irrelevant for the principal, there exists an optimal mechanism such that the utility of the principal is independent of their type.*

Another immediate consequence of Lemma 1 is that the principal cannot extract full surplus from the agent in monopoly auctions without allocation costs even when distributions are correlated. This leads to a sharp contrast to Crémer and McLean (1988) where the incentive constraints for the principal are not required. The main reason is that due to the principal's incentive constraints, the transfers cannot depend on principal's additional information on agent's valuation, and hence the mechanism designed by the principal loses the ability to fully eliminate the information rents using signal-dependent prices.

Corollary 1. *In monopoly auctions without allocation costs, when the agent's type distribution has non-degenerate binary support* $0 < \theta_0 < \theta_1$*, the principal cannot extract full surplus from the agent.*

Next we provide an example to show that when the unknown features are payoff relevant for the principal, mechanisms with fully revealing information structures can be strictly suboptimal for the principal. In this example, the principal has binary actions, selling the good or not, and there are three different features where we use the value of the features to denote the cost of selling the good. The cost of not selling is normalized to 0. Note that there exists two features ($\omega = 0.5$ and 0.8 in Table 1) such that the principal's value is positively correlated with the agent's value for the item, i.e., higher agent type occurs with higher probability conditional on the principal's value for the item being higher given the realized feature. We show that when the feature is fully revealed to the principal, the mechanism that maximizes principal's payoff is to provide the same allocation and transfer rules for both features $\omega = 0.5$ and 0.8. We then show that the principal can strictly improve her payoff by pooling those two features to relax the incentive constraints from the other feature $\omega = 0.2$.

Proposition 4. *There exists a principal-agent setting with private values and correlated types such that any mechanism with a fully revealing information structure is strictly suboptimal.*

$\omega \backslash \theta$	0.6	0.9
0.2	0.15	0.2
0.5	0.2	0.1
0.8	0.15	0.2

Table 1: Joint distribution over features and agent's types.

3.4 Cardinalities of Sets of Induced Interim Beliefs

Here we study the cardinalities of the sets of principal interim beliefs concerning the underlying feature that are induced by information structures used in optimal mechanisms. From Theorem 1, it follows that in IAPV environments, there is an optimal mechanism whose information structure is fully revealing and thus is such that the set of principal posterior beliefs that are induced with strictly positive probability has a weakly smaller cardinality than the set of possible features. However, we will see that there are environments outside of these classes in which every optimal mechanism must induce strictly more principal interim beliefs than the number of possible features. One takeaway from this is that, in some environments, in order to effectively balance tailoring the results of the mechanism with the underlying feature with satisfying incentive constraints for the principal, optimal mechanisms must adopt intricate information structures that garble features in a relatively sophisticated manner. In such environments, certain simple but natural information structures such as those obtained by partitioning the set of features and simply revealing which partition element contains the underlying feature cannot feature in an optimal mechanism.

There is a natural comparison with the common finding from the information design literature that, in standard setups with finite sets of possible states, there is guaranteed to be an optimal information structure in which the number of distinct posterior beliefs that are induced with strictly positive probability is at most the number of states. These findings can be seen as a consequence of Carathéodory's theorem and the fact that the main constraints in the associated design problems are "Bayes plausibility" constraints which concern the induced distribution over posteriors averaging out to the prior.

Similar considerations to those used to demonstrate these findings of the standard information design literature can be used to show that, in our setup, all quasilinear environments with transfers and just one possible agent type have optimal mechanisms in which the number of interim beliefs induced in the principal is weakly less than the number of distinct possible features.[9] Intuition for this result is as follows. In all quasilinear environments with

[9]There are environments in this class, which consequently have optimal mechanisms in which the number of interim beliefs induced in the principal is weakly less than the number of distinct possible features, such that every optimal mechanism must induce at least as many interim beliefs in the

transfers and just one possible agent type, every mechanism that respects the principal IC constraints and maximizes the total expected surplus across all feasible mechanisms can be modified into an optimal feasible mechanism by uniformly shifting the transfer rule so that the agent's IR constraint is satisfied with equality. Moreover, in the alternative problem of maximizing the ex-ante expected surplus across the set of mechanisms that satisfy the principal IC constraint, Carathéodory's theorem can be used, in a manner similar to its use in the standard information design literature, to show that there is an optimal mechanism which induces weakly fewer interim beliefs in the principal than the number of features. This is because Carathéodory's theorem implies that, for every mechanism that is incentive compatible for the principal, there is a mechanism which is incentive compatible for the principal, induces weakly fewer interim beliefs in the principal than the number of features, and results in a weakly higher ex-ante expected surplus than the original mechanism. In particular, by Carathéodory's theorem, for every mechanism that induces finitely many distinct interim beliefs in the principal, there is an alternative mechanism which (1) only induces interim beliefs that are also induced by the original mechanism, (2) induces weakly fewer interim beliefs than the number of features, (3) for each interim belief induced, has the same conditional outcome as in the original mechanism, and (4) gives a weakly higher ex-ante expected surplus than the original mechanism. Moreover, for every mechanism that is incentive compatible for the principal and induces finitely many distinct interim beliefs in the principal, the alternative mechanism justified by Carathéodory's theorem described above must be incentive compatible for the principal. A formal version of this argument, provided in Appendix A.4, gives the following.

Proposition 5. *For all quasilinear environments with transfers, if* $|\Theta| = 1$*, then there is an optimal mechanism* $(\sigma, \mathcal{M})$ *such that* $|\cup_{\omega\in\Omega} supp(\sigma(\omega))| \leq |\Omega|$.

However, in some environments outside of this class, there are effectively other constraints which preclude applications of Carathéodory's theorem that show that there is an optimal mechanism that induces weakly fewer interim beliefs in the principal than the number of features. The following example consists of a quasilinear environment with transfers in which there are 3 possible features, 2 possible agent types, and there is a mechanism in which exactly 4 signals are induced with strictly positive probability that does strictly better than every mechanism that has weakly fewer than 3 signals.[10]

Example 1. *The principal's type set* $\omega = \{\omega_1, \omega_2, \omega_3\}$ *has precisely* 3 *elements, the agent's type set* $\theta = \{\theta_1, \theta_2\}$ *has precisely* 2 *elements, and the allocation set* $Y = \{y_1, y_2, y_3, y_4, y_5\}$

principal as the number of features and be not fully revealing. OA (2.1) presents such an example.

[10] OA (2.3) presents a non-quasilinear environment with 3 possible features and just 1 agent type in which every optimal mechanism must induce at least 4 interim beliefs in the principal.

has precisely 5 *elements. The prior distribution* $\lambda \in \Delta(\Omega \times \Theta)$ *is such that the principal's type and the agent's type are statistically independent, the marginal distribution over the principal's type* $\lambda_\Omega \in \Delta(\Omega)$ *is given by* $\lambda_\Omega[\omega_1] = \lambda_\Omega[\omega_2] = \lambda_\Omega[\omega_3] = 1/3$*, and the marginal distribution over the agent's type* $\lambda_\Theta \in \Delta(\Theta)$ *is given by* $\lambda_\Theta[\theta_1] = 4/5$ *and* $\lambda_\Theta[\theta_2] = 1/5$*. The utilitiess to the principal and the agent, net of transfers, from the various allocations are given in the following table. (The table is such that, for each* $(\omega, \theta, y) \in \Omega \times \Theta \times Y$*, the first number in the corresponding pair of numbers gives the principal's utility while the second number gives the agent's utility.)*

ω_1	y_1	y_2	y_3	y_4	y_5
θ_1	$0, 1$	$-1.2, 0$	$0, 0$	$-1, 0$	$0, 0$
θ_2	$0, 10$	$-1.2, 0$	$0, 0$	$-1, -10000$	$0, 0$

ω_2	y_1	y_2	y_3	y_4	y_5
θ_1	$-1.2, 0$	$0, 1$	$-1.2, 0$	$0, 0$	$0, 0$
θ_2	$-1.2, 0$	$0, 10$	$-1.2, 0$	$0, 0$	$0, 0$

ω_3	y_1	y_2	y_3	y_4	y_5
θ_1	$-.1, 0$	$-1, -1000$	$-.1, 1$	$-.9, -1000$	$0, -1000$
θ_2	$-.1, 0$	$-1, -10000$	$-.1, 10$	$-.9, -10000$	$0, -10000$

Table 2: The utilities net of transfers for Example 1.

In this environment, note that, regardless of the agent type $\theta \in \{\theta_1, \theta_2\}$, the efficient allocation is y_1 when the underlying feature is ω_1, y_2 when the underlying feature is ω_2, and y_3 when the underlying feature is ω_3. While the principal would like to be able to implement a mechanism that results in the efficient allocation with probability 1 and fully extracts the surplus, no such mechanism is feasible due to the type θ_2 agent obtaining 10 times the value of the type θ_1 from each pair of underlying feature and associated efficient allocation. However, the principal can implement mechanisms that come close to fully extracting surplus for the principal by reducing the information rent of the type θ_2 rent by having y_4 occur with small but strictly positive probability conditional on (ω_1, θ_1) since y_4 gives the type θ_2 agent much greater disutility than the type θ_1 agent when the underlying feature is ω_1.

OA (2.2) shows that, in this environment, there is a mechanism that induces 4 interim beliefs in the principal and achieves an ex-ante expected utility for the principal that is very close to the ex-ante expected surplus that would be generated by the efficient allocation rule. The underlying information structure is such that, for each underlying feature, there

is a degenerate interim belief that puts probability 1 on the feature and occurs with high probability conditional on the associated feature. The non-degenerate interim belief puts probability $1/2$ on feature ω_1 and probability $1/2$ on feature ω_2 and occurs with relatively small probability conditional on ω_1 or ω_2. Conditional on the degenerate interim beliefs, the efficient allocations are implemented with probability 1, while, conditional on the non-degenerate interim belief, allocation y_4 is implemented with probability 1. The transfers are chosen so that the various incentive compatibility and individual rationality constraints hold and the principal fully extracts the ex-ante expected surplus generated by the associated allocation rule and information structure.

OA (2.2) further shows that this mechanism achieves a strictly higher ex-ante expected utility for the principal than every feasible mechanism that induces at most 3 interim beliefs in the principal. Intuitively, every feasible mechanism that achieves at least the ex-ante expected utility for the principal as the mechanism discussed above must have an information structure that is close to fully revealing in that, for each underlying feature $\omega \in \Omega$, the associated conditional distribution over posteriors must put very high probability on beliefs that put very high probability on ω, and there is a sufficiently high probability of y_4 conditional on (ω_1, θ_1). Thus, among the feasible mechanisms that induce at most 3 interim beliefs in the principal, only those which induce three distinct interim beliefs $\lambda_{\Omega,1}, \lambda_{\Omega,2}, \lambda_{\Omega,3} \in \Delta(\Omega)$ such that, for each $i \in \{1,2,3\}$, $\lambda_{\Omega,i}[\omega_i]$ is sufficiently high could possibly sustain as high an ex-ante expected utility to the principal as the mechanism discussed in the preceding paragraph. However, a consequence of principal incentive compatibility is that, for all such feasible mechanisms that induce precisely 3 interim beliefs in the principal and generate a sufficiently high probability of y_4 conditional on (ω_1, θ_1), the probability of $\{y_2, y_4, y_5\}$ conditional on (ω_3, θ_1) must meet a certain threshold. The extreme inefficiency y_2, y_4, and y_5 conditional on ω_3 then precludes the ex-ante expected utility of the principal from such a mechanism from being weakly greater than that mechanism that induces 4 interim beliefs in the principal discussed in the preceding paragraph.

4 Optimal Mechanisms for General Payoffs

In this section, we provide characterizations of the optimal mechanisms without the restriction to quasi-linear payoffs. In particular, we will provide sufficient conditions such that the optimal mechanism fully reveals the features to the principal, and illustrate the properties of the optimal mechanisms when fully revealing the features is strictly suboptimal. The omitted proofs for this section will be provided in Appendix B.

4.1 Binary Feature Space

In this section, we consider the case where the feature space Ω is binary and show that there exists an optimal mechanism that fully reveals the features to the principal. A canonical application of binary feature space in the field of industrial organization occurs when the feature represents the quality of a product available for sale, which can be either high or low. In this context, the manufacturer can design ex ante mechanisms for monitoring product qualities and contracting with downstream firms.

Theorem 2. *If $|\Omega| = 2$, there is an optimal mechanism in which the feature is fully revealed.*

Intuitively, for any mechanism M that is incentive compatible for the principal and potentially only reveals partial information about the features, the principal can construct another mechanism $\widetilde{M}$ that fully reveals the features and simulates the mapping from reports to distribution over outcomes based on the original mechanism M. Therefore, the constructed mechanism induces the same distribution over outcomes for all pair of feature and agent's type. Moreover, the following lemma shows that the constructed mechanism $\widetilde{M}$ is also feasible, i.e., it is incentive compatible for both the principal and the agent, and individually rational for the agent. Theorem 2 follows immediately from this observation.

Lemma 2. *If $|\Omega| = 2$, given any feasible mechanism, there is another feasible mechanism in which the feature is fully revealed and which induces the same outcome as the original mechanism.*

The main challenge for Lemma 2 is to show that the constructed mechanism is incentive compatible for the principal. For binary feature space, we assume $\Omega = \{0, 1\}$ without loss of generality. In any mechanism M, the posterior belief of the principal can be represented as a single number in $[0, 1]$, representing the posterior probability that the true feature is 1. Based on the incentive constraints in mechanism M for the principal, higher posterior belief implies that the interim utility of the principal conditional on the state being 1 is higher. Moreover, when the feature space is binary, the distribution over posterior beliefs condition on the true feature being 1 first order stochastically dominates distribution over posterior beliefs condition on the true feature being 0. Therefore, in mechanism $\widetilde{M}$, when the true feature is 1, reporting 1 leads to higher interim utility for the principal since the expected value of a monotone function increases in first order stochastic dominance. Similarly, when the true feature is 0, reporting 0 leads to higher interim utility for the principal, and hence mechanism $\widetilde{M}$ is incentive compatible for the principal.

In the above discussion, the key property in the binary feature model that facilitates the derivation of such results is the ability to attain the properties of monotone likelihood

ratios and, consequently, first-order stochastic dominance for free. In Section 4.2, we will show that when the feature space is not binary, such properties do not hold, and in general, mechanisms with fully revealing information structures can be strictly suboptimal.

4.2 Suboptimality of Fully Revealing Information Structures

In this section, we show that when the feature space is not binary, in general there are environments in which mechanisms with fully revealing information structures are strictly suboptimal. However, even though mechanisms with partially revealing information structures are optimal, we show that the principal never strictly prefers mechanisms with fully uninformative information structures.

We first focus on a independent private value environment. That is,

$$U(\omega, \theta, x) = U(\omega, \theta', x), \quad \forall \omega \in \Omega, \forall \theta, \theta' \in \Theta, \forall x \in X,$$
$$V(\omega, \theta, x) = V(\omega', \theta, x), \quad \forall \omega, \omega' \in \Omega, \forall \theta \in \Theta, \forall x \in X.$$

Therefore, we omit θ in the notation of principal's utility and ω in the notation of agent's utility in such environments.

To simplify, we further restrict the outcome space to be $X = \{0, 1\}$ and the agent's type space to be $\Theta = \{\theta\}$. In this binary action model, let the payoff differences of the principal between two allocations be $d(\omega) \triangleq U(\omega, 1) - U(\omega, 0)$. We focus on the non-degenerate utility of the agent where $V(\theta, 0) < 0$ and $V(\theta, 1) > 0$. This is because otherwise, either the individual rationality constraint of the agent can never be satisfied, or the principal can easily implement the first best solution. We show that even in such degenerate environments, the mechanisms with fully revealing information structures is strictly suboptimal for the principal if the feature space is rich.

Proposition 6. *In independent private value environments with degenerate agent type and binary allocations, any mechanism $M = (S, \sigma, \mathbf{x})$ with fully revealing information structure (S, σ) is strictly suboptimal for the principal if*

1. *$U(\omega, x) \geq 0$ for all $\omega \in \Omega, x \in X$;*
2. *$F_\Omega[\{\omega \in \Omega : d(\omega) \geq 0\}] < \frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)}$; and*
3. *there exists $\omega, \omega', \omega''$ in the support of F_Ω such that $d(\omega'') > 0 > d(\omega') > d(\omega)$.*

We first interpret the conditions in Proposition 6. Condition 1 implies that the utility of the principal for any outcome is higher than her outside option. Therefore, it is without loss of generality to focus on mechanisms that incentivize the agent to never take the

outside option. Note that requiring that agent has incentives not to take outside option in the optimal mechanism is a necessary condition for fully revealing information structure to be strictly suboptimal, because otherwise for any information structure, there exists a mechanism with such information structure that is weakly optimal (by implementing outcome 0 for any report of the agent to incentive the agent to take the outside option). Our condition 1 is one sufficient condition to ensure that this can never happen. Condition 2 implies that the mechanism that implements the first best is not individually rational for the agent. To see this, in the first best mechanism, the information structure is fully revealing, and outcome 1 is chosen for feature ω if and only if $d(\omega) \geq 0$. Therefore, the probability outcome 1 is chosen is $F_\Omega[\{\omega \in \Omega : d(\omega) \geq 0\}]$. It is easy to verify that the inequality in condition 2 is equivalent to the utility of the agent in the first best mechanism being negative. This condition in necessary to introduce the tension between maximizing the expected payoff of the principal, and the individually rational constraint for the agent.

Finally, the most substantial assumption is condition 3. It rules out the binary feature space considered in Theorem 2. The high level intuition is that, when there are multiple distinct features such that the principal strictly prefers allocation 0, it is beneficial to pool features with lesser inclinations towards choosing allocation 0 alongside those that favor allocation 1. This information structure helps increases the probability allocation 1 is chosen for features with larger $d(\omega)$ relative to features with smaller $d(\omega)$, and hence increasing the principal's expected payoff.

To prove Proposition 6, we first show that when the features are fully revealed to the principal, the mechanism has to treat all features with negative values to the principal equally in order to satisfies the incentive constraint for the principal. However, by using the information structure that pools features that have largest negative values for the principal with features that have positive values, the principal can increase the expected payoff by shifting the allocation probability for $x = 1$ from features with low values to features with high values. This is strictly beneficial for the principal under the three conditions in Proposition 6.[11]

5 Principal Interim Individual Rationality Constraints

Here we study effects of the principal interim individual rationality constraints. In particular, whereas the rest of the paper has predominantly focused on Program (OPT), in this

[11] In Appendix 3, we show that although mechanisms with fully revealing information structures is strictly suboptimal, mechanisms with fully uninformative information structures are never strictly optimal for the principal

section we focus on Program (OPT-IR). Some settings in which Program (OPT-IR) might be particularly natural include those in which the principal cannot commit themselves to forego their outside option after they have learned information about the underlying feature. Perhaps a more compelling class of settings for study of Program (OPT-IR) are those in which, rather than the principal actively participating in the mechanism essentially as an agent themself, there is a single agent that learns about the underlying feature (in a mannger completely controlled by the information structure chosen to be part of the mechanism) and whose utility function completely aligns with the principal. In this way, whereas the incentive compatibility constraints would be equivalent to those for the principal if the principal were to take the agent's role, the agent might naturally have interim outside options which the mechanism must incentivize them to forego.

