THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques - Toulouse School of Economics**

Présentée et soutenue par

ZHU Shuguang

le mardi 19 juin 2018

Three essays on mechanism design, information design and collective decision-making

---

Discipline : **Sciences Economiques**

Spécialité : **Droit Privé et Sciences Criminelles**

Unité de recherche : **TSE-R**

Directeur de thèse : Monsieur Takuro YAMASHITA, Professeur, Université Toulouse 1 Capitole

JURY

Rapporteurs Monsieur João CORREIA DA SILVA, Professeur, University of Porto
Monsieur Takakazu HONRYO, Professeur, University of Mannheim

Suffragants Monsieur Renato GOMES, Professeur, Université Toulouse 1 Capitole
Monsieur Takuro YAMASHITA, Professeur, Université Toulouse 1 Capitole

Summary

This thesis investigates several topics in Microeconomic Theory, with a focus on incorporating information control into mechanism design, checking the robustness of mechanisms, and providing a foundation for inconsistent collective decision-making. This work helps to optimize information transmission and acquisition in organizational communications, advertisement and policy design. It also sheds light on how inconsistent group decisions derive from heterogeneity in group members, and proposes ways to restore efficiency. The thesis consists of three chapters, each of which is self-contained and can be read separately.

The first chapter studies a mechanism design environment where the principal has control over the agents' information about a payoff-relevant state. The principal commits to an information disclosure policy where each agent observes a private signal, while the principal directly observes neither the true state nor the signal profile. Examples include (1) assessing whether a new product matches consumers' preferences through their feedback on sample product trials, and (2) gathering intelligence by authorizing investigators to collect various aspects of information. I establish optimality of individually uninformative and aggregately revealing disclosure policy, where (i) each agent obtains no new information about the state after observing any realization of his own signal, but (ii) the principal can nevertheless infer the true state from the agents' reports about their signals. Furthermore, this optimal disclosure policy admits simple and intuitive implementation (such as certain types of blinded experiments, or restrictions on access to certain information) under additional assumptions. If attention is restricted to linear settings, I characterize a class of environments (including those satisfying the standard regularity conditions in mechanism design) where an equivalence result holds between private disclosure and public disclosure.

The second chapter, co-authored with Takuro Yamashita, is motivated by Chung and Ely (2007), who establish maxmin and Bayesian foundations for dominant-strategy mechanisms in private-value auction environments. We first show that similar foundation results for ex post mechanisms hold true even with interdependent values if the interdependence is only cardinal. Conversely, if the environment exhibits ordinal interdependence, which is typically the case with multi-dimensional environments, then in general, ex post mechanisms do not have foundation. That is, there exists a non-ex-post mechanism that achieves strictly higher expected revenue than the optimal ex post mechanism, regardless of the agents' high-order beliefs.

The third chapter shows that dynamic inconsistency in collective decision-making can derive from heterogeneity in group members' outside options (i.e. opportunity costs that individuals have to pay in order to join the group), even if individuals share the same exponentially discounting time preference. This model of endogenous dynamic inconsistency facilitates the analysis of welfare consequences, since time-consistent individual preferences allow for a well-defined measurement of social welfare. We further characterize the optimal Bayesian-persuasion information disclosure policy, which takes the form of upper revealing rules, to alleviate the welfare distortion caused by inconsistent collective decisions. Our framework proves to be highly adaptable to various contexts, including provision of public facilities and assignment on team work.

Résumé

Cette thèse étudie plusieurs sujets dans la théorie microéconomique, en mettant l'accent sur l'intégration du contrôle de l'information dans la conception des mécanismes, la vérification de la robustesse des mécanismes et la création d'une base pour une prise de décision collective incohérente. Ce travail permet d'optimiser la transmission et l'acquisition de l'information dans les communications organisationnelles, la publicité et la conception de politiques. Il met également en lumière la façon dont les décisions de groupe inconsistantes découlent de l'hétérogénéité des membres du groupe et propose des moyens de restaurer l'efficacité. La thèse comprend trois chapitres, chacun étant autonome et pouvant être lu séparément.

Le premier chapitre étudie un environnement de conception de mécanisme dans lequel le principal a le contrôle sur les informations des agents concernant un état pertinent. Le principal s'engage à une politique de divulgation d'informations où chaque agent observe un signal privé, tandis que le principal n'observe directement ni l'état vrai ni le profil du signal. Les exemples incluent (1) l'évaluation si un nouveau produit correspond aux préférences des consommateurs grâce à leurs commentaires sur les essais de produits échantillon, et (2) la collecte de renseignements en autorisant les enquêteurs à recueillir divers aspects de l'information. J'établis l'optimalité d'une politique de divulgation individuellement non informative et révélatrice, où (i) chaque agent n'obtient aucune nouvelle information sur l'état après avoir observé la réalisation de son propre signal, mais (ii) le principal peut néanmoins déduire l'état réel des rapports des agents sur leurs signaux. En outre, cette politique de divulgation optimale admet une mise en œuvre simple et intuitive (comme certains types d'expériences en aveugle, ou des restrictions sur l'accès à certaines informations) sous des hypothèses supplémentaires. Si l'attention est limitée aux paramètres linéaires, je caractérise une classe d'environnements (y compris ceux qui satisfont aux conditions de régularité standard dans la conception des mécanismes) où un résultat d'équivalence est maintenu entre la divulgation privée et la divulgation publique.

Le deuxième chapitre, co-écrit avec Takuro Yamashita, est motivé par Chung et Ely (2007), qui établissent les fondements maxmin et bayésien des mécanismes de stratégie dominante dans les environnements d'enchères à valeur privée. Nous montrons d'abord que les résultats de fondation similaires pour les mécanismes ex post restent vrais même avec des valeurs interdépendantes si l'interdépendance n'est que cardinale. Inversement, si l'environnement présente une interdépendance ordinale, ce qui est typiquement le cas avec les environnements multidimensionnels, alors en général, les mécanismes ex post n'ont pas de fondement. C'est-à-dire qu'il existe un mécanisme non ex post qui réalise des recettes attendues strictement plus élevées que le mécanisme ex post optimal, quelles que soient les croyances élevées des agents.

Le troisième chapitre montre que l'incohérence dynamique dans la prise de décision collective peut provenir de l'hétérogénéité des options extérieures des membres du groupe (c.-à-d. Coûts d'opportunité que les individus doivent payer pour rejoindre le groupe) même si les individus partagent le même temps exponentiel préférence. Ce modèle d'incohérence dynamique endogène facilite l'analyse des conséquences sur le bien-être, puisque les préférences individuelles en fonction du temps permettent une mesure bien définie du bien-être social. Nous caractérisons en outre la politique de divulgation d'informations bayésienne-persuasion optimale, qui prend la forme de règles révélatrices supérieures, pour atténuer la distorsion du bien-être causée par des décisions collectives incohérentes. Notre cadre s'avère très adaptable à divers contextes, tels que la fourniture d'équipements publics et l'affectation au travail d'équipe.

Acknowledgments

First of all, I would like to show my deepest gratitude to my supervisor, Takuro Yamashita, a respectable, responsible and resourceful economist, who has provided me with inspiration and guidance at every stage of my research work. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. The unforgettable experience of working and teaching with my supervisor will become the treasure of my life which I shall definitely benefit from in my future career as a researcher.

I would like to thank my parents for their unconditional love and support. I am also grateful to my wife, Xiaoling, who has provided me with moral and emotional support along the way, without which I could not have overcome the difficulties of long-distance relationship during my six years' study in France.

A very special gratitude goes out to Jacques Crémer, Johannes Hörner and Tilman Börgers for their extremely valuable comments and suggestions on this thesis. I'm also obliged to them for their recommendation letters and strong support, without which I might have had a much tougher time during the job market.

I would like to thank Toulouse School of Economics for its great research environment, with so many talented scholars and doctoral students. I would also like to thank the Jean-Jacques Laffont Foundation and Professor Takuro Yamashita's ERC Starting Grant for providing the funding for my study in Toulouse.

With a special mention to University of Mannheim and University of Michigan, where I was invited to present the papers included in the thesis. I was impressed by their ideal academic atmosphere, and have greatly benefited from comments and discussions with the professors in both Mannheim and Michigan.