The set of optimal mechanisms depended heavily on whether PIR constraints were being imposed: Without the imposition of such constraints, there were optimal mechanisms with fully revealing information structures (in fact, every information structure is used in an optimal mechanism), but, with the imposition of the PIR constraints, only fully uninformative information structures are used in optimal mechanisms. To illustrate this, consider a simple example of bilateral trade.

Similar to the example in the introduction, suppose that there are three possible qualities of the good: The good could be of low quality L, middle quality M, or high quality H with equal probabilities. As before, both the principal and agent are unaware of the good's quality, the set of possible transfers equals the set of real numbers, and both the principal's utility and the agent's utility are quasi-linear in the transfer between them. For allocation $y = 0$ that corresponds to no trade, we normalize the payoffs such that $u(\omega, 0) = u(\omega, 0)$ for all feature $\omega \in \{L, M, H\}$. Moreover, for allocation $y = 1$ that corresponds to trade, the payoffs are

$$u(\omega, 1) = \begin{cases} -3 & \omega = L; \\ -4 & \omega = M; \\ -8 & \omega = H, \end{cases} \quad \text{and} \quad v(\omega, 1) = \begin{cases} 1 & \omega = L; \\ 5 & \omega = M; \\ 7 & \omega = H. \end{cases}$$

Thus, trade is efficient for the M quality good but inefficient for both the L quality good and the H quality good. Here, with the class of incentive compatible and (agent) individually rational mechanisms that utilize fully revealing information structures, the best the principal can do is to have no trade occur with probability 1 and achieve an expected payoff of 0. However, the principal can do strictly better than this and, indeed, extract the full expected surplus possible under efficient trade with a particular mechanism whose

information structure is neither fully uninformative nor fully revealing. More specifically, an information structure in which, whenever the good has middle quality M, the principal almost surely learns that the good has quality M with probability 1, and, whenever the good has "extreme" quality L or H, the principal almost surely updates to the belief given by their prior conditional on the good not being of quality M can be combined with a trading protocol in which the principal unilaterally decides between executing full trade at a price of 5, the agent's willingness to pay for a quality M good, or no trade at a price of 0. This would be incentive compatible, due to the principal's expected payoff, ignoring transfers, from trade after learning the good has "extreme" quality being -5.5, which would lead them to reject trade at a price of 5, and individually rational, both for the agent and for the principal at the interim stage.

A key aspect of this example is the dependence of the agent's utility on the underlying feature. In fact, for all IAPV quasilinear environments with transfers, in the case where the outside option is also available for the principal as an allocation, i.e., $o \in X$, there is a mechanism with a fully revealing mechanism that is optimal regardless of whether PIR constraints are imposed.

Proposition 7. *In all IAPV quasilinear environments with transfers, if $o \in X$, there is a mechanism with a fully revealing information structure that is an optimal solution to Program (OPT-IR) as well as Program (OPT).*

The existence of such a mechanism follows from the existence of an optimal mechanism that maximizes the principal's ex-ante payoff, which automatically satisfies the principal's individual rationality constraints in IAPV quasilinear environments with transfers when $o \in X$, given any information structure (Mylovanov and Tröger, 2014). That is, the optimal solution for Program (OPT-IR) coincides with the optimal solution for Program (OPT). Combining this with our Theorem 1 directly implies Proposition 7.

While our main result for IAPV environments extends to the setup in which PIR constraints are imposed, our main result for binary-feature environments does not extend. (For details about how other results extend, see Appendix C.) In particular, there are private-values environments outside of the quasi-linear family in which $|\Omega| = 2$ and $|\Theta| = 1$ such that every optimal mechanism must have at least 3 signals induced with strictly positive probability when principal interim individual rationality constraints are imposed. The following example provides such an environments.

Example 1. The principal's type set $\Omega = \{\omega_1, \omega_2\}$ has precisely 2 elements, the agent's type set $\Theta = \{\theta\}$ has precisely 1 element, and the allocation set $X = \{x_1, x_2, x_3\}$ has precisely 3 elements. The prior distribution over the principal's type $\lambda \in \Delta(\Omega)$ is such that

$\lambda[\omega_1] = \lambda[\omega_2] = 1/2$. The payoffs to the principal and the agent from the various allocations are given in the following table. (The table is such that, for each $(\omega, \theta, x) \in \Omega \times \Theta \times X$, the first number in the corresponding pair of numbers gives the principal's payoff while the second number gives the agent's payoff.) □

ω_1	x_1	x_2	x_3
	$0, 1$	$-100, -10$	$-1, 100$

ω_2	x_1	x_2	x_3
	$0, 1$	$100, -10$	$1, 100$

Table 3: The payoffs for Example 1.

In this environment, the principal would like to have a high probability of x_2 occurring conditional on ω_2. The issue is that the agent would obtain a fairly large disutility from such a conditional outcome and would have to be compensated in order for their IR constraints to be satisfied. It turns out that the optimal mechanism involves such a high probability of x_2 occurring conditional on ω_2, and the optimal way for the agent to be compensated for this is for there to be (1) conditional on ω_1, strictly positive probabilities of x_1 and x_3 as well as a 0 probability of x_2, and (2) conditional on ω_2, a 0 probability of x_1 and a strictly positive probability of x_3. For there to be a strictly positive probability of x_3 conditional on ω_1, the expected utility of the principal conditional on ω_1 must be strictly negative, so, in every optimal mechanism, there must be an interim belief that is induced with strictly positive probability by pooling ω_1 and ω_2. However, every optimal mechanism must be such that, for each feature $\omega \in \{\omega_1, \omega_2\}$, there must be a strictly positive probability of the degenerate belief δ_ω being induced conditional on ω, owing to the need for x_1 to occur with strictly positive probability conditional on ω_1 and 0 probability conditional on ω_2 and x_2 to occur with 0 probability conditional on ω_1 and a strictly positive probability conditional on ω_2. The formal details are given in Appendix C.1.

6 Discussion

In this paper, we provide sufficient conditions for environments both with and without quasi-linear transfers, delineating scenarios where full learning proves either optimal or strictly suboptimal for the principal. This serves as an initial exploration into endogenous principal learning problems where the principal lacks the ability to commit to how acquired information is adopted in the mechanism. Our findings also open numerous avenues for future research.

First, in environments where full learning is strictly suboptimal, a natural question arises: What is the optimal information structure for the principal? Our results illustrate

that this structure can be highly intricate, and the number of signals can be strictly larger than the number of states. It is intriguing to investigate whether there exists an upper bound on the maximum number of signals for the optimal information structure and whether we can uncover any intuitive economic properties of it.

In our paper, our primary focus has been on environments characterized by pure adverse selection. However, in numerous real-world scenarios, moral hazard concerns also play a significant role. For instance, consider a situation where a firm contracts an agent whose efforts are indispensable for the success of a project. In such cases, the firm may opt to conduct private investigations into the project's profitability before finalizing the contract. This introduces another layer of complexity as the agent's efforts may not be perfectly observable by the firm. Thus, understanding the interplay between endogenous principal learning and moral hazard is crucial for comprehensively analyzing principal-agent relationships in various economic settings. This warrants further exploration which is beyond the scope of our current study.

References

Akbarpour, M. and Li, S. (2020). Credible auctions: A trilemma. *Econometrica*, 88(2):425–467.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Bei, X. and Huang, Z. (2011). Bayesian incentive compatibility via fractional assignments. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 720–733. SIAM.

Bergemann, D., Bonatti, A., and Smolin, A. (2018). The design and price of information. *American economic review*, 108(1):1–48.

Bergemann, D., Brooks, B., and Morris, S. (2015). The limits of price discrimination. *American Economic Review*, 105(3):921–957.

Bergemann, D., Heumann, T., and Morris, S. (2022). Screening with persuasion. *arXiv preprint arXiv:2212.03360*.

Bergemann, D. and Pesendorfer, M. (2007). Information structures in optimal auctions. *Journal of economic theory*, 137(1):580–609.

Bulow, J. and Roberts, J. (1989). The simple economics of optimal auctions. *Journal of Political Economy*, 97(5):1060–1090.

Clark, D. (2024a). Contracting with private information, moral hazard, and limited commitment. *Working Paper*.

Clark, D. (2024b). The informed principal with agent moral hazard. *Working Paper*.

Clark, D. and Yang, K. H. (2024). Partially informed disclosure. *Working Paper*.

Creane, A. (1998). Ignorance is bliss as trade policy. *Review of International Economics*, 6(4):616–624.

Crémer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257.

Daskalakis, C., Papadimitriou, C., and Tzamos, C. (2016). Does information revelation improve revenue? In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 233–250.

Deimen, I. and Szalay, D. o. (2019). Delegated expertise, authority, and communication. *American Economic Review*, 109(4):1349–1374.

Deng, Y., Hartline, J., Mao, J., and Sivan, B. (2021). Welfare-maximizing guaranteed dashboard mechanisms. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 370–370.

Ebrahim-Khanjari, N., Hopp, W., and Iravani, S. M. (2012). Trust and information sharing in supply chains. *Production and Operations Management*, 21(3):444–464.

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

Ferreira, M. V. and Weinberg, S. M. (2020). Credible, truthful, and two-round (optimal) auctions via cryptographic commitments. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 683–712.

Gal-Or, E. (1987). First mover disadvantages with private information. *The Review of Economic Studies*, 54(2):279–292.

Haghpanah, N. and Siegel, R. (2023). Pareto-improving segmentation of multiproduct markets. *Journal of Political Economy*, 131(6):1546–1575.

Hartline, J. D., Kleinberg, R., and Malekian, A. (2015). Bayesian incentive compatibility via matchings. *Games and Economic Behavior*, 92:401–429.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, 61(4):561–574.

Ivanov, M. (2010). Informational control and organizational design. *Journal of Economic Theory*, 145(2):721–751.

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

Kartik, N. and Zhong, W. (2023). Lemonade from lemons: Information design and adverse selection. *arXiv preprint arXiv:2305.02994*.

Kessler, A. S. (1998). The value of ignorance. *The RAND Journal of Economics*, pages 339–354.

Koessler, F. and Skreta, V. (2023). Informed information design. *Journal of Political Economy*, 131:3186–3232.

Kreutzkamp, S. and Lou, Y. (2024). Persuasion without ex-post commitment. *R&R at Journal of Economic Theory*.

Li, H. and Shi, X. (2017). Discriminatory information disclosure. *American Economic Review*, 107(11):3363–3385.

Li, Y. (2022). Selling data to an agent with endogenous information. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 664–665.

Li, Y. and Xu, B. (2024). Test design towards unanimous consent. *working paper*.

Lyu, Q. and Suen, W. (2022). Information design in cheap talk. *arXiv preprint arXiv:2207.04929*.

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

Maskin, E. and Tirole, J. (1992). The principal-agent relationship with an informed principal, ii: Common values. *Econometrica: Journal of the Econometric Society*, pages 1–42.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73.

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

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

Özer, Ö., Zheng, Y., and Chen, K.-Y. (2011). Trust in forecast information sharing. *Management Science*, 57(6):1111–1137.

Pavan, A. and Tirole, J. (2023). Knowing your lemon before you dump it. *working paper*.

Roesler, A.-K. and Szentes, B. (2017). Buyer-optimal learning and monopoly pricing. *American Economic Review*, 107(7):2072–2080.

Seifert, D. (2003). *Collaborative planning, forecasting, and replenishment: How to create a supply chain advantage*. AMACOM Div American Mgmt Assn.

Taneva, I. and Wiseman, T. (2024). Strategic ignorance and information design. Technical report, working paper.

Wei, D. and Green, B. (2024). (reverse) price discrimination with information design. *American Economic Journal: Microeconomics*, 16(2):267–295.

Yang, K. H. (2022). Selling consumer data for profit: Optimal market-segmentation design and its consequences. *American Economic Review*, 112(4):1364–1393.

A Omitted Analysis for Section 3

A.1 Theorem 1

Since the feature spaces are finite, let $\Omega = \{\omega_1, \ldots, \omega_n\}$. Let f_i be the probability of feature ω_i given marginal distribution F_Ω. Consider an imaginary market allocation problem with m resources where each resource $\mathbf{x}_j$ has supply q_j. We assume that $\sum_{j\in[m]} q_j = 1$. Consider a distribution scheme z_{ij} such that $\sum_{j\in[n]} z_{ij} = f_i$ and $\sum_{i\in[n]} z_{ij} = q_j$. Intuitively, z_{ij} is the demand from feature ω_i for resource $\mathbf{x}_j$. The principal's value for resource $\mathbf{x}_j$ given feature ω_i is v_{ij}. The principal's utility is additive across all resources and each feature ω_i must consume demand equals f_i.[12] Let

$$
\begin{aligned}
z^* = \arg\max_{z} \quad & \sum_{i\in[n],j\in[m]} v_{ij}\cdot z_{ij} \qquad (\Psi)\\
\text{s.t.} \quad & \sum_{j\in[m]} z_{ij} = f_i, \quad \forall i \in [n],\\
& \sum_{i\in[n]} z_{ij} = q_j, \quad \forall j \in [m],\\
& z_{ij} \geq 0, \quad \forall i \in [n], j \in [m].
\end{aligned}
$$

In this market, the efficient allocation assigns resource $\mathbf{x}_j$ to feature ω_i with probability $\frac{z^*_{ij}}{f_i}$.

Lemma 3. *There exists a price vector $\{p_i\}_{i\in[n]}$ such that the efficient allocation can be implemented incentive compatibly.*

Proof. Note the the optimization program (Ψ) is a linear program. The Lagrange dual of this program is

$$
\begin{aligned}
\mathcal{L} = \min_{\lambda,\beta}\max_{z} \quad & \sum_{i\in[n],j\in[m]} v_{ij}\cdot z_{ij} + \sum_{i\in[n]} \lambda_i \left(f_i - \sum_{j\in[m]} z_{ij}\right) + \sum_{j\in[m]} \beta_j \left(q_j - \sum_{i\in[n]} z_{ij}\right)\\
\text{s.t.} \quad & z_{ij} \geq 0, \quad \forall i \in [n], j \in [m].
\end{aligned}
$$

By reordering the terms, the Lagrange objective is

$$
\mathcal{L}(\lambda,\beta,z) = \sum_{i\in[n],j\in[m]} (v_{ij} - \lambda_i - \beta_j) z_{ij} + \sum_{i\in[n]} f_i\lambda_i + \sum_{j\in[m]} q_j\beta_j.
$$

[12] The restriction that each feature must consume demand exactly equals f_i is because the values v_{ij} may be negative for some i, j. This condition can be relaxed if all values are non-negative.

By the optimality condition and complementary slackness, $v_{ij} - \lambda_i - \beta_j \leq 0$ and the equality holds if $z_{ij} > 0$ for any i, j. To interpret this condition, the Lagrange parameter β_j can be viewed as the per-unit price for resource j and λ_i is the maximum per-unit utility of feature ω_i from purchasing the resource. That is, feature ω_i's expected utility is maximized by z^* given price β_j. Consider the price $p_i = \frac{1}{f_i} \sum_{j \in [m]} z^*_{ij} \beta_{ij}$. Each feature ω_i's utility for deviating to ω'_i is

$$\begin{aligned}
&\sum_{j \in [m]} v_{ij} \cdot \frac{z^*_{i'j}}{f_i} - \frac{1}{f_i} \sum_{j \in [m]} z^*_{i'j} \beta_{i'j} \\
&\leq \frac{1}{f_i} \sum_{j \in [m]} z^*_{ij} \lambda_i = \sum_{j \in [m]} v_{ij} \cdot \frac{z^*_{ij}}{f_i} - \frac{1}{f_i} \sum_{j \in [m]} z^*_{ij} \beta_{ij}
\end{aligned}$$

Therefore, given price vector $\{p_i\}$, each feature will have incentive to report truthfully. ■

Now we revisit our original problem. For any information structure σ and any mechanism M, fixing any agent's type θ, let $\mathbf{x}^\theta_j$ be the distribution over outcome in mechanism M given feature ω_j and type θ. $\mathbf{x}^\theta_j$ can be viewed as resource j in the imaginary problem where the supply of resource j is f_j. Consider the fully revealing information structure and another mechanism $\widehat{M}$ that redistributes the resource efficiently for the principal given each agent's type. By Lemma 3, there exists a transfer profile $\{p^\theta_i\}_{i \in [n]}$ such that the efficient allocation can be implemented incentive compatibly for all principal's types, which equals the underlying features given fully revealing information structures. Moreover, by shifting the transfers by a constant (depending on the agent's type θ) for all principal's types, we can also ensure that the expected transfer of the agent are the same in both mechanisms. Since mechanism $\widehat{M}$ implements the same distribution over outcome for each agent's type, the incentive constraints and the individually ration constraints are satisfied for the agent. Finally, since the equilibrium welfare of mechanism $\widehat{M}$ is weakly higher than that for mechanism M, and the agent's expected utility remains unchanged, the principal's expected utility weakly increases. Therefore, revealing full information is optimal for the principal.

A.2 Agent Interdependent Values

To prove Proposition 2, we introduce the following notations. For any quantile $q \in [0, 1]$, let $\omega(q) \triangleq \inf_{\omega'} \{F_\Omega(\omega') \geq 1 - q\}$ be the feature that corresponds to quantile q. Let $H(q) = \int_0^q (c(\omega(z)) - \omega(z))\, \mathrm{d}z$, and let $\underline{H}(q)$ be the convex hull of H. Let $\mathcal{I}$ be the set of intervals such that $\underline{H}(q) \neq H(q)$. Note that $\mathcal{I} = \emptyset$ if $c(\omega) - \omega$ is non-increasing in ω for all ω in the support of F_Ω and $\mathcal{I} = \{\Omega\}$ if $c(\omega) - \omega$ is increasing in ω for all ω in the support of F_Ω. Let

$m = \mathbf{E}_{F_\Omega}[\omega]$ and $m_c = \mathbf{E}_{F_\Omega}[c(\omega) - \omega]$. The proof of Proposition 2 relies on the following lemma.

Lemma 4. *In the lemon's problem, fixing the utility function of both the principal and the agent and the distribution over unknown features, there exists a mechanism with fully revealing information structure that is optimal for each agent's type distribution if and only if there exists quantile q as well as a measure τ such that*

- $\tau(z) \leq F_\Omega(z)$ *for any* $z \subseteq \Omega$ *and* $\tau(\Omega) = q$*;*
- $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq q \cdot m$*;*
- $\int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) > \underline{H}(1) - \underline{H}(q)$.

Essentially, τ is a sub measure of F_Ω with total probability q. This lemma shows that by pooling features given measure τ and selling the item to the agent given those features, there exists a type distribution of the agent, which as illustrated in the following proof is a point mass distribution on value $\underline{H}'(q)$, such that the principal receives strictly higher expected payoff compared to any mechanism with fully revealing information structures.

Proof of Lemma 4. We first characterize the ex ante optimal mechanism of the principal given any fixed information structure. Note that in this setting, it is without loss to focus on the case that the principal's type is her posterior mean of the features under the given information structure, and hence to simplify the exposition in this characterization, we also use ω to denote the principal's type. By Envelope Theorem, the agent's interim utility in an incentive compatible mechanism is

$$V(\theta) = V(0) + \int_0^\theta \mathbf{y}(z) \, \mathrm{d}z$$

and hence by setting $V(0) = 0$. the expected payoff of the principal is

$$\begin{aligned} &\mathbf{E}_{(\omega,\theta)\sim F}\left[\omega \cdot (1 - \mathbf{y}(\omega,\theta)) + (c(\omega) + \theta) \cdot \mathbf{y}(\omega,\theta) - \frac{1 - F_\Theta(\theta)}{f_\Theta(\theta)} \cdot \mathbf{y}(\omega,\theta)\right] \\ &= \mathbf{E}_{\omega\sim F_\Omega}[\omega] + \mathbf{E}_{(\omega,\theta)\sim F}[(\varphi(\omega) + \phi(\theta)) \cdot \mathbf{y}(\omega,\theta)] \end{aligned}$$

where $\varphi(\omega) = c(\omega) - \omega$ and $\phi(\theta) = \theta - \frac{1-F_\Theta(\theta)}{f_\Theta(\theta)}$.[13]

[13] When the agent's value distribution is discrete, the density function does not exist and the above virtual value function is not well defined. However, one can define the virtual value function as the derivative on the revenue curve (Bulow and Roberts, 1989), and the same characterization extends.

If $\varphi(\omega)$ is monotone non-increasing in ω and $\phi(\theta)$ is monotone non-decreasing in θ, the optimal mechanism can be implemented by allocation rule where $\mathbf{y}(\omega, \theta) = 1$ if and only if $\varphi(\omega) + \phi(\theta) \geq 0$. Such allocation rule can be implemented in an incentive compatible way because it is weakly decreasing in ω and weakly increasing in θ.

However, if those monotonicity conditions are violated, ironing is necessary to characterize the optimal mechanism (Myerson, 1981; Bulow and Roberts, 1989). Specifically, recall $\underline{H}(q)$ is the convex hull of the integration of the surplus function $\varphi(\omega)$ in quantile space. Let $R(q) = q \cdot \theta(q)$ and let $\bar{R}(q)$ be its concave hull. The ironed surplus of the principal is $\underline{\varphi}(\omega) = \underline{H}'(q(\omega))$ and the ironed virtual value of the agent is $\bar{\phi}(\theta) = \bar{R}'(q(\theta))$. The expected payoff of the principal of any incentive compatible mechanism is upper bounded by

$$\mathbf{E}_{\omega \sim F_\Omega}[\omega] + \mathbf{E}_{(\omega,\theta)\sim F}\left[(\underline{\varphi}(\omega) + \bar{\phi}(\theta)) \cdot \mathbf{y}(\omega, \theta)\right]$$

and the upper bound is attained under the optimal mechanism. Next we prove the lemma.

If: Consider the case where the agent's distribution is a point mass at type $\theta = \underline{H}'(q)$. In this case, under full information, the optimal allocation is $\mathbf{y}(\omega, \theta) = 1$ if and only if $F_\Omega(\omega) \geq 1 - q$. The principal's ex ante payoff under this mechanism is

$$\begin{aligned}
&\mathbf{E}_{\omega \sim F_\Omega}[\omega] + \mathbf{E}_{(\omega,\theta)\sim F}\left[(\underline{\varphi}(\omega) + \theta) \cdot \mathbb{1}\left[F_\Omega(\omega) \geq 1 - q\right]\right] \\
&= \mathbf{E}_{\omega \sim F_\Omega}[\omega] + \underline{H}(1) - \underline{H}(q) + \theta \cdot q.
\end{aligned}$$

However, consider a partially informative information structure with binary signal $\{0, 1\}$ such that the principal receives signal 0 given the sub-measure τ and receives signal 1 otherwise. Moreover, consider an allocation rule that allocates the item if and only if the principal's signal is 0. Note that $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq q \cdot m$ implies that the posterior mean given signal 0 is at most m, and hence the posterior mean given signal 1 is at least m. Therefore, such allocation is non-increasing in principal's posterior mean, and hence this allocation rule can be implemented as an incentive compatible mechanism. Moreover, the ex ante payoff of the principal under this mechanism when receiving partial information is

$$\mathbf{E}_{\omega \sim F_\Omega}[\omega] + \int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) + \theta \cdot q,$$

which is strictly higher than the ex ante payoff under full information according to the assumption in Lemma 4.

Only if: We prove this by contradiction. For any information structure (S, σ), let $\underline{H}^\sigma(q)$ be the convex hull of the principal's integration of surplus under information structure σ. It is easy to verify that

$$\underline{H}^\sigma(1) = \underline{H}(1) = m_c \quad \text{and} \quad \underline{H}^\sigma(q) \geq \underline{H}(q), \forall q.$$

Consider any mechanism $M = (S, \sigma, \mathbf{y}, t)$, let q_θ^M be the probability the item is allocated to the agent with type θ in mechanism M. The ex ante payoff of the principal under this mechanism is

$$\begin{aligned}&\mathbf{E}_{\omega\sim F_\Omega}[\omega] + \mathbf{E}_{(\omega,\theta)\sim F}\left[(\underline{\varphi}(\omega) + \bar{\phi}(\theta)) \cdot \mathbf{y}(\omega, \theta)\right] \\ &= \mathbf{E}_{\omega\sim F_\Omega}[\omega] + \mathbf{E}_{\theta\sim F_\Theta}\left[\bar{\phi}(\theta) \cdot q_\theta^M + \underline{H}^\sigma(1) - \underline{H}^\sigma(q_\theta^M)\right]\end{aligned}$$

Now consider the mechanism $\widehat{M}$ with fully revealing information structure which only allocates the item if and only if $F_\Omega(\omega) \leq 1 - q_\theta^M$. Since this allocation is non-decreasing in principal's type, it also be implemented as an incentive compatible mechanism under full information structure. The ex ante payoff of the principal is

$$\mathbf{E}_{\omega\sim F_\Omega}[\omega] + \mathbf{E}_{\theta\sim F_\Theta}\left[\bar{\phi}(\theta) \cdot q_\theta^M + \underline{H}(1) - \underline{H}(q_\theta^M)\right]$$

which is weakly higher than the ex ante payoff under mechanism M since $\underline{H}^\sigma(q) \geq \underline{H}(q)$ for all q. ■

Proof of Proposition 2. For the "if" direction, if condition (1) is satisfied, for any quantile q and any measure τ with $\tau(\Omega) = q$ such that $\tau(z) \leq F_\Omega(z)$ for any $z \subseteq \Omega$ and $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq q \cdot m$, the expected surplus $\int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega)$ is maximized by greedily assign probabilities on low types. The maximum surplus is therefore $\underline{H}(1) - \underline{H}(q)$. By Lemma 4, full information is optimal. If condition (2) is satisfied, $\underline{H}$ is linear function, and by combining it with the linearity of $c(\omega) - \omega$, we have that any τ with $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq q \cdot m$ satisfies $\int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) \leq \underline{H}(1) - \underline{H}(q)$. Again by Lemma 4, full information is optimal.

For the "only if" direction, suppose both conditions are violated. We first consider the case $\mathcal{I} \neq \emptyset$ and $\mathcal{I} \neq \{\Omega\}$. In this case, there exists $I \in \mathcal{I}$ such that $I \neq \Omega$. Let $\underline{\omega}$ be the lowest type in I and let $\bar{\omega}$ be the highest type in I. Let z_I be the average of $c(\omega) - \omega$ for $\omega \in I$ given the prior distribution. Since $I \neq \Omega$, there exists a type $\omega \in \Omega \backslash I$. We first consider the case $\omega < \underline{\omega}$ and the case where $\omega > \bar{\omega}$ can be proved analogously. In this case, there exists $\hat{\omega} \in \Omega$ such that $c(\hat{\omega}) - \hat{\omega} > z_I$. Letting $\Omega_0 = \{\omega : \omega < \underline{\omega}\}$, consider a measure

τ such that for any $z \subseteq \Omega$,

$$\tau(z) = \begin{cases} F_\Omega(z \cap \Omega_0) & \hat{\omega} \notin z \\ F_\Omega(z \cap \Omega_0) + \epsilon & \hat{\omega} \in z. \end{cases}$$

Let $q = \tau(\Omega)$. It is easy to verify that with sufficiently small $\epsilon > 0$, $\tau(z) \leq F_\Omega(z)$ for any $z \subseteq \Omega$ and $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq q \cdot m$. Moreover,

$$\begin{aligned} &\int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) = \int_{\Omega_0} (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) + \epsilon \cdot (c(\hat{\omega}) - \hat{\omega}) \\ &> \int_{\Omega_0} (c(\omega) - \omega) \, \mathrm{d}F_\Omega(\omega) + \epsilon \cdot z_I = \underline{H}(1) - \underline{H}(q). \end{aligned}$$

Therefore, measure τ satisfies the conditions in Lemma 4 and hence full information is strictly suboptimal.

Finally, suppose $\mathcal{I} = \Omega$. Since $c(\omega) - \omega$ is not linear in ω, it is easy to show that there exists a probability measure τ on Ω such that $\int_\Omega \omega \, \mathrm{d}\tau(\omega) \leq m$ and $\int_\Omega (c(\omega) - \omega) \, \mathrm{d}\tau(\omega) > m_c$. Note that $\mathcal{I} = \Omega$ implies that $\underline{H}$ is a linear function and $\underline{H}(1) - \underline{H}(0) = m_c$. Therefore, there exists a sufficiently small probability $q > 0$ such that the probability measure $q \cdot \tau$ satisfies the conditions in Lemma 4, and hence full information is strictly suboptimal. ■

A.3 Correlated Types

The proof of Lemma 1 relies on the following assortative inequality.

Lemma 5. *For any distribution F supported on $[0, 1]$ and any increasing sequence r_s for $s \in [0, 1]$, we have*

$$\frac{\mathbf{E}_{s \sim F}[s \cdot r_s]}{\mathbf{E}_{s \sim F}[s]} \geq \frac{\mathbf{E}_{s \sim F}[(1 - s) \cdot r_s]}{\mathbf{E}_{s \sim F}[1 - s]}.$$

Proof. We define two new probability measures F_+, F_- on $[0, 1]$ such that

$$F_+(s) = \frac{\int_0^s z \, \mathrm{d}F(z)}{\mathbf{E}_{z \sim F}[z]} \quad \text{and} \quad F_-(s) = \frac{\int_0^s (1 - z) \, \mathrm{d}F(z)}{\mathbf{E}_{z \sim F}[1 - z]}, \quad \forall s \in [0, 1].$$

The inequality in Lemma 5 is equivalent to the statement that the expectation of r_s is weakly larger given measure F_+ compared to F_-. Note that it is sufficient to show that F_+ first order stochastically dominates F_-, i.e., $F_+(s) \leq F_-(s)$ for all $s \in [0, 1]$.

Note that for any $s \in [0,1]$, we have

$$F_-(s) = \frac{\int_0^s (1-z)\,\mathrm{d}F(z)}{\mathbf{E}_{z\sim F}[1-z]} = \frac{F(s) - \int_0^s z\,\mathrm{d}F(z)}{1-\mathbf{E}_{z\sim F}[z]}.$$

By rearranging the terms, it is easy to verify that $F_+(s) \le F_-(s)$ is equivalent to

$$\int_0^s z\,\mathrm{d}F(z) \le F(s)\cdot \mathbf{E}_{z\sim F}[z]\,.$$

The above inequality holds since $\int_0^s z\,\mathrm{d}F(z) = F(s)\cdot \mathbf{E}_{z\sim F}[z \,|\, z \le s]$ and the fact that $\mathbf{E}_{z\sim F}[z \,|\, z \le s] \le \mathbf{E}_{z\sim F}[z]$ for any $s \in [0,1]$. ■

Proof of Lemma 1. Since the unknown features are payoff irrelevant for the principal, and we have assumed private values in this environment, we denote $u(y)$ as the principal's value given any allocation $y \in Y$ and $u(\mathbf{y})$ as the expected value given a distribution over allocations.

When agent has binary type $\{\theta_0, \theta_1\}$, Let q_1 be the marginal probability of type θ_1 and let $q_0 = 1 - q_1$. For any mechanism $M = (S, \sigma, \mathbf{y}, t)$, the signal, or equivalently the principal's type, can be denoted as a real number in $[0,1]$, representing the probability the agent's type is θ_1 conditional on the principal's type. Since mechanism M is incentive compatible for the principal and the principal's payoff does not depend on the allocation, it is easy to verify that $u(\mathbf{y}(s,\theta_0)) + t(s,\theta_0)$ is weakly decreasing in s and $u(\mathbf{y}(s,\theta_1)) + t(s,\theta_1)$ is weakly increasing in s.

Now consider another mechanism $\widehat{M} = (S, \sigma, \mathbf{y}, \hat{t})$ such that for any $s \in S$,

$$\hat{t}(s,\theta_0) = \hat{t}_0 - u(\mathbf{y}(s,\theta_0)), \quad \text{and} \quad \hat{t}(s,\theta_1) = \hat{t}_1 - u(\mathbf{y}(s,\theta_1))$$

where

$$\hat{t}_0 \triangleq \frac{1}{q_0}\cdot \mathbf{E}_\sigma[(1-s)\cdot(u(\mathbf{y}(s,\theta_0)) + t(s,\theta_0)) \,|\, \theta_0]\,,$$
$$\hat{t}_1 \triangleq \frac{1}{q_1}\cdot \mathbf{E}_\sigma[s\cdot(u(\mathbf{y}(s,\theta_1)) + t(s,\theta_1)) \,|\, \theta_1]\,.$$

Intuitively, $u(\mathbf{y}(s,\theta)) + t(s,\theta)$ is the utility of the principal after accounting for the cost of allocation. Compared to the original mechanism M, mechanism $\widehat{M}$ adjusts the transfers such that the realized utility of the agent is a constant regardless of the reported signal s and the expected transfer of each agent type is not affected. Therefore, the incentives of the principal is completely eliminated in mechanism $\widehat{M}$. Moreover, it is easy to verify that mechanism $\widehat{M}$ is individual rational for the agent and the expected revenue for the principal

remains unchanged. Next it is sufficient to verify that $\widehat{M}$ is incentive compatible for the agent.

Let $V(\theta, \theta'; M)$ be the expected utility of agent type θ for reporting θ' under mechanism M. The expected difference in utility loss for misreporting the type from θ_1 to θ_0 is

$$\begin{aligned} &V(\theta_1, \theta_1; M) - V(\theta_1, \theta_0; M) - (V(\theta_1, \theta_1; \widehat{M}) - V(\theta_1, \theta_0; \widehat{M})) \\ &= -\frac{1}{q_1} \cdot \mathbf{E}_\sigma[s \cdot (u(\mathbf{y}(s, \theta_0)) + t(s, \theta_0)) \,|\, \theta_1] + \hat{t}_0 \leq 0, \end{aligned}$$

where the inequality holds by applying Lemma 5 and the fact that $u(\mathbf{y}(s, \theta_0)) + t(s, \theta_0)$ is weakly decreasing in s. Therefore, the IC constraint for misreporting from θ_1 to θ_0 in mechanism M implies the IC constraint for misreporting from θ_1 to θ_0 in mechanism $\widehat{M}$. Similarly, the IC constraint for misreporting from θ_1 to θ_0 in mechanism M implies the IC constraint for misreporting from θ_1 to θ_0 in mechanism $\widehat{M}$. Combining the two observations above, mechanism $\widehat{M}$ is also incentive compatible for the agent. ■

Proof of Proposition 3. By Lemma 1, there exists an optimal mechanism $M = (S, \sigma, \mathbf{y}, t)$ such that the utility of the principal does not depend on their type. In this case, there exists another mechanism $\widehat{M} = (\hat{S}, \hat{\sigma}, \hat{\mathbf{y}}, \hat{t})$ with fully revealing information structure $(\hat{S}, \hat{\sigma})$ and allocation and transfer rule $(\hat{\mathbf{y}}, \hat{t})$ that first garbles the reported signal $\hat{s} \in \hat{S}$ according to σ, and then apply $\mathbf{y}, t$ on the garbled information. The constructed mechanism $\widehat{M}$ is incentive compatible and individually rational for the agent, and generate the same expected revenue as M since the distribution over outcomes is not changed in both settings. Moreover, since the utility of the principal does not depend on their report, the principal's incentives are also preserved. ■

Proof of Corollary 1. Suppose by contradiction there exists an optimal mechanism that extracts full surplus. By Lemma 1, there exists an optimal mechanism with transfers independent from the principal's report that extracts full surplus. In particular, in order to extract full surplus, the allocation is 1 regardless of the report from both the principal and the agent, and thus the transfer of the agent always equals his value for the item. However, in this case, the mechanism is not incentive compatible for the agent since type θ_1 would like to deviate the report to θ_0, a contradiction. ■

Proof of Proposition 4. Consider the principal-agent instance with the joint distribution illustrated in Table 1. The payoff maximizing mechanism with fully revealing information structure can be solved by a simple linear program. In particular, the optimal mechanism sells the item to the agent when his type is $\theta = 0.9$, or when $\theta = 0.6$ and $\omega = 0.2$. Moreover,

the transfers in the optimal mechanism takes the form illustrated in Table 4. It is easy to verify that the expected payoff of the principal is 0.204 under full information.

$\omega\backslash\theta$	0.6	0.9
0.2	0.656	0.572
0.5	-0.024	0.932
0.8	-0.024	0.932

Table 4: Transfer function.

Note that in fact in this mechanism, both features 0.5 and 0.8 are treated equally. A feasible choice of the principal is to pool both features 0.5 and 0.8 to alleviate the incentive constraints of the principal. The allocation and transfer functions of an feasible mechanism for pooling 0.5 and 0.8 is illustrated in Table 5. The expected revenue of this mechanism is $0.24 > 0.204$. ■

$\omega\backslash\theta$	0.6	0.9
0.2	1	1
0.5 or 0.8	0	$\frac{2}{3}$

(a) Allocation function.

$\omega\backslash\theta$	0.6	0.9
0.2	1.44	0.04
0.5 or 0.8	-0.36	$\frac{88}{75}$

(b) Transfer function.

Table 5: Allocation and transfer function under pooling information.