And finally, last but by no means least, I would like to express my gratitude to many people I've talked and discussed with in Toulouse, or on other occasions outside of Toulouse, whose help on this thesis can never be exaggerated. Thank you so much Yu Awaya, Alessandro Bonatti, Gabriel Carroll, Alexandre de Cornière, João Correia-da-Silva, Daniel Garrett, Christian Gollier, Renato Gomes, Xintong Han, Christian Hellwig, Takakazu Honryo, Doh-Shin Jeon, Daniel Krähmer, Lan Lan, Tsz Kin Leung, Raphaël Levy, Ying Liang, Heng Liu, Qingmin Liu, Thomas Mariotti, David Miller, Andras Niedermayer, Di Pei, Jérôme Renault, Andrew Rhodes, Alex Smolin, Chengsi Wang, Chunan Wang, Gang Wang, and many others.

Thanks for all your encouragement!

Contents

Chapter 1

Private Disclosure with Multiple Agents[1]

Shuguang Zhu[2]

April 15, 2018

Abstract

We study a mechanism design environment where the principal has control over the agents' information about a payoff-relevant state. The principal commits to an information disclosure policy where each agent observes a private signal, while the principal directly observes neither the state nor the signal profile. We prove the optimality of *individually uninformative* and *aggregately revealing* disclosure policy, where (i) each agent obtains no new information about the state after observing any realization of his own signal, but (ii) the principal can infer the true state from the agents' reports about their signals. Furthermore, this optimal disclosure policy admits simple and intuitive implementation (such as certain type of blinded experiments) under additional assumptions. If attention is restricted to linear settings, we characterize a class of environments (including those satisfying the standard regularity conditions in mechanism design) where an equivalence result holds between private disclosure and public disclosure.

[1]I am indebted to Takuro Yamashita for his inspiration and guidance at every stage of my research work. I also thank, in alphabetical order, Yu Awaya, Tilman Börgers, Gabriel Carroll, Jacques Crémer, Daniel Garrett, Renato Gomes, Christian Hellwig, Johannes Hörner, Doh-Shin Jeon, Daniel Krähmer, Heng Liu, Qingmin Liu, Thomas Mariotti, David Miller, Jérôme Renault, Andrew Rhodes, Alex Smolin, Chunan Wang, and participants in TSE Student Workshop 2017, EWMES 2017 and SAEe 2017 for extremely valuable comments and suggestions on this project.

[2]Toulouse School of Economics, University of Toulouse Capitole, France. shuguang.zhu@tse-fr.eu

1.1 Introduction

We consider a broad range of situations where the principal (denoted by "She"; while "He" denotes the agent) has considerable control over agents' information about a payoff-relevant state, before she designs the mechanism. The principal can reveal some signals about the state to agents in the manner of private disclosure, where signals are individually chosen and secretly sent to each agent[3]; however, the principal cannot directly observe either the state or the signal profile.

In practice, the assumption that the principal designs a disclosure policy of something she cannot observe is well founded. Imagine a situation (due to Lewis and Sappington, 1994) where the seller lacks information about potential buyers' tastes for her new product. By offering product samples, the seller is able to control the amount of information about her product that is available to different buyers. After gathering feedback, the seller gains a better knowledge of whether the product characteristics match with buyers' preferences. Another example is about how an intelligence agency sends off investigators or spies to collect confidential information. By authorizing each person to investigate only a particular aspect, the intelligence agency effectively controls what and how precisely investigators learn about the targeted information.

We work in a general environment where the principal can implement stochastic outcomes, not necessarily with the help of transfers. Our model fits various classical settings, including private value, interdependent value, and a wide class of objectives. At the first stage, the principal can commit to any *information disclosure policy* which generates private signals to each agent. At the second stage, the principal can commit to any mechanism whose outcome depends on agents' reports about their private types and signals. The principal's problem is to design a *private disclosure mechanism*, consisting of the information disclosure policy and the associated mechanisms, so as to optimize her objective function. Thus, in terms of methodology, we are closer to the Bayesian persuasion model (e.g., Kamenica and Gentzkow, 2011; Kolotilin, Mylovanov, Zapechelnyuk, and Li, 2017; Yamashita, 2017) than to the cheap-talk model adopted in the informed-principal literature.[4] However, unlike the Bayesian persuasion

[3]Our private disclosure setting can easily accommodate the case of public disclosure, through sending perfectly correlated signals to all agents.

[4]See Myerson (1983), Maskin and Tirole (1990, 1992) and Mylovanov and Tröger (2012). In these papers, only cheap-talk signals are available for the principal during the information disclosure process. Thus, the following mechanism in equilibrium induced by each realization of the state – which is privately observed by the principal – must be the best choice for the principal among all feasible mechanisms given the agents' equilibrium strategies. This requirement imposes additional incentive compatible constraints on the principal side, which

model where the sender's payoff is exogenously given, in our paper the sender (who is the principal) has an endogenous payoff, which is defined as the expected payoff of the mechanism design after the disclosure process.

Our main finding is that, in the optimal private disclosure mechanism, individual agent gets *no new information* about the state after observing one's own signal, while the principal can infer the realization of the state from agents' reports. We name this property "*individually uninformative and aggregately revealing*" (or IUAR for short). We also show that through IUAR disclosure policy, the principal gets the same payoff as if she could directly observe the state and implement *state-varying* allocation rules. To the best of our knowledge, we are the first to prove the optimality of IUAR disclosure policy. A comparison between our result and other disclosure policies (e.g., Eső and Szentes, 2007; Bergemann and Pesendorfer, 2007) is made in the next session.

We encounter difficulties of multidimensional screening (e.g., Rochet and Choné, 1998; Haghpanah and Hartline, 2014; Carroll, 2017) when solving the principal's problem. This is because allocation rules at the second stage must provide enough incentives for agents to truthfully report both their types and signals. To circumvent this difficulty, we first consider a relaxed problem where the principal has *strong control* over the disclosure process in the sense that she can observe the whole signal profile.

In this relaxed problem, the principal can make allocation rules contingent on the information only if it is disclosed in the signal profile. Intuitively, the more accurate the signal is, the better she can align the outcome with the state, while the harder it becomes to make agents truthfully report their types. In other words, the principal faces a tradeoff between *flexibility* and *implementability* (e.g., Eső and Szentes, 2007; Yamashita, 2017). However, we find that the optimal private disclosure policy is (1) *individually uninformative*: conditional on observing any signal, each agent's posterior belief about the state coincides with the common prior; and (2) *aggregately revealing*: agents' signals are correlated in a particular way such that the true state can be pinned down by the whole signal profile (Proposition 1). In short, with strong control over the disclosure process, the principal is free from the tradeoff.

Next, we consider the original problem where the principal has *weak control* over the disclosure process in the sense that she has no access to the signal profile, as in Eső and Szentes (2007), Bergemann and Pesendorfer (2007) and Li and Shi (2017). It turns out that the principal can exploit the correlation among signals to elicit agents' truthful reports for free. More specifically, when there are four or more agents, without introducing any additional assumption, the optimal private disclosure policy with strong control can be implemented in

makes the problem intractable without restrictions on the environment.

weak-control settings based on the "$(N-1)$ majority rule" (Theorem 1). When there are three (or two) agents, by allowing the principal to punish all agents by a uniformly worst outcome (or to use arbitrary transfers), we can still implement the IUAR disclosure policy (Theorem 2, 3).

We point out that the optimality of IUAR disclosure policy is quite robust, thus can be applied to various contexts. Besides the general payoff environment, we also allow correlations among agents' types and the state. We can even relax, to a certain extent, the no-communication assumption among agents, which is a key assumption for private disclosure. In Section 1.6.3 we provide a variation of the IUAR disclosure policy that is immune to information sharing among any subset of agents (excluding a very limited number of agents) (Theorem 5).[5]

We also show that the IUAR disclosure policy can be implemented through a more practical way, called *sample-product approach* (Section 1.5). The basic process is to (i) conduct individual trials where each consumer is offered a randomized Choice Pair (similar to a blinded experiment), and (ii) collect consumers' reports about their preferred choices. In particular, two states are *distinguishable* if we can find two consumers and a Choice Pair, such that feedbacks coincide in one state but are different in the other state. We prove that the IUAR information structure can be built on consumers' feedbacks under a richness assumption on the sample product, which says any pair of states are distinguishable (Theorem 4).

It is worth noting that the optimality of IUAR disclosure policy does not exclude the possibility of some other disclosure policies to be optimal. Actually, the full disclosure policy (together with the allocation rule) in some papers is optimal only when it achieves the same expected payoff for the principal as in our optimal private disclosure mechanism. If we restrict our attention to linear settings with independent private information (as in, e.g., Eső and Szentes, 2007; Yamashita, 2017), then we can characterize a class of environments (including those satisfying the standard regularity conditions) where an equivalence result holds between private disclosure (which is IUAR at optimality) and public disclosure (which is full disclosure at optimality).

The rest of the paper is organized as follows. Section 1.2 reviews the related literature. Section 1.3 presents the setup. In Section 1.4, we first solve the optimal mechanism given by a relaxed problem where the principal has strong control over the disclosure process, and then prove that it is also implementable in the original problem. Section 1.5 provides a sample-product approach to implement the IUAR disclosure policy. Section 1.6 establishes

[5]In Appendix B.2.8, we show that the IUAR disclosure policy is robust to the presence of certain number of faulty agents in the sense of Eliaz (2002).

the equivalence result between public and private disclosure, and discusses the assumptions on the environment. Section 1.7 concludes and provides several applications. Omitted proofs can be found in Appendix A and the Supplemental Material (Appendix B).

1.2 Related Literature

The optimality of IUAR property appears to conflicts with the recurring finding of the optimality of full disclosure in literature. One reason is that only public signals are considered in these papers. In the investigation of winner's curse, Milgrom and Weber (1982), followed by Ottaviani and Prat (2001), discover the linkage principle, which suggests that the seller should commit to fully revealing her information that is affiliated with the buyers' valuations. Yamashita (2017) shows that it is optimal for a mechanism designer to disclose all the relevant information in a class of linear environments where the principal's and agents' information are affiliated. A common feature of these papers is that flexibility effect dominates implementability effect, resulting in the optimality of full disclosure. What makes our private disclosure mechanism distinct from these papers is that: a public signal can only induce a degenerate distribution of agents' posterior belief over the others' signals; but on receiving a private signal, each agent will form a posterior belief about the opponents' signals which is correlated with his posterior belief about the state. This correlation provides the principal more (or more precisely, full) flexibility to adjust the allocations to the true state.

Another reason that sets our result apart from the literature is that we don't impose ad hoc restrictions on the information structure available to the principal. Eső and Szentes (2007) consider an auction environment where an informed seller commits to a private disclosure policy before conducting an auction; moreover, they restrict attention to independently distributed private signals among the agents. They suggest that the seller should make available all her information to the buyers, for the sake of improving efficiency. However, this is because their independence assumption on agents' signals essentially excludes the possibility for the seller to exploit the correlation of private signals to achieve the aggregately revealing property.

Bergemann and Pesendorfer (2007) also consider an auction environment where the seller can jointly decide the accuracy of bidders' information about their valuation and the allocations. Their model is a special case of ours, where (i) agents' private information is only consisting of the signals sent by the principal, and (ii) conditional on one's signal, each agent's valuation is independent of the other agents' signals. (See Footnote 19.) They show that optimal information structures can be represented by monotone partitions. However, if we relax this restriction on the information structure and consider the general disclosure policy as in our setup, then the optimality of IUAR disclosure policy also applies to their auction environment.

Regarding the IUAR information structure, the most closely related work is Liu (2015), who defines the "*individually uninformative correlating device*" to characterize the correlations implicitly captured by partition models for incomplete information games.[6] Particularly, no individual updates the likelihood ratio between two arbitrary states upon any private observation generated from the correlating device. In fact, the IUAR disclosure policy satisfies this definition. Despite this common feature, the context Liu considers is distinct from ours. In Liu (2015), the individually uninformative information structure is exogenously given as an assumption, in order to decompose an arbitrary partition model into the conjunction of a non-redundant partition model (which shares the same set of hierarchies of beliefs), and a correlation device (which specifies the strategic correlations). While in our paper, we provide a rationale for using such information structure, because it is derived from the maximizing problem.

Börgers, Hernando-Veciana, and Krähmer (2013) define the complementarity (or substitutability) of two signals, which describe a phenomenon where one signal could become more (or less) valuable when the other signal becomes available. They show that in an auction environment, information disclosure does not increase the seller's expected revenue if signals are complements (since this will increase bidders' information advantage); while disclosing the information does no harm to the seller if signals are substitutes (because this can reduce bidders' information advantage). In our paper, the necessity of private signals is in line with their findings. Particularly, the signals privately observed by agents in the IUAR disclosure policy are essentially *complements* for each other, thus allowing agents to acquire information about the others' signals would undermine the implementation of the mechanism.

Krähmer (2017) studies information design in auctions with the assumptions: each bidder's information structure is drawn from a bidder-specific collection of random variables, where each random variable is a signal only informative about this bidder's valuation for the object; and the auctioneer can commit to any correlated distribution of the signal profiles, so that bidders' posterior valuations get correlated which allows full-rent extraction through the Crémer-McLean mechanism (where transfers are necessary). By including a fully-revealing signal in each agent's set of possible signals, and choosing a sequence of correlated distributions converging to the degenerate distribution where fully-revealing signals are selected for all bidders with probability 1, the auctioneer can *approximately* achieve the first-best outcome, and the disclosure policy is *almost aggregately revealing*.

[6] See Appendix B.2.2 for a discussion about how it is related to our paper.

1.3 The Model

1.3.1 Payoff environment

We consider an environment with a finite set $I = \{1,2,\ldots,N\}$ of risk-neutral agents, where $N \geq 2$, and a finite set $\mathscr{A}(\ni a)$ of social alternatives. Each agent $i \in I$ has a private type v_i drawn from a finite subset of d_i-dimensional Euclidean space, denoted by $V_i \subseteq \mathbb{R}^{d_i}$, which is independently distributed according to the probability measure $F_i(v_i)$. Let $V = \prod_{i=1}^N V_i$, and $F_V = \prod_{i=1}^N F_i$ be the joint distribution over V.

The information controlled by the risk-neutral principal – that is, the state of the world – is denoted by $\theta \in \Theta \subseteq \mathbb{R}^d$, which is endowed with a probability measure $F_0(\theta)$. Assume that $|\Theta| = T < \infty$,[7] then we write $F_0(\theta = \theta_t) = \alpha_t > 0$, for $t = 1,2,\ldots,T$, where $\sum_{t=1}^T \alpha_t = 1$. The principal and agents share a common prior for $(v,\theta) \in V \times \Theta$, where F_0 and $(F_i)_{i\in I}$ are mutually independently distributed. It's worth noting that such mutual independence assumption is just for ease of presentation. In Appendix B.2.3, we prove that our main result (Theorem 1) remains valid even if there exists certain correlation among $(F_0, F_1, \ldots, F_N)$.

The utility of each agent i is given by $u_i(a,v,\theta)$, and the principal's utility is given by $u_0(a,v,\theta)$. The principal can implement any lottery of the social alternatives, thus a feasible allocation $x \in \Delta(\mathscr{A})$ is a probability measure over $\mathscr{A}$. Agents and the principal evaluate allocation x according to the expected utility, which is given by

$$u_i(x,v,\theta) = \int_{a\in\mathscr{A}} u_i(a,v,\theta)dx(a), \text{ for } i = 0,1,\ldots,N.^8$$

1.3.2 Information disclosure policy

The signal space is denoted by $M = \prod_{i=1}^N M_i$, where each M_i collects all possible signals m_i that agent i can privately observe, which potentially enables the principal to induce different posterior beliefs over θ across agents.[9] The *information disclosure policy* is defined as (M, Ξ),

[7] See Subsection 1.6.2 for discussions about the case where we relax the finiteness assumptions on the set of states, as well as the set of social alternatives and agents' type spaces.

[8] Essentially, the critical assumption we need is the linearity of utility functions with respect to allocations, as well as the convexity of the set of feasible allocations.

[9] The possibility of inducing different posteriors in more practical contexts is studied in literature. For example, Appendix A of Anderson and Renault (2006) provides a model of advertising products with multiple characteristics. Because consumers have heterogeneous preferences on product characteristics, it is feasible for sellers to communicate different information about final payoffs to different consumers, by describing particular product characteristics.

where the measurable mapping $\Xi : \Theta \to \Delta(M)$ specifies the distribution of private signal profiles generated by the disclosure policy under each state of the world. Ξ and F_0 induce a joint distribution over $\Theta \times M$, denoted by Φ, such that for any $A \subseteq \Theta$ and $B \subseteq M$, we have

$$\int_{A\times B} d\Phi(\theta,m) = \int_{\theta\in A}\int_{m\in B} d\Xi_\theta(m)dF_0(\theta).$$

Let $\Lambda \in \Delta(M)$ be the marginal distribution of Φ over M. Similarly, we define $\Lambda_i \in \Delta(M_i)$ as the marginal distribution of Φ over M_i.