A.4 Proof of Proposition 5

Lemma 6. *If $|\Theta| = 1$, then, for every incentive compatible and individually rational mechanism $(S, \sigma, \mathcal{M})$ that satisfies $|S| \in \mathbb{N}$, there is an incentive compatible and individually rational mechanism $(\widetilde{S}, \tilde{s}, \widetilde{\mathcal{M}})$ that satisfies $|\widetilde{S}| \leq |\Omega|$ and gives the principal a weakly higher expected payoff than $(S, \sigma, \mathcal{M})$.*

Proof. Let $\lambda_\Omega \in \Delta(\Omega)$ denote the prior probability distribution over the principal's type and θ denote the sole element of Θ. Consider an arbitrary incentive compatible and individually rational mechanism $(S, \sigma, \mathcal{M})$ that satisfies $|S| \in \mathbb{N}$. For every $s \in S$, let $p[s] = \sum_{\omega\in\Omega} \lambda_\Omega[\omega]\sigma(\omega)[s]$ denote the ex-ante probability that signal s occurs under signal structure (S, σ), let $\lambda_{(S,\sigma)}(s) \in \Delta(\Omega)$ denote the posterior belief of the principal upon observing signal s with information structure (S, σ), and let $U_{(S,\sigma,\mathcal{M})}(s) = \mathbb{E}_{\mathcal{M}(s,\theta)}[\mathbb{E}_{\omega\sim\lambda_{(S,\sigma)}(s)}[u(\omega, \theta, x)] + t]$ and $V_{(S,\sigma,\mathcal{M})}(s) = \mathbb{E}_{\mathcal{M}(s,\theta)}[\mathbb{E}_{\omega\sim\lambda_{(S,\sigma)}(s)}[v(\omega, \theta, x)] - t]$ denote the respective expected utilities of the principal and the agent under the mechanism $(S, \sigma, \mathcal{M})$ conditional upon signal s. Observe that (1) $\sum_{s\in S} p[s]\lambda_{(S,\sigma)}(s) = \lambda_\Omega$, (2)

The principal's expected utility from the mechanism $(S, \sigma, \mathcal{M})$, denoted by $U(S, \sigma, \mathcal{M})$, must satisfy $U(S, \sigma, \mathcal{M}) = \sum_{s \in S} p[s] U_{(S,\sigma,\mathcal{M})}(s)$, and (3) The agent's expected utility from the mechanism $(S, \sigma, \mathcal{M})$, denoted by $V(S, \sigma, \mathcal{M})$, must satisfy $V(S, \sigma, \mathcal{M}) = \sum_{s \in S} p[s] V_{(S,\sigma,\mathcal{M})}(s)$. By standard arguments involving Caratheodory's Theorem, there exists an $N \in \mathbb{N}$, $f : \{1, ..., N\} \to [0, 1]$, $g : \{1, ..., N\} \times \{1, ..., |\Omega|\} \to (0, 1]$, and $h : \{1, ..., N\} \times \{1, ..., |\Omega|\} \to S$ such that (1) $\sum_{n \in \{1,...,N\}} f(n) = 1$, (2) For all $n \in \{1, ..., N\}$, $\sum_{m \in \{1,...,|\Omega|\}} g(n, m) = 1$ and $\sum_{m \in \{1,...,|\Omega|\}} g(n, m) \lambda_{(S,\sigma)}(h(n, m)) = \lambda_\Omega$, and (3) For all $s \in S$, $\sum_{n \in \{1,...,N\}} f(n) \sum_{m \in \{1,...,|\Omega|\}} g(n, m) \mathbb{1}_s(h(n, m)) = p[s]$. Thus, it follows that

$$
\begin{aligned}
&U(S, \sigma, \mathcal{M}) + V(S, \sigma, \mathcal{M}) = \sum_{s \in S} p[s] U_{(S,\sigma,\mathcal{M})}(s) + \sum_{s \in S} p[s] V_{(S,\sigma,\mathcal{M})}(s) \\
&= \sum_{s \in S} p[s] (U_{(S,\sigma,\mathcal{M})}(s) + V_{(S,\sigma,\mathcal{M})}(s)) \\
&= \sum_{s \in S} \sum_{n \in \{1,...,N\}} f(n) \sum_{m \in \{1,...,|\Omega|\}} g(n, m) \mathbb{1}_s(h(n, m)) (U_{(S,\sigma,\mathcal{M})}(s) + V_{(S,\sigma,\mathcal{M})}(s)) \\
&= \sum_{n \in \{1,...,N\}} f(n) \sum_{m \in \{1,...,|\Omega|\}} g(n, m) (U_{(S,\sigma,\mathcal{M})}(h(n, m)) + V_{(S,\sigma,\mathcal{M})}(h(n, m))).
\end{aligned}
$$

Hence, there must be some $n \in \{1, ..., N\}$ such that $\sum_{m \in \{1,...,|\Omega|\}} g(n, m) (U_{(S,\sigma,\mathcal{M})}(h(n, m)) + V_{(S,\sigma,\mathcal{M})}(h(n, m))) \geq U(S, \sigma, \mathcal{M}) + V(S, \sigma, \mathcal{M})$. Fix such an $n \in \{1, ..., N\}$. For each $(s, \theta) \in S \times \Theta$, let $\widehat{\mathcal{M}}(s, \theta) \in \Delta(X \times \mathbb{R})$ denote the probability distribution that is obtained from the probability distribution $\mathcal{M}(s, \theta) \in \Delta(X \times \mathbb{R})$ by shifting t to $t + \sum_{m \in \{1,...,|\Omega|\}} g(n, m) V_{(S,\sigma,\mathcal{M})}(h(n, m)) - V(S, \sigma, \mathcal{M})$ for each $(x, t) \in X \times \mathbb{R}$. Now, consider the mechanism $(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$ in which (1) $\widetilde{S} = \{1, ..., |\Omega|\}$, (2) $\tilde{s} : \Omega \to \widetilde{S}$ is given by $\tilde{\sigma}(\omega)[s] = g(n, s) h(n, s)[\omega] / (\sum_{s \in \widetilde{S}} g(n, s) h(n, s)[\omega])$ for all $s \in \widetilde{S}$, and (3) $\widetilde{\mathcal{M}} : \widetilde{S} \times \Theta \to \Delta(X \times \mathbb{R})$ is given by $\widetilde{\mathcal{M}}(s, \theta) = \widehat{\mathcal{M}}(h(n, s), \theta)$ for all $(s, \theta) \in \widetilde{S} \times \Theta$. Observe that, for each $s \in \widetilde{S}$, the ex-ante probability that signal s occurs under signal structure $(\widetilde{S}, \tilde{\sigma})$ is $g(n, s)$ and the posterior belief of the principal upon observing signal s with information structure $(\widetilde{S}, \tilde{\sigma})$ is $h(n, s)$. Furthermore, the principal's expected utility under $(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$, denoted by $U(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$, satisfies $U(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}}) = \sum_{s \in \{1,...,|\Omega|\}} g(n, s) U_{(S,\sigma,\mathcal{M})}(h(n, s)) + \sum_{s \in \{1,...,|\Omega|\}} g(n, s) V_{(S,\sigma,\mathcal{M})}(h(n, s)) - V(S, \sigma, \mathcal{M}) \geq U(S, \sigma, \mathcal{M})$. Additionally, because of the incentive compatibility of $(S, \sigma, \mathcal{M})$, it follows that $(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$ is incentive compatible for the principal. Moreover, as the expected payoff of the agent under $(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$ equals their expected utility under $(S, \sigma, \mathcal{M})$ and $(S, \sigma, \mathcal{M})$ is individually rational for the agent, $(\widetilde{S}, \tilde{\sigma}, \widetilde{\mathcal{M}})$ must be individually rational for the agent. ■

Proof of Proposition 5. Throughout this argument, for every mechanism $(S, \sigma, \mathcal{M})$, let $U(S, \sigma, \mathcal{M})$ denote the expected payoff of the principal under mechanism $(S, \sigma, \mathcal{M})$ and

$V(S, \sigma, \mathcal{M})$ denote the expected payoff of the agent under mechanism $(S, \sigma, \mathcal{M})$.

Let $(S^*, \sigma^*, \mathcal{M}^*)$ be an optimal incentive compatible and individually rational mechanism and let θ denote the sole element of Θ. Standard arguments show that, for all $n \in \mathbb{N}$, there is an incentive compatible mechanism $(S_n, \sigma_n, \mathcal{M}_n)$ such that $|S_n| \in \mathbb{N}$ and $U(S_n, \sigma_n, \mathcal{M}_n) + V(S_n, \sigma_n, \mathcal{M}_n) > U(S^*, \sigma^*, \mathcal{M}^*) + V(S^*, \sigma^*, \mathcal{M}^*) - 1/n$. This mechanism can be modified by uniformly shifting transfers to obtain a mechanism $(\widehat{S}_n, \hat{\sigma}_n, \widehat{\mathcal{M}}_n)$ that is incentive compatible, individually rational, and satisfies $U(\widehat{S}_n, \hat{\sigma}_n, \widehat{\mathcal{M}}_n) > U(S^*, \sigma^*, \mathcal{M}^*) - 1/n$. By Lemma 6, for all $n \in \mathbb{N}$, it must be that there exists an incentive compatible and individually rational mechanism $(\widetilde{S}_n, \tilde{\sigma}_n, \widetilde{\mathcal{M}}_n)$ such that $|\widetilde{S}_n| \leq |\Omega|$ and $U(\widetilde{S}_n, \tilde{\sigma}_n, \widetilde{\mathcal{M}}_n) \geq U(S^*, \sigma^*, \mathcal{M}^*) - 1/n$. Without loss of generality, suppose that, for all $n \in \mathbb{N}$, $|\widetilde{S}_n| = \{1, ..., |\widetilde{S}_n|\}$. Furthermore, since an appropriate subsequence could be identified, it is without loss of generality to assume that there is some $M \in \mathbb{N}$ such that (1) $M \leq |\Omega|$ and $\widetilde{S}_n = \{1, ..., M\}$ for all $n \in \mathbb{N}$, (2) $\lim_{n\to\infty} \tilde{s}_n(\omega) \in \Delta(\{1, ..., M\})$ exists for all $\omega \in \Omega$, and (3) $\lim_{n\to\infty} \mathcal{M}_n(s, \theta) \in \Delta(X \times \mathbb{R})$ exists for all $(s, \theta) \in \{1, ..., M\} \times \Theta$. Consider the mechanism $(\widetilde{S}^*, \tilde{\sigma}^*, \widetilde{\mathcal{M}}^*)$ given by $\widetilde{S}^* = \{1, ..., M\}$, $\tilde{\sigma}^*(\omega) = \lim_{n\to\infty} \tilde{\sigma}_n(\omega)$ for all $\omega \in \Omega$, and $\widetilde{\mathcal{M}}^*(s, \theta) = \lim_{n\to\infty} \widetilde{\mathcal{M}}_n(s, \theta)$ for all $(s, \theta) \in S \times \Theta$. By continuity, since, for each $n \in \mathbb{N}$, $(\widetilde{S}_n, \tilde{\sigma}_n, \widetilde{\mathcal{M}}_n)$ is incentive compatible and individually rational, $(\widetilde{S}^*, \tilde{\sigma}^*, \widetilde{\mathcal{M}}^*)$ itself must be incentive compatible and individually rational. Moreover, by continuity, since $U(S^*, \sigma^*, \mathcal{M}^*) - 1/n < U(\widetilde{S}_n, \tilde{\sigma}_n, \widetilde{\mathcal{M}}_n) \leq U(S^*, \sigma^*, \mathcal{M}^*)$ for all $n \in \mathbb{N}$, it follows that $U(\widetilde{S}^*, \tilde{\sigma}^*, \widetilde{\mathcal{M}}^*) = \lim_{n\to\infty} U(\widetilde{S}_n, \tilde{\sigma}_n, \widetilde{\mathcal{M}}_n) = U(S^*, \sigma^*, \mathcal{M}^*)$, so $(\widetilde{S}^*, \tilde{\sigma}^*, \widetilde{\mathcal{M}}^*)$ must be optimal. Finally, observe that, by construction, $|\widetilde{S}^*| \leq |\Omega|$. ■

B Omitted Analysis for Section 4

B.1 Binary Features

Proof of Lemma 2. For any feasible mechanism $M = (S, \sigma, \mathbf{x})$, consider another mechanism $\widetilde{M} = (\Omega, \tilde{\sigma}, \tilde{\mathbf{x}})$ with fully revealing information structure in which $\tilde{\sigma}(\omega) = \delta_\omega$ for all $\omega \in \Omega$. Moreover, the allocation rule $\tilde{\mathbf{x}} : \Omega \times \Theta \to \Delta(X)$ is given by

$$\tilde{\mathbf{x}}(\omega, \theta) = \mathbf{E}_{s\sim\sigma(\omega)}[\mathbf{x}(s, \theta)] .$$

for all $(\omega, \theta) \in \Omega \times \Theta$. By the construction of mechanism $\widetilde{M}$, both mechanism M and $\widetilde{M}$ induce the same distribution over outcomes for all $(\omega, \theta) \in \Omega \times \Theta$.

For every $\omega \in \Omega$, let $F_\Theta(\omega) \in \Delta(\Theta)$ denote the conditional distribution over the agent'stype when the principal'stype is ω. Likewise, for every $\theta \in \Theta$, let $F_\Omega(\theta) \in \Delta(\Omega)$ denote the conditional distribution over the principal'stype when the agent'stype is θ.

We now establish agent incentive compatibility and individual rationality. Observe that, for any $\theta, \theta' \in \Theta$, by the construction of mechanism $\widetilde{M}$, the interim utility of type θ for reporting θ' is

$$\begin{aligned}\mathcal{V}(\theta', \theta; \widetilde{M}) &\triangleq \mathbf{E}_{\omega \sim F_\Omega(\theta)}\left[\mathbf{E}_{x \sim \tilde{\mathbf{x}}(\omega, \theta')}[V(\omega, \theta, x)]\right] \\ &= \mathbf{E}_{\omega \sim F_\Omega(\theta)}\left[\mathbf{E}_{s \sim \sigma(\omega)}\left[\mathbf{E}_{x \sim \mathbf{x}(s, \theta')}[V(\omega, \theta, x)]\right]\right] = \mathcal{V}(\theta', \theta; M).\end{aligned}$$

Let $\mathcal{V}(\theta; \widetilde{M}) \triangleq \mathcal{V}(\theta, \theta; \widetilde{M})$ be the interim utility of the agent for truthful reporting. Since mechanism M is incentive compatible and individually rational for the agent, it follows that

$$\mathcal{V}(\theta; M) \geq \max\left\{\mathcal{V}(\theta', \theta; M), 0\right\}$$

for all $\theta, \theta' \in \Theta$. Therefore, we have

$$\mathcal{V}(\theta; \widetilde{M}) \geq \max\left\{\mathcal{V}(\theta', \theta; \widetilde{M}), 0\right\}$$

for all $\theta, \theta' \in \Theta$. Hence mechanism $\widetilde{M}$ is also incentive compatible and individually rational for the agent.

We finally establish principal incentive compatibility. Suppose without loss of generality that $\Omega = \{0, 1\}$ and $S = [0, 1]$. Moreover, for any $s \in S$, the posterior belief of the principal for receiving signal s is $(1 - s)\delta_0 + s\delta_1$. For any $\omega \in \Omega$, let

$$\mathcal{U}_\omega(s; M) \triangleq \mathbf{E}_{\theta \sim F_\Theta(\omega)}\left[\mathbf{E}_{x \sim \mathbf{x}(s, \theta)}[U(\omega, \theta, x)]\right]$$

be the interim utility of the principal in mechanism M condition on the true feature being ω. Since mechanism M is principal incentive compatible, it follows that, for all $s, s' \in S$,

$$(1 - s) \cdot \mathcal{U}_0(s; M) + s \cdot \mathcal{U}_1(s; M) \geq (1 - s) \cdot \mathcal{U}_0(s'; M) + s \cdot \mathcal{U}_1(s'; M).$$

By standard arguments, for all $s, s' \in S$ such that $s \geq s'$,

$$\mathcal{U}_1(s; M) \geq \mathcal{U}_1(s'; M) \quad \text{and} \quad \mathcal{U}_0(s; M) \leq \mathcal{U}_0(s'; M).$$

Let the interim utility of the principal in mechanism $\widetilde{M}$ for reporting $\omega' \in \Omega$ be

$$\mathcal{U}(\omega', \omega; \widetilde{M}) \triangleq \mathbf{E}_{s \sim \sigma(\omega')}\left[\mathbf{E}_{\theta \sim F_\Theta(\omega)}\left[\mathbf{E}_{x \sim \mathbf{x}(s, \theta)}[U(\omega, \theta, x)]\right]\right] = \mathbf{E}_{s \sim \sigma(\omega')}[\mathcal{U}_\omega(s; M)]$$

and let $\mathcal{U}(\omega; \widetilde{M}) \triangleq \mathcal{U}(\omega, \omega; \widetilde{M})$. Recall that $\sigma(\omega)$ is the distribution over signals given information structure σ when the true feature is ω. In binary feature space, we know that

$\sigma(1)$ first order stochastically dominates $\sigma(0)$, which further implies that

$$\mathcal{U}(1;\widetilde{M}) = \mathbf{E}_{s\sim\sigma(1)}[\mathcal{U}_1(s;M)] \geq \mathbf{E}_{s\sim\sigma(0)}[\mathcal{U}_1(s;M)] = \mathcal{U}(0,1;\widetilde{M}),$$
$$\mathcal{U}(0;\widetilde{M}) = \mathbf{E}_{s\sim\sigma(0)}[\mathcal{U}_0(s;M)] \geq \mathbf{E}_{s\sim\sigma(1)}[\mathcal{U}_0(s;M)] = \mathcal{U}(1,0;\widetilde{M}).$$

Thus, mechanism $\widetilde{M}$ is also incentive compatible for the principal. ■

B.2 Suboptimality of Fully Revealing Information Structures

We prove Proposition 6 by directly characterizing the optimal mechanism in this setting. For any threshold feature ω and probability p, we define the information structure $(S, \sigma_{\omega,p})$ with $S = \{0,1\}$ as

$$\sigma_{\omega,p}(\omega') = \begin{cases} \delta_1 & d(\omega') > d(\omega) \\ p\cdot\delta_1 + (1-p)\cdot\delta_0 & d(\omega') = d(\omega) \\ \delta_0 & d(\omega') < d(\omega). \end{cases}$$

Let $\underline{\omega}$ be the feature that minimizes $d(\underline{\omega})$ subject to the constraints that there exists $p \in [0,1]$ such that $\mathbf{E}_{\sigma_{\underline{\omega},p}}[d(\omega) \mid s = 1] \geq 0$. Moreover, let $\underline{p} \in [0,1]$ be the maximum probability such that the above inequality holds. Intuitively, $\underline{\omega}$ is the threshold feature value and $\underline{p}$ is the threshold probability such that the principal weakly prefers allocation 1 over 0 when receiving a signal that pools features with payoff differences above $\underline{\omega}$.

If $\mathbf{Pr}_{\sigma_{\underline{\omega},\underline{p}}}[s=1] < \frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)}$, let $\omega^* = \underline{\omega}$ and $p^* = 1$. Otherwise, let ω^* be the feature that maximizes $d(\omega^*)$ subject to the constraints that $d(\underline{\omega}) \leq d(\omega^*) < 0$ and there exists $p \in [0,1]$ such that $\mathbf{Pr}_{\sigma_{\omega^*,p}}[s=1] \geq \frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)}$. Moreover, let $p^* \in [0,1]$ be the minimum probability such that the above inequality holds. Essentially, ω^* and p^* provides incentives for the agent to accept the mechanism given information structure σ_{ω^*,p^*}. Let

$$\hat{p} = \frac{1}{1-\mathbf{Pr}_{\sigma_{\omega^*,p^*}}[s=1]} \cdot \max\left\{0, \frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)} - \mathbf{Pr}_{\sigma_{\omega^*,p^*}}[s=1]\right\}. \tag{1}$$

Lemma 7. *In independent private value environments with degenerate agent type and binary allocations, mechanism $M = (S, \sigma_{\omega^*,p^*}, \mathbf{x})$ with $S = X = \{0,1\}$ is optimal if*

- $\mathbf{x}(s,\theta) = o$ *for all* $s \in S$ *when* $\mathbf{E}[U(\omega,1)] - (1-\hat{p})\cdot\mathbf{E}_{\sigma_{\omega^*,p^*}}[d(\omega)\cdot\mathbb{1}[s=0]] < 0$*;*
- $\mathbf{x}(1,\theta) = 1$ *with probability 1 and* $\mathbf{x}(0,\theta) = 1$ *with probability* $\hat{p}$ *when* $\mathbf{E}[U(\omega,1)] - (1-\hat{p})\cdot\mathbf{E}_{\sigma_{\omega^*,p^*}}[d(\omega)\cdot\mathbb{1}[s=0]] \geq 0$.

Essentially, Lemma 7 implies that if the principal wants to incentivize the agent to participate, the information structure in the optimal mechanism is $(S, \sigma_{\omega^*,p^*})$ with $S = \{0, 1\}$. Moreover, the item may be allocated to the agent with strictly positive probability even if the principal's received signal is 0 in order to satisfies the agent's individual rationality constraint. The expected payoff of this mechanism is $\mathbf{E}[U(\omega, 1)] - (1-\hat{p}) \cdot \mathbf{E}_{\sigma_{\omega^*,p^*}}[d(\omega) \cdot \mathbb{1}[s = 0]]$ and hence the principal wants to incentivize the agent to participate if and only if the above term is non-negative.

Proof of Lemma 7. We first characterize the optimal mechanism when the agent is incentivized to accept the mechanism.

In this binary allocation model, for any feasible mechanism $M = (S, \sigma, \mathbf{x})$, let $S_0 = \{s \in S : \mathbf{E}_\sigma[d(\omega) \mid s] < 0\}$ and let $S_1 = \{s \in S : \mathbf{E}_\sigma[d(\omega) \mid s] \geq 0\}$. That is, principal with signal in S_0 would strictly prefer allocation 0 to 1, and principal with signal in S_1 would weakly prefer allocation 1 to 0. Moreover, for any feasible mechanism M, principal with signal s would strictly prefer the report with lowest probability allocation 1 is chosen if $s \in S_0$ and weakly prefer the report with highest probability allocation 1 is chosen if $s \in S_1$. Note that it is without loss to maximize the probability allocation 1 is chosen, such that $\mathbf{x}(1, \theta) = 1$, when the principal is indifferent since it both improves the principal's expected payoff and alleviates the agent's individual rationality constraint. Therefore, it is without loss to function on feasible mechanisms with allocation rule such that $\mathbf{x}(s, \theta) = \mathbf{x}(s', \theta)$ if $s, s' \in S_0$ and $\mathbf{x}(s, \theta) = 1$ if $s \in S_1$. In this case, it is also without loss of generality to focus on signal space where $S = \{0, 1\}$ where allocation 1 is chosen when $s = 1$.

By the individual rationality constraints for the agent, the ex ante probability that allocation 1 is chosen in the mechanism should be at least $\frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)}$. Moreover, to satisfy the incentive constraints for the principal, the posterior belief over the payoff difference given signal 1 should be non-negative, i.e., $\mathbf{E}_\sigma[d(\omega) \mid s = 1] \geq 0$. To maximizes the principal's expected payoff subject to these constraints, the optimal information structure is σ_{ω^*,p^*}, which is attained by greedily pooling features with highest payoff differences into signal 1 until one of the constraints is binding. Moreover, $\hat{p} = \mathbf{x}(0, \theta)$ is chosen to respect the agent's individual rationality constraints, which can be verified to take the form in Equation (1).

Finally, the expected payoff of the principal in the optimal mechanism that incentivize the agent to accept is

$$\mathbf{E}[U(\omega, 1)] - (1 - \hat{p}) \cdot \mathbf{E}_{\sigma_{\omega^*,p^*}}[d(\omega) \cdot \mathbb{1}[s = 0]].$$

Therefore, the principal prefers this mechanism compared to taking the outside option if and only if the above term is non-negative. ■

Proof of Proposition 6. We first characterize the optimal mechanism with fully revealing information structures. Let $\Omega_1 = \{\omega \in \Omega : d(\omega) \geq 0\}$ and $\Omega_0 = \{\omega \in \Omega : d(\omega) < 0\}$. Similar to the proof of Lemma 7, the optimal mechanism $\bar{M} = \{\Omega, \bar{\sigma}, \bar{\mathbf{x}}\}$ with fully revealing information structure $\bar{\sigma}$ satisfies that $\bar{\mathbf{x}}(\omega, \theta) = 1$ for all $\omega \in \Omega_1$, and $\bar{\mathbf{x}}(\omega, \theta) = 1$ with probability p for all $\omega \in \Omega_0$, where $p \in [0, 1]$ is the minimum probability such that the agent's individual rationality constraint is satisfied, i.e., $p \cdot F_\Omega[\Omega_0] + F_\Omega[\Omega_1] = \frac{-V(\theta,0)}{V(\theta,1)-V(\theta,0)}$. By condition 2, such p exists and is in the interior of $(0, 1)$.

In comparison to the optimal mechanism $M = (S, \sigma_{\omega^*,p^*}, \mathbf{x})$ characterized in Lemma 7, the ex ante probability allocation 1 is chosen is the same in both mechanisms. However, condition 2 together with the fact that there exist ω such that $d(\omega) > 0$ in condition 3 implies that $d(\omega^*) < 0$, and hence in mechanism M allocation 0 is chosen with higher probability for features with lower payoff differences, instead of distributed uniformly across types in Ω_0 as in mechanism $\bar{M}$. This strictly improves the expected payoff of the principal under condition 3 due to the rearrangement inequality.

Finally, condition 1 implies that the expected payoff of the principal is strictly positive in the optimal mechanism. Therefore, expected payoff of the principal in the optimal mechanism is strictly higher than the optimal mechanism with fully revealing information structures. ■

C Omitted Analysis for Section 5

C.1 Omitted Analysis of Example 1

Claim 1. *In the environment given in Example 1, when principal interim individual rational constraints are imposed, every optimal mechanism $(S, \sigma, \mathcal{M})$ must satisfy $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_1]] = 200/209$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_1]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] = 200/209$, and $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$.*

Proof. First, we establish the following preliminary fact.

$$\left\{\left(0, \frac{9}{209}, \frac{200}{209}, \frac{9}{209}\right)\right\} = \underset{(z_1,z_2,z_3,z_4)\in[0,1]^4}{\arg\max} -50z_1 - \frac{1}{2}z_2 + 50z_3 + \frac{1}{2}z_4$$
$$\text{s.t. } z_1 + z_2 \leq 1,$$
$$z_3 + z_4 \leq 1, \tag{2}$$
$$z_4 \geq z_2,$$
$$1 - \frac{11}{2}z_1 + \frac{99}{2}z_2 - \frac{11}{2}z_3 + \frac{99}{2}z_4 \geq 0.$$

To see this, observe that, for all (z_1, z_2, z_3, z_4) that maximize the given objective subject to the given constraints, it must be that $z_1 = 0$ and $z_3 + z_4 = 1$. Additionally, for all (z_1, z_2, z_3, z_4) that satisfy the given constraints and satisfy $z_2 < z_4$, there are ways in which z_2 could be increased, z_4 could be decreased, and z_3 could be increased that would result in a strictly higher value of the objective while still preserving the constraints. Hence, for all (z_1, z_2, z_3, z_4) that maximize the given objective subject to the given constraints, it must be that $z_2 = z_4$. Combining these findings with the facts that $-(1/2)a + 50(1-a) + (1/2)a = 50 - 50a$ for all $a \in \mathbb{R}$, $1 + (99/2)a - (11/2)(1-a) + (99/2)a = -9/2 + (209/2)a$ for all $a \in \mathbb{R}$, and

$$\left\{\frac{9}{209}\right\} = \arg\max_{a \in [0,1]} 50 - 50a \quad \text{s.t.} \quad -\frac{9}{2} + \frac{209}{2}a \geq 0,$$

gives (2) as a consequence.

Now, observe that, for an arbitrary incentive compatible and individually rational mechanism $(S, \sigma, \mathcal{M})$, the ex-ante expected payoff to the principal is $-50\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] - (1/2)\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] + 50\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] + (1/2)\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]]$ and the ex-ante expected payoff to the agent is $1 - (11/2)\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] + (99/2)\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] - (11/2)\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] + (99/2)\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]]$. Additionally, since $u(\theta_1, \phi, x_1) = u(\theta_2, \phi, x_2) = 0$, $u(\theta_1, \phi, x) = -u(\theta_2, \phi, x) < 0$ for all $x \in X \setminus \{x_1\}$, $\lambda[\theta_1] = \lambda[\theta_2] = 1/2$, and $(S, \sigma, \mathcal{M})$ is individually rational, it must be that $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] \geq \mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]]$. Thus, we have that, every incentive compatible and individually rational mechanism must result in a weakly lower ex-ante expected payoff to the principal than 10000/2009, the corresponding optimal value in the problem given by (2). Additionally, every mechanism $(S, \sigma, \mathcal{M})$ in which either $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_1]] \neq 200/209$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] \neq 0$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] \neq 9/209$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_1]] \neq 0$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] \neq 200/209$, or $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] \neq 9/209$ must result in a strictly lower ex-ante expected payoff to the principal than 10000/2009.

Now we complete the argument by showing that there are incentive compatible and individually rational mechanisms that achieve an ex-ante expected payoff to the principal of 10000/209, the optimal value in the problem given by (2). Consider a signal set $S^* = \{s_1, s_2, s_3\}$ with precisely 3 signals and the mechanism $(S^*, \sigma^*, \mathcal{M}^*)$ in which $\sigma^* : \Theta \to \Delta(S^*)$ is given by

$$\sigma^*(\theta) = \begin{cases} \frac{200}{209}\delta_{s_1} + \frac{9}{209}\delta_{s_3} & \text{if } \theta = \theta_1, \\ \frac{200}{209}\delta_{s_2} + \frac{9}{209}\delta_{s_3} & \text{if } \theta = \theta_2, \end{cases}$$

for all $\theta \in \Theta$ and the allocation rule $\mathcal{M}^* : S^* \times \Phi \to \Delta(X)$ is given by

$$\mathcal{M}^*(s,\phi) = \begin{cases} \delta_{x_1} & \text{if } s = s_1, \\ \delta_{x_2} & \text{if } s = s_2, \\ \delta_{x_3} & \text{if } s = s_3, \end{cases}$$

for all $(s,\phi) \in S \times \Phi$. Observe that 10000/209 is the principal's expected payoff from $(S^*, \sigma^*, \mathcal{M}^*)$. Additionally, with information structure (S^*, σ^*), δ_{θ_1} is the posterior belief of the principal upon observing s_1, δ_{θ_2} is the posterior belief of the principal upon observing s_2, and $(1/2)\delta_{\theta_1} + (1/2)\delta_{\theta_2}$ is the posterior belief of the principal upon observing s_3. Thus, $(S^*, \sigma^*, \mathcal{M}^*)$ is incentive compatible and individually rational for the principal. Moreover, the expected payoff of the agent from $(S^*, \sigma^*, \mathcal{M}^*)$ is 0. Hence, $(S^*, \sigma^*, \mathcal{M}^*)$ is incentive compatible and individually rational. ■

Claim 2. *In the environment given in Example 1, every individually rational mechanism $(S, \sigma, \mathcal{M})$ that induces at most 2 signals with strictly positive probability cannot satisfy $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_1]] = 200/209$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_1]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] = 200/209$, and $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$.*

Proof. Consider an arbitrary mechanism $(S, \sigma, \mathcal{M})$ that induces at most 2 signals with strictly positive probability. Suppose that $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_1]] = 200/209$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_1]] = 0$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] = 200/209$, and $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] = 9/209$. Then there must be signals $s_1 \in S$ and $s_2 \in S$ such that $\sigma(\theta_1)[s_1] = 1$, $\sigma(\theta_2)[s_2] = 1$, $\mathcal{M}(s_1,\phi)[x_1]] = 200/209$, $\mathcal{M}(s_1,\phi)[x_1]] = 9/209$, and $\mathcal{M}(s_2,\phi)[x_2]] = 200/209$. Consequently, $(S, \sigma, \mathcal{M})$ would not be interim individually rational for the principal upon observing signal s_1, since the resulting interim expected payoff would be $-9/209$. Thus, it follows that, for every individually rational mechanism $(S, \sigma, \mathcal{M})$ that induces at most 2 signals with strictly positive probability, either $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_1]] \neq 200/209$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_2]] \neq 0$, $\mathbb{E}_{s\sim\sigma(\theta_1)}[\mathcal{M}(s,\phi)[x_3]] \neq 9/209$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_1]] \neq 0$, $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_2]] \neq 200/209$, or $\mathbb{E}_{s\sim\sigma(\theta_2)}[\mathcal{M}(s,\phi)[x_3]] \neq 9/209$ must hold. ■

ONLINE APPENDIX

OA 1 Proof of Proposition 1

Proposition 1 follows immediately from Lemmas 8 and 9 below. Here we establish some preliminaries that will be useful for the proofs in this section. For an arbitrary direct mechanism $(\sigma, \mathbf{x}) \in \mathcal{M}$, let $\mathbf{q}(\sigma, \mathbf{x}) \in \Delta(\Omega \times \Delta(\Omega) \times \Delta(X)^{\Theta})$ be the probability distribution over $\Omega \times \Delta(\Omega) \times \Delta(X)^{\Theta}$ obtained by first drawing $\omega \in \Omega$ according to F_{Ω}, then drawing $s \in \Delta(\Omega)$ according to $\sigma(\omega)$, and then, with probability 1, drawing $\xi \in \Delta(X)^{\Theta}$ given by $\xi(\theta) = \mathbf{x}(\omega, \theta)$ for all $\theta \in \Theta$. Furthermore, for arbitrary $(\sigma, \mathbf{x}) \in \mathcal{M}$, let $\mathbf{p}(\sigma, \mathbf{x}) = \text{marg}_{\Delta(\Omega)\times\Delta(X)^{\Theta}}(\mathbf{q}(\sigma, \mathbf{x}))$. Observe that, for arbitrary $(\sigma, \mathbf{x}) \in \mathcal{M}$, $\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \mathbf{E}_{(s,\xi)\sim\mathbf{p}(\sigma,\mathbf{x})}\left[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]\right]$.

Lemma 8. *For all environments in which X is compact and U and V are continuous, there exist solutions to the problem given by (OPT).*

Proof. Let $(\sigma_n, \mathbf{x}_n)_{n\in\mathbb{N}} \in (\mathcal{M}^F)^{\mathbb{N}}$ be such that $\lim_{n\to\infty} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma_n(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \sup_{(\sigma,\mathbf{x})\in\mathcal{M}^F} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]$. For all $n \in \mathbb{N}$, let $q_n = \mathbf{q}(\sigma_n, \mathbf{x}_n)$. Without loss of generality, suppose that $(q_n)_{n\in\mathbb{N}} \in \Delta(\Omega \times \Delta(\Omega) \times \Delta(X)^{\Theta})^{\mathbb{N}}$ is convergent and let $q \in \Delta(\Omega \times \Delta(\Omega) \times \Delta(X)^{\Theta})$ be such that $\lim_{n\to\infty} q_n = q$. For all $n \in \mathbb{N}$, let $p_n = \mathbf{p}(\sigma_n, \mathbf{x}_n)$, and further let $p = \text{marg}_{\Delta(\Omega)\times\Delta(X)^{\Theta}}(q)$. Observe that $\lim_{n\to\infty} p_n = p$ and

$$\begin{aligned}
&\mathbf{E}_{(s,\xi)\sim p}\left[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]\right] \\
&= \lim_{n\to\infty} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma_n(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] \\
&= \sup_{(\sigma,\mathbf{x})\in\mathcal{M}^F} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right].
\end{aligned}$$

Let $\Xi = \text{supp}(\text{marg}_{\Delta(X)^{\Theta}}(p))$.

We will establish that

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] = \max_{\tilde{\xi}\in\Xi} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\tilde{\xi}(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1. \quad (3)$$

Since Ξ is a compact metric space, there is a countable dense subset of Ξ. Let $\{\xi_m \in \Xi\}_{m\in\mathbb{N}} \subseteq \Xi$ be a countable dense subset of Ξ. Note that (3) is implied by

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1 \quad (4)$$

holding for all $m \in \mathbb{N}$. To see this, observe that, by standard arguments,

$$\{(s,\xi) \in \Delta(\Omega) \times \Delta(X)^{\Theta} : \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] = \max_{\tilde{\xi}\in\Xi} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\tilde{\xi}(\theta)}\left[U(\omega,\theta,x)\right]\right]\}$$

$$= \cap_{m\in\mathbb{N}} \{(s,\xi) \in \Delta(\Omega) \times \Delta(X)^{\Theta} : \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right]\},$$

so (4) holding for all $m \in \mathbb{N}$ implies (3). Fix arbitrary $m \in \mathbb{N}$. There exists $(\xi_{m,n})_{n\in\mathbb{N}} \in (\Delta(X)^{\Theta})^{\mathbb{N}}$ such that $\xi_{m,n} \in \text{supp}(\text{marg}_{\Delta(X)^{\Theta}}(p_n))$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \xi_{m,n} = \xi_m$. Consider arbitrary $n \in \mathbb{N}$. By PIC, for all $s \in \Delta(\Theta)$,

$$\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_{m,n}(\theta)}\left[U(\omega,\theta,x)\right]\right],$$

so

$$\mathbb{P}_{(s,\xi)\sim p_n}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_{m,n}(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1.$$

By standard arguments then, for all $\varepsilon \in \mathbb{R}_{++}$,

$$\lim_{n\to\infty} \mathbb{P}_{(s,\xi)\sim p_n}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right] - \varepsilon] = 1,$$

which, combined with the fact that $\lim_{n\to\infty} p_n = p$ under the relevant topology of weak convergence, gives

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right] - \varepsilon] = 1.$$

Consequently, (4) holds for m and, by the arbitrariness of $m \in M$ and the equivalence stated earlier, we conclude that (3) holds.