The conditional distribution over Θ given any $m \in M$ is denoted by $\Psi_m(\theta) \in \Delta(\Theta)$. From the standard results in probability theory,[10] $\Psi_m(\theta)$ is well-defined, provided that Θ is a complete, separable metric space (or in other words, a Polish space). Moreover, for any $m \in M$ and $A \subseteq \Theta$, we have

$$\int_{\theta\in A} d\Psi_m(\theta) = \frac{\int_{\theta\in A} d\Phi(\theta,m)}{\int_{\theta\in\Theta} d\Phi(\theta,m)}.$$

Similarly, for each i, we define $\Psi_{m_i} \in \Delta(\Theta \times M_{-i})$ as the conditional distribution over $\Theta \times M_{-i}$ given any $m_i \in M_i$.

1.3.3 Mechanism

The principal selects (without knowing the state of the world) the information disclosure policy, which generates a privately observed signal for each agent. We assume that the principal has "*weak control*" over the disclosure process in the sense that she cannot observe the realization of m. This setting can be rationalized by assuming that the principal hires a third party to conduct a series of individual experiments for each agent. Essentially, each agent i's private information includes the "exogenous" type v_i and the "endogenous" type m_i. By the revelation principle we can restrict attention to direct mechanisms, where messages from agents to the principal are drawn from $V \times M$. The *associated direct mechanism* consists of a tuple (V,M,x), where $x : V \times M \to \Delta(\mathscr{A})$ specifies the lottery of social alternatives that will be implemented by the principal on receiving agents' reports about their private information.

Each agent i with v_i after observing his own signal m_i will form a posterior belief Ψ_{m_i} about θ and the other agents' signals m_{-i}. When agent i reports $(\hat{v}_i,\hat{m}_i)$ and all the other agents report truthfully, the interim utility of agent i is defined as

$$U_i(\hat{v}_i,\hat{m}_i;v_i,m_i) = \int_{v_{-i}}\int_{\theta,m_{-i}} u_i\big(x(\hat{v}_i,v_{-i},\hat{m}_i,m_{-i}),v_i,v_{-i},\theta\big)d\Psi_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i}).$$

We consider Bayesian Nash equilibrium where truth-telling by all agents constitutes the equilibrium strategy profile. Since agents have multi-dimensional private information, $(\hat{v}_i,\hat{m}_i)$

[10] See Faden (1985), for example.

will be a misreport as long as $\hat{v}_i \neq v_i$ or $\hat{m}_i \neq m_i$. Thus, the mechanism has to satisfy agents' *Bayesian incentive compatibility* (BIC) constraints

$$U_i(v_i, m_i; v_i, m_i) \geq U_i(\hat{v}_i, \hat{m}_i; v_i, m_i), \quad \forall (v_i, m_i) \neq (\hat{v}_i, \hat{m}_i),$$

and *interim individual rationality* (IIR) constraints

$$U_i(v_i, m_i; v_i, m_i) \geq 0, \quad \forall (v_i, m_i) \in V \times M.$$

1.3.4 Principal's problem

The principal chooses the *private disclosure mechanism*, denoted by (Ξ, x), to maximize her ex ante expected utility, which is consisting of the information disclosure policy (M, Ξ) and the associated direct mechanism (V, M, x). The timing is as follows:

1. The principal makes public (Ξ, x), to which we assume that the principal can commit during the whole game.

2. Each agent i privately observes his own signal m_i generated by (M, Ξ), and then simultaneously reports $\hat{v}_i$ and $\hat{m}_i$ to the principal.

3. After observing the whole report profile $(\hat{v}, \hat{m})$, the principal implements the social alternatives according to $x(\hat{v}, \hat{m})$.

We assume for the moment that agents cannot communicate during the game. In Section 1.6.3 we relax this assumption and allow a subset of agents to share information about their signals. We also rule out the possibility of collusion among agents. The principal's problem, denoted by (P), is defined as follows:

$$\begin{aligned}
(P) \qquad & \sup_{\Xi, x} \quad \int_\theta \int_m \int_v u_0\big(x(v,m), v, \theta\big) dF_V(v) d\Xi_\theta(m) dF_0(\theta) \\
& s.t. \quad \forall i, (m_i, v_i) \neq (m_i', v_i'): \\
BIC_{m_i, v_i \to m_i', v_i'} \quad & \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i, v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\
& \geq \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i', v_{-i}, m_i', m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}), \\
IIR_{m_i, v_i} \quad & \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i, v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \geq 0.
\end{aligned}$$

We can immediately see the tradeoff faced by the principal: choosing more informative disclosure policy could make it more difficult to satisfy the BIC constraints and leave more information rent to the agents; however, less informative signals would increase the difficulty of implementing the proper allocation for each state, because the principal relies on the information revealed by the disclosure policy to determine the allocation.

It is worth noting that the signal profile is realized *before* agents make any report. As a result, the principal cannot make her disclosing strategy contingent on agents' private types v. One may wonder whether the principal can do strictly better in an alternative setting where she can commit to an information disclosure policy $\Xi : \Theta \times V \to \Delta(M)$ and ask each agent i to report v_i before the signal profile is realized. In Appendix B.2.1, we show that the optimal private disclosure mechanism in Theorem 1, 2, 3 already achieves the best outcome in this alternative setting. Thus, it is without loss of generality to adopt the framework of this paper.

1.4 Optimal Private Disclosure Mechanism

In this section, we characterize the optimal private disclosure mechanism defined by (P). To give an outline of how we proceed, we first consider a relaxed problem where the principal has "*strong control*" over the information disclosure process, in the sense that she directly observes the signal profile m received by all agents. Then we show that the principal can exploit the particular information structure derived from the relaxed problem to elicit agents' truthful reports about the signal profile.

1.4.1 Relaxed problem: Strong control

Let (P_1) be the relaxed problem of (P), where the principal has *strong control* over the disclosure process, so that agents cannot lie about their private signals. Since we have fewer BIC constraints, the value of the relaxed problem – that is, the principal's maximum expected payoff in (P_1) – is as an upper bound of the original problem (P).

$$
\begin{aligned}
(P_1) \qquad & \sup_{\Xi,\, x} \quad \int_\theta \int_m \int_v u_0\big(x(v,m), v, \theta\big) dF_V(v) d\Xi_\theta(m) dF_0(\theta) \\
& s.t. \quad \forall i, m_i, v_i \neq v_i' : \\
BIC_{v_i \to v_i' | m_i} \qquad & \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i, v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\
& \geq \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i', v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}), \\
IIR_{v_i | m_i} \qquad & \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x(v_i, v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \geq 0.
\end{aligned}
$$

Now we construct the following problem, denoted by (P^*), where no information disclosure policy is involved, but the principal can privately observe the true state and commit to an allocation rule $x : V \times \Theta \to \Delta(\mathscr{A})$.[11]

$$
\begin{aligned}
(P^*) \qquad & \sup_{x} && \int_\theta \int_v u_0\big(x(v,\theta),v,\theta\big) dF_V(v) dF_0(\theta) \\
& s.t. && \forall i, v_i \neq v_i' : \\
BIC_{v_i \to v_i'} \quad & && \int_\theta \int_{v_{-i}} u_i\big(x(v_i,v_{-i},\theta),v,\theta\big) dF_{-i}(v_{-i}) dF_0(\theta) \\
& && \geq \int_\theta \int_{v_{-i}} u_i\big(x(v_i',v_{-i},\theta),v,\theta\big) dF_{-i}(v_{-i}) dF_0(\theta) \\
IIR_{v_i} \quad & && \int_\theta \int_{v_{-i}} u_i\big(x(v_i,v_{-i},\theta),v,\theta\big) dF_{-i}(v_{-i}) dF_0(\theta) \geq 0.
\end{aligned}
$$

We can prove that the value of (P^*) is an upper bound of the value of (P_1). The basic idea is that, by pooling the signals for each agent, we can relax the incentive constraints and enlarge the set the implementable mechanisms.

Lemma 1. (P^*) is an relaxed problem of (P_1), and thus an relaxed problem of (P).

Proof. See Appendix A.1. □

Skreta (2011) considers a similar information disclosure problem to (P_1), except that the principal can observe the state and make use of her information to determine the allocation (even without disclosing it to the agents). Due to the inscrutability principle (Myerson, 1983), it is optimal for the principal to reveal no information to the agents, which is exactly given by (P^*) if we restate her results in our notations. However, Proposition 1 says that the principal's ability to directly observe and utilize the true state is *not* necessary to implement the optimal mechanism. In other words, the value of (P_1) can achieve the upper bound given by (P^*), even though the principal does not have more information about θ than m.