Let $\chi : \Delta(\Omega) \to \Delta(\Xi)$ be a regular conditional probability distribution of $\Delta(X)^{\Theta}$ given $\Delta(\Omega)$ for the probability distribution p. Further, let $\boldsymbol{\xi} : \Delta(\Omega) \to \Delta(X)^{\Theta}$ be the measurable mapping such that, for all $s \in \Delta(\Omega)$, $\boldsymbol{\xi}(s)(\theta)[\widetilde{X}] = \mathbf{E}_{\xi\sim\chi(s)}\left[\mathbb{P}_{x\sim\xi(\theta)}[\widetilde{X}]\right]$ for all measurable $\widetilde{X} \subseteq X$ for all $\theta \in \Theta$. By the Kuratowski and Ryll-Nardzewski measurable selection theorem, there exists some measurable mapping $\boldsymbol{\xi}' : \Delta(\Omega) \to \Xi$ such that $\boldsymbol{\xi}'(s) \in \arg\max_{\xi\in\Xi} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\boldsymbol{\xi}(s)(\omega)(\theta)}\left[U(\omega,\theta,x)\right]\right]$ for all $s \in \Delta(\Omega)$. By (3), there exists some measurable $S \subseteq \Delta(\Omega)$ such that $p[S] = 1$ and $\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\boldsymbol{\xi}(s)(\theta)}\left[U(\omega,\theta,x)\right]\right] = \max_{\xi\in\Xi} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]$. Let $\sigma : \Omega \to \Delta(\Delta(\Omega))$ be the information structure such that, for all $\omega \in \Omega$,

$$\sigma(\omega)[\widetilde{S}] = \frac{\mathbf{E}_{s\sim\text{marg}_{\Delta(\Omega)}(p)}\left[s[\omega]\right]}{F_{\Omega}[\omega]}$$

for all measurable $\widetilde{S} \subseteq \Delta(\Omega)$. Let $\mathbf{x} : \Delta(\Omega) \times \Theta \to \Delta(X)$ be the allocation rule given by

$$\mathbf{x}(s,\theta) = \begin{cases} \boldsymbol{\xi}(s)(\theta) & \text{if } s \in S, \\ \boldsymbol{\xi}'(s)(\theta) & \text{if } s \notin S, \end{cases}$$

for all $(s,\theta) \in \Delta(\Omega) \times \Theta$. By construction, $\mathbf{p}(\sigma,\mathbf{x}) = p$. Thus, $\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \mathbf{E}_{(s,\xi)\sim p}\left[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \sup_{(\tilde{\sigma},\tilde{\mathbf{x}})\in\mathcal{M}^F}\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\tilde{\sigma}(\omega)}\left[\mathbf{E}_{x\sim\tilde{\mathbf{x}}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]$. We conclude the proof by showing that $(\sigma,\mathbf{x}) \in \mathcal{M}^F$.

As $\mathbf{x}_n(s,\theta)[o] = \mathbf{x}_n(s',\theta)[o]$ for all $n \in \mathbb{N}, s, s' \in \Delta(\Omega), \theta \in \Theta$, it must be that $\xi(\theta) = \xi'(\theta)$ for all $\xi, \xi' \in \Xi, \theta \in \Theta$. Therefore, $\mathbf{x}(s,\theta) = \mathbf{x}(s',\theta)$ for all $s, s' \in \Delta(\Omega), \theta \in \Theta$ and so $(\sigma,\mathbf{x})$ satisfies the relevant consistency constraints. Moreover, by construction, $s \in \arg\max_{s'\in\Delta(\Omega)}\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\mathbf{x}(s',\theta)}\left[U(\omega,\theta,x)\right]\right]$ for all $s \in \Delta(\Omega)$, so $(\sigma,\mathbf{x})$ satisfies the PIC constraints. Finally, for all $(\tilde{\sigma},\tilde{x}) \in \mathcal{M}$, $\mathbf{E}_{(\omega,s)\sim H(\theta,\tilde{\sigma})}\left[\mathbf{E}_{x\sim\tilde{\mathbf{x}}(s,\theta)}\left[V(\omega,\theta,x)\right]\right] = \mathbf{E}_{(s,\xi)\sim\mathbf{p}(\tilde{\sigma},\tilde{x})}\left[\sum_{\omega\in\Omega}\mathbf{F}_\Theta(\omega)[\theta]\mathbf{E}_{x\sim\xi(\theta')}\left[V(\omega,\theta,x)\right]\right]/F_\Theta[\theta]$. Combining this with the fact that, for all $n \in \mathbb{N}$, $(\sigma_n,\mathbf{x}_n)$ satisfies the AIC and AIR constraints gives

$$\mathbf{E}_{(s,\xi)\sim p_n}\left[\sum_{\omega\in\Omega}\mathbf{F}_\Theta(\omega)[\theta]\mathbf{E}_{x\sim\xi(\theta)}\left[V(\omega,\theta,x)\right]\right] \geq \max\left\{\max_{\theta'\in\Theta}\mathbf{E}_{(s,\xi)\sim p_n}\left[\sum_{\omega\in\Omega}\mathbf{F}_\Theta(\omega)[\theta]\mathbf{E}_{x\sim\xi(\theta')}\left[V(\omega,\theta,x)\right]\right], 0\right\}$$

for all $n \in \mathbb{N}, \theta \in \Theta$. From this and the fact that $\lim_{n\to\infty} p_n = p$, it follows that

$$\mathbf{E}_{(s,\xi)\sim p}\left[\sum_{\omega\in\Omega}\mathbf{F}_\Theta(\omega)[\theta]\mathbf{E}_{x\sim\xi(\theta)}\left[V(\omega,\theta,x)\right]\right] \geq \max\left\{\max_{\theta'\in\Theta}\mathbf{E}_{(s,\xi)\sim p}\left[\sum_{\omega\in\Omega}\mathbf{F}_\Theta(\omega)[\theta]\mathbf{E}_{x\sim\xi(\theta')}\left[V(\omega,\theta,x)\right]\right], 0\right\}$$

for all $\theta \in \Theta$ and, thus, $(\sigma,\mathbf{x})$ satisfies the AIC and AIR constraints. ■

Lemma 9. *For all environments in which X is compact and U and V are continuous, there exist solutions to the problem given by (OPT-IR).*

Proof. Let $(\sigma_n,\mathbf{x}_n)_{n\in\mathbb{N}} \in (\mathcal{M}^{F,IR})^{\mathbb{N}}$ be such that

$$\begin{aligned} &\lim_{n\to\infty}\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma_n(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] \\ &= \sup_{(\sigma,\mathbf{x})\in\mathcal{M}^{F,IR}}\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]. \end{aligned}$$

For all $n \in \mathbb{N}$, let $q_n = \mathbf{q}(\sigma_n,\mathbf{x}_n)$. Without loss of generality, suppose that $(q_n)_{n\in\mathbb{N}} \in \Delta(\Omega\times\Delta(\Omega)\times\Delta(X)^\Theta)^{\mathbb{N}}$ is convergent and let $q \in \Delta(\Omega\times\Delta(\Omega)\times\Delta(X)^\Theta)$ be such that $\lim_{n\to\infty} q_n = q$. For all $n \in \mathbb{N}$, let $p_n = \mathbf{p}(\sigma_n,\mathbf{x}_n)$, and further let $p = \operatorname{marg}_{\Delta(\Omega)\times\Delta(X)^\Theta}(q)$.

Observe that $\lim_{n\to\infty} p_n = p$ and

$$\begin{aligned}
&\mathbf{E}_{(s,\xi)\sim p}\left[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]\right] \\
&= \lim_{n\to\infty} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma_n(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] \\
&= \sup_{(\sigma,\mathbf{x})\in\mathcal{M}^F} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right].
\end{aligned}$$

Let $\Xi = \mathrm{supp}(\mathrm{marg}_{\Delta(X)^\Theta}(p))$.

Furthermore, for all $n \in \mathbb{N}$, let $\hat{p}_n =\in \Delta(\Delta(\Omega)\times\Delta(X)^\Theta)^{\mathbb{N}}$ be the probability distribution over $\Delta(\Omega)\times\Delta(X)^\Theta$ obtained by first drawing drawing $s \in \Delta(\Omega)$ according to $U(\Delta(\Omega))$ and then, with probability 1, drawing $\xi \in \Delta(X)^\Theta$ given by $\xi(\theta) = \mathbf{x}(\omega,\theta)$ for all $\theta \in \Theta$. Without loss of generality, suppose that $(\hat{p}_n)_{n\in\mathbb{N}} \in \Delta(\Delta(\Omega)\times\Delta(X)^\Theta)^{\mathbb{N}}$ is convergent and let $\hat{p} \in \Delta(\Delta(\Omega)\times\Delta(X)^\Theta)$ be such that $\lim_{n\to\infty}\hat{p}_n = \hat{p}$. Let $\widehat{\Xi} = \mathrm{supp}(\mathrm{marg}_{\Delta(X)^\Theta}(\hat{p}))$.

We will establish that

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] = \max_{\tilde{\xi}\in\Xi\cup\widehat{\Xi}} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\tilde{\xi}(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1. \tag{5}$$

Since $\Xi\cup\widehat{\Xi}$ is a compact metric space, there is a countable dense subset of $\Xi\cup\widehat{\Xi}$. Let $\{\xi_m \in \Xi\cup\widehat{\Xi}\}_{m\in\mathbb{N}} \subseteq \Xi\cup\widehat{\Xi}$ be a countable dense subset of $\Xi\cup\widehat{\Xi}$. Note that (5) is implied by

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1 \tag{6}$$

holding for all $m \in \mathbb{N}$, as can be shown using a similar argument to the corresponding argument given in the proof of Lemma 8. Fix arbitrary $m \in \mathbb{N}$. There exists $(\xi_{m,n})_{n\in\mathbb{N}} \in (\Delta(X)^\Theta)^{\mathbb{N}}$ such that $\xi_{m,n} \in \mathrm{supp}(\mathrm{marg}_{\Delta(X)^\Theta}(p_n))\cup\mathrm{supp}(\mathrm{marg}_{\Delta(X)^\Theta}(\hat{p}_n))$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}\xi_{m,n} = \xi_m$. Consider arbitrary $n \in \mathbb{N}$. By PIC, for all $s \in \Delta(\Theta)$,

$$\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\mathbf{x}_n(s,\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_{m,n}(\theta)}\left[U(\omega,\theta,x)\right]\right],$$

so

$$\mathbb{P}_{(s,\xi)\sim p_n}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_{m,n}(\theta)}\left[U(\omega,\theta,x)\right]\right]] = 1.$$

By standard arguments then, for all $\varepsilon \in \mathbb{R}_{++}$,

$$\lim_{n\to\infty}\mathbb{P}_{(s,\xi)\sim p_n}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right] - \varepsilon] = 1,$$

which, combined with the fact that $\lim_{n\to\infty} p_n = p$ under the relevant topology of weak

convergence, gives

$$\mathbb{P}_{(s,\xi)\sim p}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi_m(\theta)}\left[U(\omega,\theta,x)\right]\right] - \varepsilon] = 1.$$

Consequently, (6) holds for m and, by the arbitrariness of $m \in M$ and the equivalence stated earlier, we conclude that (5) holds.

A similar argument establishes that $\mathbb{P}_{(s,\xi)\sim\hat{p}}[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}[U(\omega,\theta,x)]\right] \geq 0] = 1$. Since $\text{supp}(\text{marg}_{\Delta(\Omega)}(\hat{p})) = \Delta(\Omega)$, it follows that $\max_{\tilde{\xi}\in\widehat{\Xi}} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\tilde{\xi}(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq 0$ for all $s \in \Delta(\Omega)$.

Let $\chi : \Delta(\Omega) \to \Delta(\Xi)$ be a regular conditional probability distribution of $\Delta(X)^\Theta$ given $\Delta(\Omega)$ for the probability distribution p. Further, let $\boldsymbol{\xi} : \Delta(\Omega) \to \Delta(X)^\Theta$ be the measurable mapping such that, for all $s \in \Delta(\Omega)$, $\boldsymbol{\xi}(s)(\theta)[\widetilde{X}] = \mathbf{E}_{\xi\sim\chi(s)}\left[\mathbb{P}_{x\sim\xi(\theta)}[\widetilde{X}]\right]$ for all measurable $\widetilde{X} \subseteq X$ for all $\theta \in \Theta$. By the Kuratowski and Ryll-Nardzewski measurable selection theorem, there exists some measurable mapping $\boldsymbol{\xi}' : \Delta(\Omega) \to \Xi \cup \widehat{\Xi}$ such that $\boldsymbol{\xi}'(s) \in \arg\max_{\xi\in\Xi\cup\widehat{\Xi}} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\boldsymbol{\xi}(s)(\omega)(\theta)}\left[U(\omega,\theta,x)\right]\right]$ for all $s \in \Delta(\Omega)$. By (5), there exists some measurable $S \subseteq \Delta(\Omega)$ such that $p[S] = 1$ and $\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\boldsymbol{\xi}(s)(\theta)}\left[U(\omega,\theta,x)\right]\right] = \max_{\xi\in\Xi\cup\widehat{\Xi}} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]$. Let $\sigma : \Omega \to \Delta(\Delta(\Omega))$ be the information structure such that, for all $\omega \in \Omega$,

$$\sigma(\omega)[\widetilde{S}] = \frac{\mathbf{E}_{s\sim\text{marg}_{\Delta(\Omega)}(p)}\left[s[\omega]\right]}{F_\Omega[\omega]}$$

for all measurable $\widetilde{S} \subseteq \Delta(\Omega)$. Let $\mathbf{x} : \Delta(\Omega) \times \Theta \to \Delta(X)$ be the allocation rule given by

$$\mathbf{x}(s,\theta) = \begin{cases} \boldsymbol{\xi}(s)(\theta) & \text{if } s \in S, \\ \boldsymbol{\xi}'(s)(\theta) & \text{if } s \notin S, \end{cases}$$

for all $(s,\theta) \in \Delta(\Omega) \times \Theta$. By construction, $\mathbf{p}(\sigma,\mathbf{x}) = p$. Thus, $\mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\sigma(\omega)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \mathbf{E}_{(s,\xi)\sim p}\left[\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right]\right] = \sup_{(\tilde{\sigma},\tilde{\mathbf{x}})\in\mathcal{M}^{F,IR}} \mathbf{E}_{(\omega,\theta)\sim F}\left[\mathbf{E}_{s\sim\tilde{\sigma}(\omega)}\left[\mathbf{E}_{x\sim\tilde{\mathbf{x}}(s,\theta)}\left[U(\omega,\theta,x)\right]\right]\right]$. We conclude the proof by showing that $(\sigma,\mathbf{x}) \in \mathcal{M}^{F,IR}$.

To establish that $(\sigma,\mathbf{x})$ satisfies the relevant PIC, AIC, AIR, and Consistency constraints, very similar arguments to the corresponding arguments given in the proof of Lemma 8 can be used. Furthermore, $\mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\mathbf{x}(s,\theta)}\left[U(\omega,\theta,x)\right]\right] \geq \max_{\xi\in\widehat{\Xi}} \mathbf{E}_{(\omega,\theta)\sim G(s)}\left[\mathbf{E}_{x\sim\xi(\theta)}\left[U(\omega,\theta,x)\right]\right] \geq 0$ for all $s \in \Delta(\Omega)$, so $(\sigma,\mathbf{x})$ satisfies the relevant PIR constraints. ■

OA 2 Cardinalities of Sets of Induced Interim Beliefs for Optimal Mechanisms

OA 2.1 Signal Sets in Quasilinear Environments with Singleton Agent Type Sets

The following example provides an environment in which $|\Omega| = 3$, $|\Theta| = 1$, and there is a mechanism that does strictly better than every mechanism that has strictly fewer than 3 signals or is fully revealing. Combining this with Proposition 5, it follows that, in this environment, there is an optimal mechanism in which exactly 3 signals are induced with strictly positive probability, every optimal mechanism must have at least 3 signals induced with strictly positive probability, and every optimal mechanism must be not fully revealing.

Example 2.

Both the principal's type set $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and allocation set $X = \{x_1, x_2, x_3\}$ are trinary while the agent's type set $\Theta = \{\theta\}$ is singleton. The prior distribution over the principal's type $\lambda \in \Delta(\Omega)$ is such that $\lambda[\omega_1] = \lambda[\omega_3] = 1/4$ and $\lambda[\omega_2] = 1/2$. The payoffs to the principal and the agent, net of transfers, from the various allocations are given in the following table. (The table is such that, for each $(\omega, x) \in \Omega \times X$, the first number in the corresponding pair of numbers gives the principal's payoff while the second number gives the agent's payoff.)

ω_1	x_1	x_2	x_3
	$0, 1$	$-1, 0$	$1.1, -2$

ω_2	x_1	x_2	x_3
	$0, 0$	$.5, .1$	$-2, 0$

ω_3	x_1	x_2	x_3
	$0, 0$	$0, -12$	$1, .1$

Table 6: The payoffs net of transfers for Example 6.

□

Claim 3. *In the environment given in Example 2, the highest expected payoff for the principal across the class of incentive compatible and individually rational mechanisms with at most* 2 *signals is* $11/20$.

Proof. We first show that $11/20$ is an upper bound on the expected payoffs for the principal across all incentive compatible and individually rational mechanisms with at most 2 signals. Since it is the case that, for each mechanism, the principal's expected payoff is weakly lower than the expected surplus generated by the mechanism, we can show that $11/20$ is such

an upper bound by demonstrating that, for each subset of the allocation set of size 2, the expected surplus generated by efficiently assigning principal types to allocations within this subset is weakly less than 11/20. For the $\{x_1, x_2\}$ subset of allocations, it would be efficient from a surplus perspective to assign ω_1 and ω_3 to x_1 and assign ω_2 to x_2, and doing so would result in an expected surplus of 11/20. For the $\{x_1, x_3\}$ subset of allocations, it would be efficient to assign ω_1 and ω_2 to x_1 and assign ω_3 to x_3, and doing so would result in an expected surplus of 21/40, which is strictly less than 11/20. For the $\{x_2, x_3\}$ subset of allocations, it would be efficient to assign ω_2 to x_2 and assign ω_1 and ω_3 to x_3, and doing so would result in an expected surplus of 7/20, which is strictly less than 11/20.

We now show that there is an incentive compatible and individually rational mechanism with 2 signals that achieves an expected payoff for the principal of 11/20. Consider a signal set $S = \{s_1, s_2\}$ with precisely 2 signals and the mechanism $(S, \sigma, \mathcal{M})$ in which $\sigma : \Omega \to \Delta(S)$ is given by

$$\sigma(\omega) = \begin{cases} \delta_{s_1} & \text{if } \omega \in \{\omega_1, \omega_3\}, \\ \delta_{s_2} & \text{if } \omega = \omega_2, \end{cases}$$

for all $\omega \in \Omega$ and the allocation rule $\mathcal{M} : S \times \Theta \to \Delta(X \times \mathbb{R})$ is given by

$$\mathcal{M}(s, \theta) = \begin{cases} \delta_{(x_1, -.3)} & \text{if } s = s_1, \\ \delta_{(x_2, -.3)} & \text{if } s = s_2, \end{cases}$$

for all $(s, \theta) \in S \times \Theta$. Observe that 11/20 is the principal's expected payoff from $(S, \sigma, \mathcal{M})$. Additionally, with information structure (S, σ), $(1/2)\delta_{\omega_1} + (1/2)\delta_{\omega_3}$ is the posterior belief of the principal upon observing s_1 and δ_{ω_2} is the posterior belief of the principal upon observing s_2, so $(S, \sigma, \mathcal{M})$ is incentive compatible for the principal. Moreover, the expected payoff of the agent is 0. Thus, $(S, \sigma, \mathcal{M})$ is incentive compatible and individually rational. ■

Claim 4. *In the environment given in Example 2, the highest expected payoff for the principal across the class of incentive compatible and individually rational mechanisms that are fully revealing is* 11/20.