Proposition 1. The value of (P_1) is equal to the value of (P^*).

Before we prove the proposition, we first construct the information disclosure policy Φ^S (induced by Ξ^S and F_0, where the superscript "S" stands for "strong control"). Let $M_i = \{1,2,\dots,T\}$ for all i, and let

$$
\Phi^S(\theta_t, m) = \begin{cases} \frac{\alpha_t}{T}, & \text{if } m_{l_1} \equiv m_{l_2} + (l_1 - l_2)t \pmod{T}, \quad \forall l_1, l_2 \in I \\ 0, & \text{otherwise} \end{cases}
$$

[11] We assume that the feasible set of (P^*) is nonempty; that is, there exists one allocation rule satisfying all $(BIC_{v_i \to v_i'})$ and all (IIR_{v_i}). This is not a demanding assumption in most cases, for example, if $\mathscr{A}$ includes a social alternative $a_\emptyset$ where all agents opt out and get their own reservation utility, then the allocation rule which implements $a_\emptyset$ with probability 1 under all (v,θ) will be such a feasible candidate.

for any $(\theta_t, m) \in \Theta \times M$.[12] The following lemma summarizes several nice properties of this information structure.

Lemma 2. Φ^S satisfies the following properties:

(i) For any $m \in M$, there exists at most one θ such that $\Phi^S(\theta, m) > 0$.

(ii) For any i, m_i and θ, there exists a unique $m_{-i} \in M_{-i}$ such that $\Phi^S(\theta, m_i, m_{-i}) > 0$.

(iii) For any i and m_i, the marginal of $\Psi^S_{m_i}$ over Θ, denoted by $\Psi^S_{m_i}(\theta)$, equals F_0.

(iv) The marginal of Φ^S over Θ is F_0.

Proof. See Appendix A.2. □

Property (i) says the information policy is *aggregately revealing*, since after observing any signal profile that occurs with strictly positive probability, the principal can infer the realization of θ. This property enables the principal to make the allocation rule fully flexible with the state. Property (ii) shows that, after receiving any signal, each agent's posterior belief always exhibits a perfect correlation between the state and the opponents' signal profile (that occurs with strictly positive probability). Property (iii) proves that this disclosure policy is *individually uninformative* in the sense that, each individual signal reveals no new information about the state of the world, compared with the common prior. Property (iv) shows that Φ^S indeed meets the requirement of a feasible information disclosure policy. Next, we give the proof of Proposition 1.

Proof of Proposition 1. Let $\{x^*(v,\theta)\}_{(v,\theta)\in V\times\Theta}$ be the solution to (P^*).[13] Let $\Theta^+(m) := \{\theta \in \Theta \mid \Phi^S(\theta, m) > 0\}$. From property (i), $\Theta^+(m)$ has at most one element. Let $\theta^+(m) : M \to \Theta$ stand for its unique element if $\Theta^+(m)$ is not empty. From property (ii), we can define

[12] An equivalent way to define Φ^S is that: m_1 is uniformly distributed over $\{1,2,\ldots,T\}$ and independent of θ; while conditional on $\theta = \theta_t$, we have $m_i = m_1 + (i-1)t \mod T$ for any $i \geq 2$. The use of uniform distributions and modulo operation is crucial for individually uninformative disclosure policy. Similar idea is used to provide enough incentives for agents to take equilibrium strategies in recommendation games and communication networks. (See, e.g., Kalai, Kalai, Lehrer, and Samet, 2010; Renou and Tomala, 2012; Peters and Troncoso-Valverde, 2013)

[13] The existence of the solution to (P^*) is straightforward. Since we assume that Θ, V and $\mathscr{A}$ are all finite subsets, (P^*) is essentially a finite-dimensional linear programming problem whose feasible set is compact and nonempty. It follows that there exists a global maximizer. See Lemma 5 for the existence of optimal mechanisms without the finiteness assumptions.

a mapping $m^+_{-i} : \Theta \times M_i \to M_{-i}$, such that $\Phi^S\big(\theta, m_i, m^+_{-i}(\theta, m_i)\big) > 0$ for any (θ, m_i). We construct the allocation rule x^S:

$$x^S(v,m) = \begin{cases} x^*\big(v, \theta^+(m)\big), & \text{if } \Theta^+(m) \neq \emptyset \\ \tilde{x}, & \text{otherwise} \end{cases}$$

where $\tilde{x}$ is an arbitrary element in $\Delta(\mathscr{A})$. The choice of $\tilde{x}$ has no effect on the implementation because any m such that $\Theta^+(m) = \emptyset$ will never occur. Next, we show that the private disclosure mechanism (Φ^S, x^S), which can also be written as (Ξ^S, x^S), satisfies all the constraints in (P_1) and achieves the upper bound given by x^*. Pick any $i, m_i, v_i \neq v_i'$, since we have

$$\begin{aligned}
&\int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x^S(v_i', v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi^S_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\
&= \int_{v_{-i}} \int_{\theta} u_i\big(x^S(v_i', v_{-i}, m_i, m^+_{-i}(\theta, m_i)), v, \theta\big) d\Psi^S_{m_i}(\theta) dF_{-i}(v_{-i}) \\
&= \int_{v_{-i}} \int_{\theta} u_i\big(x^*(v_i', v_{-i}, \theta), v, \theta\big) dF_0(\theta) dF_{-i}(v_{-i}),
\end{aligned}$$

then immediately $(BIC_{v_i \to v_i' | m_i})$ and $(IIR_{v_i | m_i})$ are satisfied, due to the fact that x^* satisfies the constraints $(BIC_{v_i \to v_i'})$ and (IIR_{v_i}). On the other hand, the principal's ex ante expected utility under the private disclosure policy (Φ^S, x^S) is given by

$$\begin{aligned}
&\int_v \int_\theta \int_m u_0\big(x^S(v,m), v, \theta\big) d\Xi^S_\theta(m) dF_0(\theta) dF_V(v) \\
&= \int_v \int_\theta u_0\Big(\int_m x^S(v,m) d\Xi^S_\theta(m), v, \theta\Big) dF_0(\theta) dF_V(v) \\
&= \int_v \int_\theta u_0\Big(x^*(v,\theta), v, \theta\Big) dF_0(\theta) dF_V(v),
\end{aligned}$$

where the last equality comes from the fact that fixed any θ, any $m \in M$ satisfying $\Xi_\theta(m) > 0$ must also satisfy $\Phi(\theta, m) > 0$, and thus induces the same allocation rule $x^*(\cdot, \theta)$. To conclude, (Φ^S, x^S) constitutes a solution to problem (P_1), and achieves the value of (P^*). □

1.4.2 Original problem: Weak control

We come back to the original problem (P) where the principal cannot directly observe the signal profile. Then, multidimensional screening problems arise, since agents could misreport both their private types and signals. We prove that the solution to (P_1) can actually be implemented in (P) when there are four or more agents. This is because the IUAR disclosure policy constructed in Proposition 1 has a particular correlated structure which enables the principal to elicit truthful reports of the signal profile for free. The cases with two or three agents are discussed in the next subsection.

Theorem 1. When $N \geq 4$, the value of (P) is equal to the value of (P^*). The solution to (P), written as (Φ^W, x^W) (induced by Ξ^W, whose superscript "W" stands for "weak control"), is given by:

$$\Phi^W(\theta_t, m) = \begin{cases} \frac{\alpha_t}{K}, & \text{if } m_{l_1} \equiv m_{l_2} + (l_1 - l_2)t \pmod{K}, \quad \forall l_1, l_2 \in I \\ 0, & \text{otherwise;} \end{cases}$$

$$x^W(v, m) = \begin{cases} x^*\big(v, \theta^+(m)\big), & \text{if } \Theta^+(m) \neq \emptyset \\ x^*\big(v, \theta^+_{-i}(m)\big), & \text{if } \Theta^+(m) = \emptyset, \text{ and } \exists i \text{ such that } \Theta^+_{-i}(m) \neq \emptyset \\ \tilde{x}, & \text{otherwise,} \end{cases}$$

where $\tilde{x}$ is an arbitrary probability measure over $\mathscr{A}$, K is a prime number satisfying $K \geq \max\{T, N\}$,[14] $M_i = \{1, 2, \ldots, K\}$ for each $i \in I$, $\Theta^+_{-i}(m) := \big\{\theta_t \in \Theta \mid m_{l_1} \equiv m_{l_2} + (l_1 - l_2)t \pmod{K}, \forall l_1, l_2 \in I \setminus \{i\}\big\}$, and $\theta^+_{-i}(m)$ denotes its unique element if $\Theta^+_{-i}(m) \neq \emptyset$.