Proof. We first show that 11/20 is an upper bound on the expected payoffs for the principal across all incentive compatible and individually rational mechanisms that are fully revealing. Consider an arbitrary incentive compatible and individually rational mechanism that is fully revealing and let $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$ denote its outcome. Incentive compatibility requires that

$$-\mathbb{P}_p[x_2|\omega_1] + 1.1\mathbb{P}_p[x_3|\omega_1] + \mathbb{E}_p[t|\omega_1] \geq -\mathbb{P}_p[x_2|\omega_3] + 1.1\mathbb{P}_p[x_3|\omega_3] + \mathbb{E}_p[t|\omega_3],$$
$$\mathbb{P}_p[x_3|\omega_3] + \mathbb{E}_p[t|\omega_3] \geq \mathbb{P}_p[x_3|\omega_1] + \mathbb{E}_p[t|\omega_1].$$

Combining these inequalities gives

$$\mathbb{P}_p[x_2|\omega_3] \geq \mathbb{P}_p[x_2|\omega_1] + .1(\mathbb{P}_p[x_3|\omega_3] - \mathbb{P}_p[x_3|\omega_1]).$$

Since the principal's expected utility from p is weakly less than the expected surplus generated by p, it thus follows that the principal's expected utility from p, denoted by $U(p)$, must satisfy

$$\begin{aligned} U(p) &\leq \frac{1}{4}\left(\mathbb{P}_p[x_1|\omega_1] - \mathbb{P}_p[x_2|\omega_1] - .9\mathbb{P}_p[x_3|\omega_1]\right) + \frac{1}{2}(.6) + \frac{1}{4}\left(-12\mathbb{P}_p[x_2|\omega_3] + 1.1\mathbb{P}_p[x_3|\omega_3]\right), \\ &\leq \frac{11}{20} - \frac{13}{4}\mathbb{P}_p[x_2|\omega_1] - \frac{7}{40}\mathbb{P}_p[x_3|\omega_1] - \frac{1}{40}\mathbb{P}_p[x_3|\omega_3]. \end{aligned}$$

As $\mathbb{P}_p[x_2|\omega_1], \mathbb{P}_p[x_3|\omega_1], \mathbb{P}_p[x_3|\omega_3] \geq 0$, it follows that $11/20 - 13\mathbb{P}_p[x_2|\omega_1]/4 - 7\mathbb{P}_p[x_3|\omega_1]/40 - \mathbb{P}_p[x_3|\omega_3]/40 \leq 11/20$. Thus, it follows that $U(p) \leq 11/20$ must hold.

We now show that there is an incentive compatible and individually rational mechanism that is fully revealing and achieves an expected payoff for the principal of $11/20$. Consider the mechanism $(\Omega, \sigma, \mathcal{M})$ in which $\sigma : \Omega \to \Delta(\Omega)$ is given by $\sigma(\omega) = \delta_\omega$ for all $\omega \in \Omega$ and the allocation rule $\mathcal{M} : \Omega \times \Theta \to \Delta(X \times \mathbb{R})$ is given by

$$\mathcal{M}(\omega, \theta) = \begin{cases} \delta_{(x_1, -.3)} & \text{if } \omega \in \{\omega_1, \omega_3\}, \\ \delta_{(x_2, -.3)} & \text{if } \omega = \omega_2 \end{cases}$$

for all $(\omega, \theta) \in \Omega \times \Theta$. Observe that $11/20$ is the principal's expected payoff from $(\Omega, \sigma, \mathcal{M})$. Moreover, $(\Omega, \sigma, \mathcal{M})$ is incentive compatible for the principal and the expected payoff of the agent under $(\Omega, \sigma, \mathcal{M})$ is 0, so $(S, \sigma, \mathcal{M})$ is incentive compatible and individually rational. ■

Claim 5. *In the environment given in Example 2, there is an incentive compatible and individually rational mechanism with exactly* 3 *signals that gives an expected payoff of* $27/40$ *to the principal.*

Proof. Consider a signal set $S = \{s_1, s_2, s_3\}$ with precisely 3 signals and the mechanism $(S, \sigma, \mathcal{M})$ in which $\sigma : \Omega \to \Delta(S)$ is given by

$$\sigma(\omega) = \begin{cases} \delta_{s_1} & \text{if } \omega = \omega_1, \\ \frac{1}{2}\delta_{s_1} + \frac{1}{2}\delta_{s_2} & \text{if } \omega = \omega_2, \\ \delta_{s_3} & \text{if } \omega = \omega_3, \end{cases}$$

for all $\omega \in \Omega$ and the allocation rule $\mathcal{M} : S \times \Theta \to \Delta(X \times \mathbb{R})$ is given by

$$\mathcal{M}(s,\theta) = \begin{cases} \delta_{(x_1,-.3)} & \text{if } s = s_1, \\ \delta_{(x_2,-.3)} & \text{if } s = s_2, \\ \delta_{(x_3,-.3)} & \text{if } s = s_3, \end{cases}$$

for all $(s,\theta) \in S \times \Theta$. Observe that 27/40 is the principal's expected payoff from $(S,\sigma,\mathcal{M})$. Additionally, with information structure (S,σ), $(1/2)\delta_{\omega_1} + (1/2)\delta_{\omega_2}$ is the posterior belief of the principal upon observing s_1, δ_{ω_2} is the posterior belief of the principal upon observing s_2, and δ_{ω_3} is the posterior belief of the principal upon observing s_3. Thus, $(S,\sigma,\mathcal{M})$ is incentive compatible for the principal. Moreover, the expected payoff of the agent is 0. Thus, $(S,\sigma,\mathcal{M})$ is incentive compatible and individually rational. ■

OA 2.2 Omitted Analysis of Example 1

Let $\overline{S} = 83/30$. Observe that $\overline{S}$ is the highest feasible expected surplus in the environment given in Example 1. Further, let $\overline{U} = 414883/150270$. Observe that $\overline{S} > \overline{U} > 2.760917$.

Claim 6. *In the environment given in Example 1, there is an incentive compatible and individually rational mechanism in which exactly* 4 *signals are induced with strictly positive probability that gives an expected payoff of* $\overline{U}$ *to the principal.*

Proof. Consider a signal set $S = \{s_1, s_2, s_3, s_4\}$ with precisely 4 signals and the mechanism $(S,\sigma,\mathcal{M})$ in which $\sigma : \Omega \to \Delta(S)$ is given by

$$\sigma(\omega) = \begin{cases} \frac{9991}{10018}\delta_{s_1} + \frac{27}{10018}\delta_{s_4} & \text{if } \omega = \omega_1, \\ \frac{9991}{10018}\delta_{s_2} + \frac{27}{10018}\delta_{s_4} & \text{if } \omega = \omega_2, \\ \delta_{s_3} & \text{if } \omega = \omega_3, \end{cases}$$

for all $\omega \in \Omega$ and the allocation rule $\mathcal{M} : S \times \Theta \to \Delta(X \times \mathbb{R})$ is given by

$$\mathcal{M}(s,\theta) = \begin{cases} \delta_{(x_1,-\frac{5000}{5009})} & \text{if } (s,\theta) = (s_1,\theta_1), \\ \delta_{(x_2,-\frac{5000}{5009})} & \text{if } (s,\theta) = (s_2,\theta_1), \\ \delta_{(x_3,-\frac{5000}{5009})} & \text{if } (s,\theta) = (s_3,\theta_1), \\ \delta_{(x_4,-\frac{5000}{5009})} & \text{if } (s,\theta) = (s_4,\theta_1), \\ \delta_{(x_1,-\frac{50000}{5009})} & \text{if } (s,\theta) = (s_1,\theta_2), \\ \delta_{(x_2,-\frac{50000}{5009})} & \text{if } (s,\theta) = (s_2,\theta_2), \\ \delta_{(x_3,-\frac{50000}{5009})} & \text{if } (s,\theta) = (s_3,\theta_2), \\ \delta_{(x_5,-\frac{50000}{5009})} & \text{if } (s,\theta) = (s_4,\theta_2), \end{cases}$$

for all $(s,\theta) \in S \times \Theta$. Observe that $\overline{U}$ is the principal's expected payoff from $(S,\sigma,\mathcal{M})$. Additionally, with information structure (S,σ), δ_{ω_1} is the posterior belief of the principal upon observing s_1, δ_{ω_2} is the posterior belief of the principal upon observing s_2, δ_{ω_3} is the posterior belief of the principal upon observing s_3, and $(1/2)\delta_{\omega_1} + (1/2)\delta_{\omega_2}$ is the posterior belief of the principal upon observing s_4. Thus, $(S,\sigma,\mathcal{M})$ is incentive compatible for the principal. Moreover, the expected payoff of the type θ_1 agent from honestly reporting their type would be 0 while their expected payoff from misreporting their type as θ_2 would be $-45000/5009$. Likewise, the expected payoff of the type θ_2 agent from honestly reporting their type would be 0 while their expected payoff from misreporting their type as θ_1 would be $-45000/5009$. Thus, $(S,\sigma,\mathcal{M})$ is incentive compatible and individually rational for the agent. ■

Claim 7. *In the environment given in Example 1, for every outcome $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$ induced by incentive compatible and individually rational mechanisms, the principal's expected payoff from p must be weakly less than $299/150 + 2000\mathbb{P}_p[x_4|\omega_1,\theta_1]/3$.*

Proof. Consider an arbitrary outcome $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$ that is induced by an incentive compatible and individually rational mechanism. Let $v_1 = \mathbb{P}_p[x_1|\omega_1,\theta_1] + \mathbb{P}_p[x_2|\omega_2,\theta_1] + \mathbb{P}_p[x_3|\omega_3,\theta_1] - 1000(\mathbb{P}_p[x_2|\omega_3,\theta_1] + \mathbb{P}_p[x_4|\omega_3,\theta_1] + \mathbb{P}_p[x_5|\omega_3,\theta_1])$. Observe that $\max\{v_1,0\}$ is an upper bound on both the value of the agent and the expected surplus generated by p conditional on θ_1 and $299/30$ is an upper bound on the expected surplus generated by p conditional on θ_2. Additionally, because p is incentive compatible and individually rational, $\max\{9v_1 - 10000\mathbb{P}_p[x_4|\omega_1,\theta_1]/3, 0\}$ is a lower bound on the expected utility of the type θ_2 agent. Thus, it follows that $(4/5)\max\{v_1,0\} + (1/5)(299/30 - \max\{9v_1 - 10000\mathbb{P}_p[x_4|\omega_1,\theta_1]/3, 0\})$ is an upper bound on the principal's expected pay-

off from p. As $(4/5)\max\{v_1, 0\} + (1/5)(299/30 - \max\{9v_1 - 10000\mathbb{P}_p[x_4|\omega_1, \theta_1]/3, 0\}) \leq 299/150 + 2000\mathbb{P}_p[x_4|\omega_1, \theta_1]/3 - \max\{v_1, 0\} \leq 299/150 + 2000\mathbb{P}_p[x_4|\omega_1, \theta_1]/3$, it follows that $299/150 + 2000\mathbb{P}_p[x_4|\omega_1, \theta_1]/3$ must be an upper bound on the principal's expected payoff from p. ■

Claim 7 implies that, in order for a mechanism to obtain an ex-ante payoff for the principal at least as high as $\overline{U}$, the probability that the mechanism results in x_4 conditional on (ω_1, θ_1) must be strictly positive. This is because obtaining such a high payoff for the principal requires generating positive surplus conditional on agent type θ_1 as well as preventing the information rent for agent type θ_2 from becoming too large, which involves deterring the type θ_2 agent from mimicking the type θ_1 agent by inducing x_4 with strictly positive probability conditional on (ω_1, θ_1).

Claim 8. *In the environment given in Example 1, for every outcome $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$ induced by incentive compatible and individually rational mechanisms that induce at most 3 signals with strictly positive probability and give the principal an expected payoff at least $\overline{U}$, the principal's expected payoff from p must be weakly less than $\overline{S} - 18.8085(\mathbb{P}_p[x_4|\omega_1, \theta_1] - .0006)$.*

Proof. We first obtain some properties that must hold for all outcomes that are induced by incentive compatible and individually rational mechanisms that give the principal an expected payoff at least $\overline{U}$. Consider an arbitrary such outcome $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$. We will argue that, for every $i \in \{1, 2, 3\}$, $\min\{\mathbb{P}_p[x_i|\omega_i, \theta_1], \mathbb{P}_p[x_i|\omega_i, \theta_2]\} \geq .976$ must hold. Observe that, since the principal's expected utility from p is weakly less than the expected surplus generated under p, $u(\omega_1, \theta_1, x_1) + v(\omega_1, \theta_1, x_1) = 1$, and $u(\omega_1, \theta_1, x) + v(\omega_1, \theta_1, x) \leq 0$ for all $x \in X \setminus \{x_1\}$, $\mathbb{P}_p[x_1|\omega_1, \theta_1] \geq 1 - (\overline{S} - \overline{U})/(\lambda_\Omega[\omega_1]\lambda_\Theta[\theta_1]) = 4901/5009 > .976$ is necessary for $(S, \sigma, \mathcal{M})$ to give the principal an expected payoff at least $\overline{U}$. Similarly, since $u(\omega_1, \theta_2, x_1) + v(\omega_1, \theta_2, x_1) = 10$, and $u(\omega_1, \theta_2, x) + v(\omega_1, \theta_2, x) \leq 0$ for all $x \in X \setminus \{x_1\}$, $\mathbb{P}_p[x_1|\omega_1, \theta_2] \geq 1 - (\overline{S} - \overline{U})/(10\lambda_\Omega[\omega_1]\lambda_\Theta[\theta_2]) = 24829/25045 > .976$ must also hold for $(S, \sigma, \mathcal{M})$ to give the principal an expected payoff at least $\overline{U}$. Almost identical arguments show that $\mathbb{P}_p[x_2|\omega_2, \theta_1] \geq .976$ and $\mathbb{P}_p[x_2|\omega_2, \theta_2] \geq .976$ would also have to hold. Finally, since $u(\omega_3, \theta_1, x_3) + v(\omega_3, \theta_1, x_3) = .9$, $u(\omega_3, \theta_2, x_3) + v(\omega_3, \theta_2, x_3) = 9.9$, and $u(\omega_3, \theta, x) + v(\omega_3, \theta, x) \leq 0$ for all $(\theta, x) \in \Theta \times (X \setminus \{x_3\})$, it follows that $\mathbb{P}_p[x_3|\omega_3, \theta_1] \geq 1 - (\overline{S} - \overline{U})/(.9\lambda_\Omega[\omega_3]\lambda_\Theta[\theta_1]) = 4889/5009 > .976$ and $\mathbb{P}_p[x_3|\omega_1, \theta_2] \geq 1 - (\overline{S} - \overline{U})/(9.9\lambda_\Omega[\omega_3]\lambda_\Theta[\theta_2]) = 54619/55099 > .976$ would have to hold for $(S, \sigma, \mathcal{M})$ to give the principal an expected payoff at least $\overline{U}$. Hence, for every $i \in \{1, 2, 3\}$, $\min\{\mathbb{P}_p[x_i|\omega_i, \theta_1], \mathbb{P}_p[x_i|\omega_i, \theta_2]\} \geq .976$ must hold.

Consider now an arbitrary mechanism $(S, \sigma, \mathcal{M})$ that induces at most 3 signals with strictly positive probability and gives the principal an expected payoff at least $\overline{U}$. Let $p \in \Delta(\Omega \times \Theta \times X \times \mathbb{R})$ denote the outcome induced by $(S, \sigma, \mathcal{M})$. By the previously established properties, there must be three signals $s_1, s_2, s_3 \in S$ such that, for all $i \in \{1, 2, 3\}$, $\sigma(\omega_i)[s_i] \geq 952/976$ and $\min\{\mathcal{M}(s_i, \theta_1)[x_i], \mathcal{M}(s_i, \theta_2)[x_i]\} \geq .976$. Observe that, for every $i \in \{1, 2, 3\}$, the posterior belief of the principal upon observing s_i with information structure (S, σ) must put probability at least .952 on ω_i. Additionally, since $\sigma(\omega_1)[s_2]\mathcal{M}(s_2, \theta_1)[x_4] + \sigma(\omega_1)[s_3]\mathcal{M}(s_3, \theta_1)[x_4] \leq (24/976)(.024) < .0006$, it must be that $\mathcal{M}(s_1, \theta_1)[x_4] > \mathbb{P}_p[x_4|\omega_1, \theta_1] - .0006$. For all $i \in \{1, 2, 3\}$ and $x \in X$, let $\lambda_{(S,\sigma)}(s_i) \in \Delta(\Omega)$ denote the posterior belief of the principal upon observing s_i with information structure (S, σ) and $u(s_i, x) = \mathbb{E}_{\omega \sim \lambda_{(S,\sigma)}(s_i)}[u(\omega, \theta_1, x)]$ denote the expected value of the principal from allocation x given agent type θ_1 conditional upon observing s_i. (Note that, for all $i \in \{1, 2, 3\}$ and $x \in X$, it is also the case that $u(s_i, x)$ equals the expected value of the principal from allocation x given agent type θ_2 conditional upon observing s_i.) Since $\lambda_{(S,\sigma)}(s_1)[\omega_1], \lambda_{(S,\sigma)}(s_2)[\omega_2] \geq .952$, it follows that $u(s_1, x_1) - u(s_3, x_1) = u(s_1, x_3) - u(s_3, x_3) \in [.0328, .1576]$, $u(s_3, x_2) - u(s_1, x_2), u(s_3, x_4) - u(s_1, x_4) \in [.004, .296]$, and $u(s_1, x_5) - u(s_3, x_5) = 0$. Incentive compatibility requires that

$$\sum_{x \in X} \left(\frac{4}{5}\mathcal{M}(s_1, \theta_1)[x] + \frac{1}{5}\mathcal{M}(s_1, \theta_2)[x] \right) u(s_1, x) + \frac{4}{5}\mathbb{E}_{\mathcal{M}(s_1,\theta_1)}[t] + \frac{1}{5}\mathbb{E}_{\mathcal{M}(s_1,\theta_2)}[t]$$
$$\geq \sum_{x \in X} \left(\frac{4}{5}\mathcal{M}(s_3, \theta_1)[x] + \frac{1}{5}\mathcal{M}(s_3, \theta_2)[x] \right) u(s_1, x) + \frac{4}{5}\mathbb{E}_{\mathcal{M}(s_3,\theta_1)}[t] + \frac{1}{5}\mathbb{E}_{\mathcal{M}(s_3,\theta_2)}[t]$$

and

$$\sum_{x \in X} \left(\frac{4}{5}\mathcal{M}(s_3, \theta_1)[x] + \frac{1}{5}\mathcal{M}(s_3, \theta_2)[x] \right) u(s_3, x) + \frac{4}{5}\mathbb{E}_{\mathcal{M}(s_3,\theta_1)}[t] + \frac{1}{5}\mathbb{E}_{\mathcal{M}(s_3,\theta_2)}[t]$$
$$\geq \sum_{x \in X} \left(\frac{4}{5}\mathcal{M}(s_1, \theta_1)[x] + \frac{1}{5}\mathcal{M}(s_1, \theta_2)[x] \right) u(s_3, x) + \frac{4}{5}\mathbb{E}_{\mathcal{M}(s_1,\theta_1)}[t] + \frac{1}{5}\mathbb{E}_{\mathcal{M}(s_1,\theta_2)}[t].$$