Proof. See Appendix A.3. □

The basic idea to induce agents to truthfully report their privately observed signals is to apply the "$(N-1)$ majority rule", that is, the allocation for each agent is determined by the opponents' reported signal profile rather than his own signal. When there are four or more agents, we can show (in Lemma 11) that the truthful report of signal profile by $(N-1)$ agents can uniquely pin down the true state.[15] Meanwhile, any report of signal profile off the equilibrium path can be induced by at most one agent's unilateral deviation from truth-telling equilibrium strategy. In other words, let $\mathscr{D}_i$ collect all possible signal profiles induced by misreport of only agent i, that is,

$$\mathscr{D}_i := \bigcup_{\theta \in \Theta} \bigcup_{m_i \in M_i} \Big(\big(M_i \setminus \{m_i\}\big) \times \{m^+_{-i}(\theta, m_i)\} \Big),$$

then we have $\mathscr{D}_i$ and $\mathscr{D}_j$ are disjoint for any $i \neq j$. Thus, if only agent i misreports his signal, the principal observing $(\hat{m}_i, m_{-i})$ can infer the true state $\theta = \theta^+_{-i}(\hat{m}_i, m_{-i})$ from m_{-i} and assign

[14] We can always find such K because there are infinitely many prime numbers by Euclid's theorem. The use of prime numbers is to give a unified way of proving the theorem, because when there are a limited number of agents, choosing $|M_i| = |\Theta|$ cannot guarantee that the principal could identify which agent takes the unilateral deviation and what is the true state. For example, we assume that $K = |\Theta| = N = 4$, and the principal receives a signal profile $(1, 1, 1, 3)$. Clearly, there exists some agent who misreports his signal; however, the principal cannot distinguish between two situations that both could induce such signal profile. Specifically, either the true state is θ_4 and the true signal profile is $(1, 1, 1, 1)$, but agent 4 misreports 3; or the true state is θ_2 and the true signal profile is $(1, 3, 1, 3)$, but agent 2 misreports 1.

[15] Similar property is named *Nonexclusivity in Information* by Postlewaite and Schmeidler (1986), who adopt this property to derive the sufficient conditions for implementation of social choice correspondences in differential information economies.

the corresponding allocation $x^*(\cdot,\theta)$ which is invariant with agent i's report, so as to make agent i indifferent between telling the truth and misreporting.[16]

The IUAR disclosure policy frees the principal from the tradeoff between keeping the mechanism more flexible with the true state, and leaving less information rent for agents.[17] By making each agent's posterior belief about the others' signals (perfectly) correlated with his posterior belief about the state, the principal achieves both goals at the same time. However, in single-agent case, if we keep the aggregately-revealing property, then there is no room for uncertainty in the agent's posterior belief about the state. Thus, the existence of multiple agents is indispensable for Theorem 1.

Also, the nature of private disclosure is critical for Theorem 1. To make (Φ^W, x^W) function properly, the principal must guarantee that each agent's information about the state is restricted to his own signal. Specifically, since any two signals can uniquely pin down the true state,[18] once an agent knows another agent's signal – possibly by side communication or "peeping" – he will know the true state and may find it profitable to deviate. Thus, the principal has to completely ban information sharing among any pair of agents. In Subsection 1.6.3 we relax this assumption by constructing a variation of (Φ^W, x^W) that is robust to information sharing among any subset of agents (excluding a very limited number of agents).

It's worth noting that Theorem 1 is also robust to various assumptions on the principal's access to information. To see this, even though the principal has no private information, she can achieve the same payoff as if she could directly observe both the state and the signal profile. Thus, if the principal also observes a (noisy) signal about the state, it is still optimal for her to simply ignore this private information and conduct the mechanism developed in Theorem 1.

We don't impose particular restrictions on the set of feasible information structures that can be used by the principal. This makes our result distinct from Bergemann and Pesendorfer (2007). To see this, they consider the auction environment, where the timing is close to ours, except that bidders can only learn about their valuations $v \in V = \times_{i\in I} V_i$ through private signals $(m_i)_{i\in I}$ generated by an information structure determined by the seller. The reason why optimal information structures take the form of monotone partitions is due to their assumption on the information structure, which says conditional on agent i's signal m_i, agent i's valuation v_i

[16] If all agents truthfully report their signals, we have already proved in Proposition 1 that agents will also report their true types. As for signal profiles which only can be obtained from misreporting by at least two agents, the associated allocations are irrelevant because we only consider unilateral deviations.

[17] For example, in Ottaviani and Prat (2001), Eső and Szentes (2007) and Yamashita (2017), the flexibility effect plays a dominant role, resulting in the optimality of full disclosure.

[18] See the proof of Lemma 2-(i).

is independent of the others' private signals m_{-i}.[19] Apparently, our disclosure policy doesn't satisfy this assumption. Let $\theta = (v_i)_{i\in I}$ be the state, and construct Φ^W as before. From property (ii) of Lemma 2, fixed any m_i, there exists a bijection between V and $\{m^+_{-i}(v) \mid v \in V\}$, which means v_i cannot be independent of m_{-i}. If we remove such restriction and allow the seller to adopt arbitrary information structure, then the optimality of IUAR disclosure policy also applies to their paper.

1.4.3 Two or three agents

When $N \leq 3$, we can no longer construct the allocation rule as in case $N \geq 4$, because there could exist $i \neq j$ such that $\mathscr{D}_i \cap \mathscr{D}_j \neq \emptyset$. For instance, let $m = (m_1, m_2, m_3)$ and $m' = (m_1, m'_2, m'_3)$ be two signal profiles which will be sent with strictly positive probabilities in Φ^W, satisfying $m_2 \neq m'_2$ and $m_3 \neq m'_3$. If m is the true signal profile and agent 2 misreports m'_2, then from the "majority rule" defined in the previous subsection, the allocation for off-equilibrium-path signal profile (m_1, m'_2, m_3) is $x^W(\cdot, m_1, m'_2, m_3) = x^*(\cdot, \theta^+_{-2}(m))$. Similarly, if m' is the true signal profile and agent 3 misreports m_3, we get exactly the same off-equilibrium-path signal profile (m_1, m'_2, m_3); but the allocation should be $x^W(\cdot, m_1, m'_2, m_3) = x^*(\cdot, \theta^+_{-3}(m'))$. Notice that $\theta^+_{-2}(m) \neq \theta^+_{-3}(m')$, then x^W is not well-defined at $(\cdot, m_1, m'_2, m_3)$ if $x^*(\cdot, \theta^+_{-2}(m)) \neq x^*(\cdot, \theta^+_{-3}(m'))$. Essentially, to make (Φ^W, x^W) function properly, any unilateral deviation from truthfully reporting one's signal must not affect the principal's inference about the true state.

Definition 1 ("Innocuous Unilateral Deviation"). An aggregately revealing information disclosure policy (M, Ξ) is immune to unilateral deviation if:

(1) Undetected unilateral deviations do not affect the principal's inference about θ, that is, fixed $\forall\theta \in \Theta$, and $\forall m \in M$ such that $\Xi_\theta(m) > 0$, we have: $\Xi_{\theta'}(m'_i, m_{-i}) = 0$, $\forall i \in I$, $\forall m'_i \neq m_i$ and $\forall \theta' \neq \theta$;

(2) On detecting any unilateral deviation, the principal's inference about θ is still not affected, that is, fixed $\forall m \in M$ satisfying $\Xi_\theta(m) > 0$ for some θ, and $\forall m'_i \neq m_i$ such that

[19] This conditional independence assumption is indispensable to how they define interim utilities. To be precise, given the auction mechanism (q, t), the interim utility of bidder i who observes m_i and reports $\hat{m}_i$ is $U_i(m_i, \hat{m}_i) = \mathbb{E}_{v_i, m_{-i}}[v_i q_i(\hat{m}_i, m_{-i}) \mid m_i] - \mathbb{E}_{m_{-i}}[t_i(\hat{m}_i, m_{-i}) \mid m_i]$. Since we have

$$\mathbb{E}_{v_i, m_{-i}}[v_i q_i(\hat{m}_i, m_{-i}) \mid m_i] = \int_{v_i, m_{-i}} v_i q_i(\hat{m}_i, m_{-i}) d\Psi_{m_i}(v_i, m_{-i}) = \int_{v_i} \int_{m_{-i}} v_i q_i(\hat{m}_i, m_{-i}) d\Psi_{m_i, v_i}(m_{-i}) d\Psi_{m_i}(v_i),$$

the condition to justify $U_i(m_i, \hat{m}_i) = \mathbb{E}[v_i \mid m_i] \cdot \mathbb{E}_{m_{-i}}[q_i(\hat{m}_i, m_{-i}) \mid m_i] - \mathbb{E}_{m_{-i}}[t_i(\hat{m}_i, m_{-i}) \mid m_i]$, as in their paper, is that $\Psi_{m_i, v_i}(m_{-i}) = \Psi_{m_i}(m_{-i})$, which implies m_{-i} is independent of v_i conditional on m_i.

$\Xi_{\theta'}(m_i', m_{-i}) = 0$ for all $\theta' \in \Theta$, we have: $\Xi_{\theta'}(m_i', m_j', m_{-ij}) = 0$, $\forall j \neq i$, $\forall m_j' \in M_j$ and $\forall \theta' \neq \theta$.

From the previous results, immediately we have that (M, Ξ^W) satisfies the innocuous unilateral deviation property if and only if $N \geq 4$. In fact, when there are less than four agents, Lemma 3 shows that any IUAR disclosure policy violates the innocuous unilateral deviation property. Thus, if the individually uninformative property is indispensable for implementing the solution to (P^*), then (Φ^W, x^W) no longer work. See Appendix B.3.1 for an example. Next, we prove Theorem 1 for cases with two or three agents by introducing certain assumptions.

Lemma 3. When $N = 2, 3$, aggregately revealing and innocuous unilateral deviation imply individually revealing, that is, $\Psi_{m_i}(\theta) \in \Delta(\Theta)$ is degenerate, $\forall i \in I$ and $\forall m_i \in M_i$.

Proof. See Appendix B.1.1. □

Three agents

Notice that in (M, Φ^W), the state θ can be uniquely inferred from the truthful report of signals by any two agents, then with three agents, any unilateral deviation from the equilibrium strategy can be detected by the principal. Thus, the principal can provide enough incentives for agents to truthfully report their signals by punishing all agents with a uniformly worst social alternative if a misreport is detected.

Assumption 1. There exists $\underline{a} \in \mathscr{A}$ such that

$$\underline{a} \in \bigcap_{i,v,\theta} \arg\min_{a \in \mathscr{A}} u_i(a, v, \theta).$$

The existence of such uniformly worst alternative is a relatively mild assumption. For example, if all agents' utility functions are non-negative, then a social alternative which gives each agent 0 utility under any possible realization of (v, θ) would satisfy this assumption. Such alternative can be interpreted as punishing all agents with reservation utility, which is 0 in our model. The following theorem characterizes the optimal private disclosure mechanism with weak control for three agents.

Theorem 2. Under Assumption 1, the optimal private disclosure mechanism with *weak control* when $N = 3$ is given by $(\Phi^W, x^{W|3})$, where

$$x^{W|3}(v, m) = \begin{cases} x^*\big(v, \theta^+(m)\big), & \text{if } \Theta^+(m) \neq \emptyset \\ \underline{a}, & \text{otherwise.} \end{cases}$$

Proof. See Appendix A.4. □

Particularly, we adopt the same information disclosure policy Φ^W as in Theorem 1, keep the on-path allocation rules of x^W unaffected, but modify the off-path allocation rules as follows: any unilateral deviation from truth-telling equilibrium will induce $x^{W|3}(\cdot, m) = \underline{a}$, which means the uniformly worst social alternative $\underline{a}$ is implemented with probability 1 conditional on a misreport m is received by the principal.

Remark 1. The reason why we need the uniformly worst social alternative is that, the principal cannot distinguish the agent who misreports, given an off-equilibrium-path signal profile induced by unilateral deviation. However, if the principal can tell which agent does not lie about his signal, then we can relax Assumption 1 by requiring the existence of punishment for any pair of agents rather than all agents; that is, for $i = 1,2,3$ there exists $\underline{a}_{-i} \in \mathscr{A}$ satisfying

$$\underline{a}_{-i} \in \bigcap_{j \neq i, v, \theta} \arg\min_{a \in \mathscr{A}} u_j(a, v, \theta).$$

In general, it is an open question whether such disclosure policy exists and how to characterize it. In Appendix B.2.4, we prove that when $T = 2$ and $N = 3$, the IUAR disclosure policy Φ^W meets the requirement that the principal knows which pair of agents the liar belongs to. Then observing any off-path signal profile m induced by unilateral deviation, the principal identifies the truth teller i, and implements $x^{W|3}(\cdot, m) = \underline{a}_{-i}$, so that the liar is punished for sure. While the on-path allocation rule is the same as x^W.

Two agents

When $N = 2$, the previous construction does not work even with uniformly worst social alternative, since the principal cannot tell whether there is unilateral deviation or not. See Appendix B.3.2 for an example. Notice that the disclosure policy Φ^W induces correlated signals among the agents, then potentially we can apply the Crémer-McLean mechanism to guarantee that agents send the true signal profile to the principal.

We assume that transfers are allowed in the associated mechanisms, that is, a feasible allocation is given by $\tilde{x} = \big(x, (p_i)_{i \in I}\big) \in \Delta(\mathscr{A}) \times \mathbb{R}^N$. The utility functions of the principal and agents are transferable, which are the sum of non-monetary utilities and transfers. Let $\tilde{u}_i(\tilde{x}, v, \theta) = u_i(x, v, \theta) + p_i$ be each agent i's utility, and $\tilde{u}_0(\tilde{x}, v, \theta) = u_0(x, v, \theta) - \sum_{i \in I} p_i$ be the principal's utility. Immediately, we have that $\tilde{u}_i(\tilde{x}, v, \theta)$ is linear with respect to $\tilde{x}$ for $i = 0, 1, \ldots, N$, thus the new payoff environment is included in the original model, and all previous results remain valid. The upper bound of the value of (P) is given by the value of (P^*), which is achieved by $\tilde{x}^* = \big(x^*, (p_i^*)_{i\in I}\big)$.

We keep the information disclosure policy Φ^W unaffected, and start from the allocation rule $x^{W|2}(v,m) = \tilde{x}^*\big(v, \theta^+(m)\big)$, for all $(v,m) \in V \times M$.[20] In general, $x^{W|2}$ alone is not Bayesian incentive compatible under the disclosure policy Φ^W, because agent i observing m_i may find it profitable to report $m_i' \neq m_i$. The main idea to elicit agents' truthful reports about m is to further introduce additional transfers from agents to the principal as $(t_i)_{i\in I} : M \to \mathbb{R}^N$, besides the transfers that are already included in $\tilde{x}^*$. Now given the other agents telling the truth, the interim expected utility of agent i with (v_i, m_i) by reporting $(\hat{v}_i, \hat{m}_i)$ is given by

$$\begin{aligned}\tilde{U}_i(\hat{v}_i, \hat{m}_i; v_i, m_i) = &\int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^{W|2}(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v_i, v_{-i}, \theta\big) d\Psi^W_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\ &- \int_{m_{-i}} t_i(\hat{m}_i, m_{-i}) d\Psi^W_{m_i}(m_{-i}),\end{aligned}$$

where $\Psi^W_{m_i}(m_{-i})$ is the marginal of $\Psi^W_{m_i}(\theta, m_{-i})$ over M_{-i}. We introduce the following assumption which restricts agents' non-monetary utility functions to be bounded.

Assumption 2. There exists $H > 0$ such that for any $i \in I$, $a \in \mathscr{A}$, $v \in V$ and $\theta \in \Theta$, we have $|u_i(a, v, \theta)| < H$.

By Assumption 2, the amount of violation of each Bayesian incentive compatibility constraints under $x^{W|2}$ is finite. On the other hand, we can always construct Φ^W such that for each agent, the conditional distributions of the other agent's signal on each possible realization of his own signal are linearly independent (Lemma 13). By manipulating the transfers $(t_i)_{i\in I}$ in the same way as in the Crémer-McLean mechanism, the principal can provide the agents sufficient incentives to truthfully report m.

Theorem 3. With transferable utilities and Assumption 2, the optimal private disclosure mechanism with *weak control* achieves the value of (P^*) for $N = 2$.

Proof. See Appendix A.5. □

1.5 Implementation: Sample-product approach

In this section we build the IUAR disclosure policy in a more practical manner, called the *sample-product approach*, which is illustrated as follows.[21]

[20] Since $N = 2$, we have $\Theta^+(m) \neq \emptyset$ for all $m \in M$.

[21] In Appendix B.2.6, we provide an alternative way of implementation, that is, restrictions on agents' access to certain information.