Combining these inequalities along with the facts that $u(s_1, x_1) - u(s_3, x_1) = u(s_1, x_3) - u(s_3, x_3) \in [.0328, .1576]$, $u(s_3, x_2) - u(s_1, x_2), u(s_3, x_4) - u(s_1, x_4) \in [.004, .296]$, and

$\mathcal{M}(s_1,\theta_1)[x_2], \mathcal{M}(s_1,\theta_2)[x_2], \mathcal{M}(s_1,\theta_1)[x_4], \mathcal{M}(s_1,\theta_2)[x_4] \geq 0$ gives

$$\begin{aligned}
&.4536\left(\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_2]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_2]+\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_4]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_4]\right)\\
&+.1576\left(\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_5]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_5]\right)\\
\geq&.0328\left(\frac{4}{5}\mathcal{M}(s_1,\theta_1)[x_4]\right).
\end{aligned}$$

Hence, since $.0328(4/5)/.4536 > .057848$, it follows that

$$\begin{aligned}
&\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_2]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_2]+\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_4]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_4]\\
&+\frac{4}{5}\mathcal{M}(s_3,\theta_1)[x_5]+\frac{1}{5}\mathcal{M}(s_3,\theta_2)[x_5]\\
>&.057848\mathcal{M}(s_1,\theta_1)[x_4]
\end{aligned}$$

As $\sigma(\omega_3)[s_3] \geq 952/976$ and $(952/976)(.057848) > .056425508$, it thus follows that

$$\mathbb{P}_p[\{x_2,x_4,x_5\}|\omega_3] > .056425508\mathcal{M}(s_1,\theta_1)[x_4].$$

By this and the facts that the principal's expected utility from p is weakly less than the expected surplus generated under p, $\max_{x\in X} u(\omega,\theta,x)+v(\omega,\theta,x) > 0$ for every $(\omega,\theta) \in \Omega\times\Theta$, $v(\omega_3,\theta_1,x_2), v(\omega_3,\theta_2,x_2), v(\omega_3,\theta_1,x_4), v(\omega_3,\theta_2,x_4), v(\omega_3,\theta_1,x_5), v(\omega_3,\theta_2,x_5) \leq -1000$, and $(1/3)(.056425508)(1000) > 18.8085$, it follows that principal's expected utility from p, denoted by $U(p)$, must satisfy $U(p) \leq \overline{S} - 18.8085\mathcal{M}(s_1,\theta_1)[x_4]$. Since $\mathcal{M}(s_1,\theta_1)[x_4] > \mathbb{P}_p[x_4|\omega_1,\theta_1] - .0006$, we conclude that $U(p) \leq \overline{S} - 18.8085(\mathbb{P}_p[x_4|\omega_1,\theta_1] - .0006)$. ■

Claim 9. *In the environment given in Example 1, every incentive compatible and individually rational mechanism that induces at most 3 signals with strictly positive probability must give the principal an expected payoff strictly less than $\overline{U}$.*

Proof. Consider an arbitrary mechanism that induces at most 3 signals with strictly positive probability. Let $p \in \Delta(\Omega\times\Theta\times X\times\mathbb{R})$ denote the corresponding outcome. Since $299/150 + 2000\mathbb{P}_p[x_4|\omega_1,\theta_1]/3 \leq 133/50 < \overline{U}$ if $\mathbb{P}_p[x_4|\omega_1,\theta_1] \leq 1/1000$, it follows by Claim 7 that the mechanism must give the principal an expected payoff strictly less than $\overline{U}$ if $\mathbb{P}_p[x_4|\omega_1,\theta_1] \leq 1/1000$. Moreover, since $\overline{S} - 18.8085(\mathbb{P}_p[x_4|\omega_1,\theta_1] - .0006) < 2.76 < \overline{U}$ if $\mathbb{P}_p[x_4|\omega_1,\theta_1] \geq 1/1000$, it follows by Claim 8 that the mechanism must give the principal an expected payoff strictly less than $\overline{U}$ if $\mathbb{P}_p[x_4|\omega_1,\theta_1] \geq 1/1000$. Hence every incentive compatible and individually rational mechanism that induces at most 3 signals with strictly positive probability must give the principal an expected payoff strictly less than $\overline{U}$. ■

OA 2.3 Signal Sets in Non-Quasilinear Environments

The following example provides an environment in which $|\Omega| = 3$, $|\Theta| = 1$, and every optimal mechanism $M = (S, \sigma, \mathbf{x})$ must satisfy $|S| \geq 4$.

Example 2. *The feature set $\Omega = \{\omega_1, \omega_2, \omega_3\}$ has precisely 3 elements, the agent's type set $\Theta = \{\theta\}$ has precisely 1 element, and the allocation set $X = \{x_1, x_2, x_3, x_4\}$ has precisely 4 elements. The prior distribution over the features $F \in \Delta(\Omega)$ is such that $F[\omega_1] = F[\omega_2] = F[\omega_3] = 1/3$. The payoffs to the principal and the agent from the various allocations are given in the following table. (The table is such that, for each $(\omega, \theta, x) \in \Omega \times \Theta \times X$, the first number in the corresponding pair of numbers gives the principal's payoff while the second number gives the agent's payoff.)*

ω_1	x_1	x_2	x_3	x_4
	$1, -1$	$-1, -1$	$1, -100$	$0, 299$

ω_2	x_1	x_2	x_3	x_4
	$-10, -1$	$1, -1$	$-1, -1$	$0, -1$

ω_3	x_1	x_2	x_3	x_4
	$-10, -10000$	$-1, -10000$	$1, -1$	$-1, -10000$

Table 7: The payoffs for Example 2.

Let $\overline{U} = 149/150$. We will show that there exists a mechanism with signal size 4 that achieves expected payoff $\overline{U}$, and any mechanism with signal size at most 3 has expected payoff strictly less than $\overline{U}$.

Intuitively, for any feature ω_i, the principal's first best allocation is x_i. However, such allocation rule is not individually rational for the agent. In order to increase the agent's utility for participating in the mechanism, allocation x_4 must be chosen with sufficiently high probability when the feature is ω_1. This can be done by pooling feature ω_1 and ω_2 with small probabilities into the fourth signal and allocate x_4 for that type. Such mechanism satisfies individual rationality constraint for the agent without creating huge distortions on the first best allocation for the principal. This construction is illustrated in Claim 10 to show that the principal can achieve expected payoff $\overline{U}$ with 4 signals.

Claim 10. *In the environment given in Example 2, there is a feasible mechanism in which exactly 4 signals that gives an expected payoff of $\overline{U}$ to the principal.*

Proof. Consider a signal set $S = \{s_1, s_2, s_3, s_4\}$ with precisely 4 signals and the mechanism

$(S, \sigma, \mathbf{x})$ in which $\sigma : \Omega \to \Delta(S)$ is given by

$$\sigma(\omega) = \begin{cases} \frac{99}{100}\delta_{s_1} + \frac{1}{100}\delta_{s_4} & \text{if } \omega = \omega_1, \\ \frac{99}{100}\delta_{s_2} + \frac{1}{100}\delta_{s_4} & \text{if } \omega = \omega_2, \\ \delta_{s_3} & \text{if } \omega = \omega_3, \end{cases}$$

for all $\omega \in \Omega$ and the allocation rule $\mathbf{x} : S \times \Theta \to \Delta(X)$ is given by

$$\mathbf{x}(s, \theta) = \begin{cases} \delta_{x_1} & \text{if } s = s_1, \\ \delta_{x_2} & \text{if } s = s_2, \\ \delta_{x_3} & \text{if } s = s_3, \\ \delta_{x_4} & \text{if } s = s_4, \end{cases}$$

for all $(s, \theta) \in S \times \Theta$. Observe that $\overline{U}$ is the principal's expected payoff from $(S, \sigma, \mathbf{x})$. Additionally, with information structure (S, σ), δ_{ω_1} is the posterior belief of the principal upon observing s_1, δ_{ω_2} is the posterior belief of the principal upon observing s_2, δ_{ω_3} is the posterior belief of the principal upon observing s_3, and $(1/2)\delta_{\omega_1} + (1/2)\delta_{\omega_2}$ is the posterior belief of the principal upon observing s_4. Thus, $(S, \sigma, \mathbf{x})$ is incentive compatible for the principal. Moreover, the expected payoff of the agent is 0, so $(S, \sigma, \mathbf{x})$ is individually rational for the agent. ■

In contrast, when there is only 3 signals, in order to attain expected payoff $\overline{U}$, the allocation of the mechanism must be close to efficient. In particular, signals must be almost revealing for the features, i.e., each signal s_i would correspond to feature ω_i such that signal s_i is generated with sufficiently high probability conditional on feature ω_i for all $i \in \{1, 2, 3\}$. Moreover, in order for the agent's individual rationality constraint to be satisfied, allocation x_4 must be chosen with sufficiently high probability given signal s_1 and allocation x_3 must be chosen with sufficiently high probability given signal s_3. However, the creates an incentive for signal s_1 to misreport as s_3 since conditional on s_1, the principal would believe feature ω_1 happens with high probability, and allocation x_3 is favorable to the principal given this belief. Thus, no mechanism can implemented expected payoff of $\overline{U}$ for the principal with only 3 signals. We formalize this intuition in Claim 11.

Claim 11. *In the environment given in Example 2, for any feasible mechanism $M = (S, \sigma, \mathbf{x})$ with $|S| \leq 3$, the expected payoff of mechanism M is strictly less than $\overline{U}$.*

Combining Claim 10 and 11, the optimal mechanism must have signal space at least 4 in the environment given in Example 2.

Proof of Claim 11. Suppose towards a contradiction that there exists a feasible mechanism $M = (S, \sigma, \mathbf{x})$ with $|S| \leq 3$ such that the expected payoff of mechanism M is at least $\overline{U}$. Let $\mathbf{Pr}_M[x \,|\, \omega]$ be the probability allocation x is chosen when the true feature is ω.

We first show that the following statements must hold: (1) $\mathbf{Pr}_M[x_4 \,|\, \omega_1] \geq 1/100$; (2) $\mathbf{Pr}_M[x_1 \,|\, \omega_1], \mathbf{Pr}_M[x_2 \,|\, \omega_2], \mathbf{Pr}_M[x_3 \,|\, \omega_3] > .949696$; and (3) $\mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3] \leq .00121$.

(1) Since the mechanism is individually rational for the agent, and $V(\omega_1, \theta, x_4) = 299$ and $V(\omega, \theta, x) \leq -1$ for all $(\omega, x) \in (\Omega \times X) \setminus \{(\omega_1, x_4)\}$, we must have $299\mathbf{Pr}_M[x_4 \,|\, \omega_1]/3 - (1 - \mathbf{Pr}_M[x_4 \,|\, \omega_1]/3) \geq 0$, which implies that $\mathbf{Pr}_M[x_4 \,|\, \omega_1] \geq 1/100$.

(2) In order for the principal's expected utility to be at least $\bar{U}$, the probability each feature chooses its first best allocation must be sufficiently large. Specifically, since $U(\omega_1, \theta, x_1) = U(\omega_1, \theta, x_3) = 1$ and $U(\omega_1, \theta, x_2), U(\omega_1, \theta, x_4) \leq 0$, we have that $\mathbf{Pr}_M[\{x_1, x_3\} \,|\, \omega_1] \geq 1 - (1 - \overline{U})/F(\omega_1) = .98$. Similarly, $\mathbf{Pr}_M[x_2 \,|\, \omega_2] \geq .98 > .949696$ and $\mathbf{Pr}_M[x_3 \,|\, \omega_3] \geq .98 > .949696$ must hold.

Additionally, in order for mechanism M to be individually rational for the agent, the agent's expected utility in mechanism M is at least

$$\frac{1}{3}(-\mathbf{Pr}_M[x_1 \,|\, \omega_1] - 100(.98 - \mathbf{Pr}_M[x_1 \,|\, \omega_1]) + 299(1 - .98)) - \frac{2}{3} \geq 0$$

since $V(\omega_1, \theta, x_1) = -1$, $V(\omega_1, \theta, x_3) = -100$, $V(\omega_1, \theta, x_4) = 299$, $V(\omega, \theta, x) \leq -1$ for all $(\omega, x) \in \{\omega_2, \omega_3\} \times X$, $\mathbf{Pr}_M[\{x_1, x_3\} \,|\, \omega_1] \geq .98$, and $\mathbf{Pr}_M[x_4 \,|\, \omega_1] \geq 1/100$. This implies that $\mathbf{Pr}_M[x_1 \,|\, \omega_1] \geq 1567/1650 > .949696$.

(3) Finally, since $V(\omega_1, \theta, x_4) = 299$, $V(\omega_3, \theta, x_1) = V(\omega_3, \theta, x_2) = V(\omega_3, \theta, x_4) = -10000$, $V(\omega, \theta, x) \leq -1$ for all $(\omega, x) \in (\Omega \times X) \setminus \{(\omega_1, x_4)\}$, and $\mathbf{Pr}_M[x_1 \,|\, \omega_1] > .949696$, the individual rationality constraint of the agent implies that the agent's expected utility in mechanism M is at least

$$\begin{aligned}&\frac{1}{3}(-.949696 + (.050304)(299)) - \frac{1}{3} \\ &\quad + \frac{1}{3}(-10000\mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3] - (1 - \mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3])) \geq 0,\end{aligned}$$

which further implies that $\mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3] \leq 229/189375 < .00121$.

Since the signal space S only has cardinality 3, $\mathbf{Pr}_M[x_i \,|\, \omega_i] \geq .949696$ for all $i \in \{1, 2, 3\}$ implies that $\mathbf{x}(s_i, \theta)[x_i] \geq .949696$ and $\sigma(s_i \,|\, \omega_i) + (1 - .949696)(1 - \sigma(s_i \,|\, \omega_i)) \geq .949696$ for all $i \in \{1, 2, 3\}$. The latter further implies that $\sigma(s_i \,|\, \omega_i) \geq .947$ for all $i \in \{1, 2, 3\}$. Therefore, for any $i \in \{1, 2, 3\}$, the posterior belief of the principal upon observing s_i must

put probability at least .899 on ω_i. Additionally, since $\sigma(s_2 \,|\, \omega_1) \cdot \mathbf{x}(s_2, \theta)[x_4] + \sigma(s_3 \,|\, \omega_1) \cdot \mathbf{x}(s_3, \theta)[x_4] \leq (.053)(.050304) < .00266$, it must be that

$$\mathbf{x}(s_1, \theta)[x_4] \geq \mathbf{Pr}_M[x_4 \,|\, \omega_1] - \sigma(s_2 \,|\, \omega_1) \cdot \mathbf{x}(s_2, \theta)[x_4] - \sigma(s_3 \,|\, \omega_1) \cdot \mathbf{x}(s_3, \theta)[x_4] > .00734.$$

For all $i, j \in \{1, 2, 3\}$, let $\lambda_{s_i} \in \Delta(\Omega)$ denote the posterior belief of the principal upon observing s_i given mechanism M, and $\mathcal{U}(s_i, s_j) = \mathbf{E}_{x \sim \mathbf{x}(s_j, \theta)} \Big[\mathbf{E}_{\omega \sim \lambda_{s_i}} [U(\omega, \theta, x)] \Big]$ denote the interim payoff of the principal from reporting s_j to the mechanism after observing s_i. By simple algebraic calculation,

$$\begin{aligned}
\mathcal{U}(s_1, s_1) \leq & \lambda_{s_1}(\omega_1)(1 - \mathbf{x}(s_1, \theta)[x_4]) + (1 - \lambda_{s_1}(\omega_1))(-10\mathbf{x}(s_1, \theta)[x_1] + 1 - \mathbf{x}(s_1, \theta)[x_1]), \\
\mathcal{U}(s_1, s_3) \geq & \lambda_{s_1}(\omega_1)(\mathbf{x}(s_3, \theta)[x_3] - \mathbf{x}(s_3, \theta)[x_2]) \\
& + (1 - \lambda_{s_1}(\omega_1))(-10(1 - \mathbf{x}(s_3, \theta)[x_3]) - \mathbf{x}(s_3, \theta)[x_3]).
\end{aligned}$$

Combining these inequalities with the fact that $\mathcal{U}(s_1, s_1) \geq \mathcal{U}(s_1, s_3)$ would have to hold gives

$$\begin{aligned}
& \lambda_{s_1}(\omega_1)(1 - \mathbf{x}(s_1, \theta)[x_4] - \mathbf{x}(s_3, \theta)[x_3] + \mathbf{x}(s_3, \theta)[x_2]) \\
\geq & (1 - \lambda_{s_1}(\omega_1))(11\mathbf{x}(s_1, \theta)[x_1] + 9\mathbf{x}(s_3, \theta)[x_3] - 11) \geq 0.
\end{aligned}$$

Since $\mathbf{x}(s_3, \theta)[x_2] \leq 1 - \mathbf{x}(s_3, \theta)[x_3]$, the above inequality implies that $1 - \mathbf{x}(s_3, \theta)[x_3] \geq \frac{1}{2}\mathbf{x}(s_1, \theta)[x_4] > .00367$, which further implies that

$$\mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3] \geq \sigma(s_3 \,|\, \omega_3)(1 - \mathbf{x}(s_3, \theta)[x_3]) > (.947)(.00367) > .00347,$$

which contradicts our previous conclusion that $\mathbf{Pr}_M[\{x_1, x_2, x_4\} \,|\, \omega_3] \leq .00121$. ■

OA 3 Fully Uninformative Informative Structure

Proposition 6 implies that mechanisms with fully revealing information structures are strictly suboptimal, a natural question to ask is whether mechanisms with fully uninformative information structures can be strictly optimal for the principal. In particular, we say an information structure (S, σ) is *fully uninformative* if $\sigma(\omega) = \sigma(\omega')$ for all $\omega, \omega' \in \Omega$.

In the following proposition, we show that there is no strict benefit for the principal to stay fully ignorant. Specifically, if there exists an optimal mechanism in which the information structure is not fully uninformative, for any information structure (S, σ), there exists another optimal mechanism in which the information structure is (S, σ). The high

level intuition is that the principal can design mechanisms that ignore the report of the principal, and hence simulating the outcomes of the mechanism with fully uninformative information structures without violating the incentive constraints for the principal.

Proposition 8. *If there is an optimal mechanism in which the information structure is fully uninformative, then, for every information structure* (S, σ)*, there is an optimal mechanism in which the information structure is* (S, σ)*.*

Proof. Suppose that there is an optimal mechanism $M_u = (S_u, \sigma_u, \mathbf{x}_u)$ with fully uninformative information structure (S_u, σ_u). For any information structure (S, σ), let mechanism $M = (S, \sigma, \mathbf{x})$ be the mechanism with allocation rule

$$\mathbf{x}(\omega, \theta) = \mathbf{E}_{\omega' \sim F_\Theta} \left[\mathbf{E}_{s_u \sim \sigma_u(\omega')} [\mathbf{x}_u(s_u, \theta)] \right].$$

By the construction of mechanism M, it induces the same distribution over outcomes and hence the same expected payoff for the principal compared to mechanism M_u. Moreover, this further implies that mechanism M is also incentive compatible and individually rational for the agent since mechanism M_u is incentive compatible and individually rational for the agent. Finally, mechanism M is incentive compatible for the principal since the allocation does not depend on the principal's report. ■