A seller, either producing some physical products or holding some financial instruments, is always eager to know about the fundamentals of the market, such as the fashion trend, consumer confidence, risk tolerance and market volatility. In practice, a frequently used approach is to let consumers try some sample products and infer from their feedbacks the targeted information. Moreover, these trials are usually carefully designed in the form of *blinded experiments*, where consumers only get an overall ranking of testing experiences but hardly understand the details of the sample products. Next, we show how these ideas could apply to the implementation of IUAR disclosure policy.

Let $\mathbb{S}(\ni s)$ be the space of sample products, where each buyer i has a strict and complete preference over $\mathbb{S}$ under each $\theta \in \Theta$, denoted by $\prec_i^\theta$.[22] We assume that, the buyer can rank any pair of sample products based on his overall testing experience, but he *cannot* recognize the identity of each sample product, due to the deliberate design of trials by the seller. We say a distinct pair of states (θ', θ'') is *distinguishable* in $\mathbb{S}$, if there exist $i, j \in I$ and $s', s'' \in \mathbb{S}$ such that

$$\begin{array}{ll} s' \prec_i^{\theta'} s'' & s' \prec_j^{\theta'} s'' \\ s' \prec_i^{\theta''} s'' & s' \succ_j^{\theta''} s''. \end{array} \tag{1.1}$$

In other words, θ' and θ'' are distinguishable if we can find two buyers and a pair of sample products such that, two buyers share the same preferred choice in one state while preference discordance occurs in the other state. We say $\mathbb{S}$ is *rich* if all distinct pairs of states drawn from Θ are distinguishable in $\mathbb{S}$.

We illustrate the sample-product approach for $T = 2$. Assume that θ_1 and θ_2 are distinguishable, with buyers $1, 2$ and s', s'' satisfying $s' \prec_1^{\theta_1} s''$, $s' \prec_2^{\theta_1} s''$, $s' \prec_1^{\theta_2} s''$, and $s' \succ_2^{\theta_2} s''$. The timing is: (i) both buyers are offered the *same* randomized Choice Pair (A, B) which is independent of the state and satisfies

$$\Pr\big((A,B) = (s', s'')\big) = \frac{1}{2}, \quad \Pr\big((A,B) = (s'', s')\big) = \frac{1}{2};$$

(ii) each buyer i privately observes a test result $m_i \in \{\text{“}A \prec B\text{”}, \text{“}A \succ B\text{”}\}$ for $i = 1, 2$, and then simultaneously reports it to the seller.

This randomization guarantees that no buyer can infer the true state from one's own test result. The joint distribution of (θ, m_1, m_2) is given by Table 1.1 (where buyer 1's feedback is listed in the first column and buyer 2's feedback is listed in the first row), which replicates (M, Ξ^W) in the optimal private disclosure mechanism.

[22]For simplicity, we assume that $\prec_i^\theta$ does not depend on agents' private type profile v.

Table 1.1: Joint distribution of (θ, m_1, m_2)

$\theta=\theta_1$	$A \prec B$	$A \succ B$
$A \prec B$	$\frac{\alpha_1}{2}$	0
$A \succ B$	0	$\frac{\alpha_1}{2}$

$\theta=\theta_1$	$A \prec B$	$A \succ B$
$A \prec B$	0	$\frac{\alpha_2}{2}$
$A \succ B$	$\frac{\alpha_2}{2}$	0

Then we consider the general case with $T \geq 2$. Fixed any (θ', θ'') that are distinguishable, we get $i, j \in I$ and $s', s'' \in \mathbb{S}$ satisfying (1.1). For any $\theta \in \Theta \setminus \{\theta', \theta''\}$, exactly one of the following four cases happens: (i) $s' \prec_i^\theta s''$, $s' \prec_j^\theta s''$; (ii) $s' \succ_i^\theta s''$, $s' \succ_j^\theta s''$; (iii) $s' \prec_i^\theta s''$, $s' \succ_j^\theta s''$; (iv) $s' \succ_i^\theta s''$, $s' \prec_j^\theta s''$. Buyers i and j are offered the same randomized Choice Pair (A, B) as in the binary-state case. Immediately, when θ belongs to case (i) and (ii), we have $\Xi_\theta(A \prec B, A \prec B) = \Xi_\theta(A \succ B, A \succ B) = \frac{1}{2}$; when θ belongs to case (iii) and (iv), we have $\Xi_\theta(A \prec B, A \succ B) = \Xi_\theta(A \succ B, A \prec B) = \frac{1}{2}$. Define $I_{i,j,s',s''}(\theta) = \{\tilde{\theta} \in \Theta \mid \Xi_\theta = \Xi_{\tilde{\theta}}\}$ as the collection of states such that, when offering sample products s', s'' to buyers i, j, the distribution of feedbacks conditional on these states coincide with θ. Thus, $I_{i,j,s',s''}(\theta')$ and $I_{i,j,s',s''}(\theta'')$ constitute a partition of Θ. It follows that the seller's posterior belief about the distribution of the state is $F_0\big(\cdot \mid I_{i,j,s',s''}(\theta')\big)$ if $m_i = m_j$, and is $F_0\big(\cdot \mid I_{i,j,s',s''}(\theta'')\big)$ if $m_i \neq m_j$; while buyers' knowledge about the state is still the common prior F_0. We name this sample-product procedure $S(\theta', \theta'')$ since it separates θ' from θ''. The last step is to repeat this procedure *independently* for all pairs of (θ', θ''), and after collecting all the feedbacks, the seller can infer the true state.

Theorem 4. If Θ permits a rich $\mathbb{S}$, then the sample-product approach can implement the individually uninformative and aggregately revealing disclosure policy.

Proof. See Appendix A.6. □

1.6 Discussion

1.6.1 Private disclosure vs. Public disclosure

In a public disclosure policy, denoted by Φ^P, the realizations of signals across the agents are always the same, that is, $M_i = \mathbf{C}$ for all $i \in I$, and for any $\theta \in \Theta$, we have

$$\Phi^P(\theta, m) > 0 \quad \Longrightarrow \quad m_i = m_j, \quad \forall i \neq j.$$

Obviously, public disclosure is a special case of private disclosure, but with simpler information structure, which might be easier to implement in reality. A natural question is whether private disclosure policy can achieve a better performance than the public disclosure policy.

With multiple agents, the principal essentially has strong control over the disclosure process, because she can exploit the perfect correlation of agents' signal profile and elicit truthful reports of the public signal for free. Let (Φ^{SP}, x^{SP}) be the optimal public disclosure policy with strong control. When $N \geq 3$, the principal can apply the "$(N-1)$ majority rule" where the allocation rule with weak control $x^{WP}(v,m)$ satisfies

$$x^{WP}(v,m) = \begin{cases} x^{SP}\big(v, c \cdot \mathbf{1}_N\big), & \text{if } \exists i \in I, \text{ s.t. } m_j = c \quad \forall j \neq i \\ \tilde{x}, & \text{otherwise,} \end{cases}$$

where $\tilde{x}$ is an arbitrary element of $\Delta(\mathscr{A})$, and $\mathbf{1}_N = (1, \ldots, 1) \in \mathbb{R}^N$; while when there are two agents, the principal can punish all agents by assigning a uniformly worst social alternative, that is, $x^{WP}(v,m) = \underline{a}$ if $m_1 \neq m_2$. By the same argument as in Section 1.4.2, agents will truthfully report the public signal to the principal. Thus, the optimal public disclosure policy (Φ^{WP}, x^{WP}) achieves the performance of (Φ^{SP}, x^{SP}), given by

$$\begin{aligned}
(P_{pub}) \qquad & \sup_{\Phi,\, x} && \int_c \int_\theta \int_v u_0\big(x(v, c \cdot \mathbf{1}_N), v, \theta\big) dF_V(v) d\Psi_c(\theta) d\Lambda(c) \\
& s.t. && \forall i, c \in \mathbf{C}, v_i \neq v_i' : \\
& BIC_{v_i \to v_i'|c} && \int_{v_{-i}} \int_\theta u_i\big(x(v_i, v_{-i}, c \cdot \mathbf{1}_N), v, \theta\big) d\Psi_c(\theta) dF_{-i}(v_{-i}) \\
& && \geq \int_{v_{-i}} \int_\theta u_i\big(x(v_i', v_{-i}, c \cdot \mathbf{1}_N), v, \theta\big) d\Psi_c(\theta) dF_{-i}(v_{-i}) \\
& IIR_{v_i|c} && \int_{v_{-i}} \int_\theta u_i\big(x(v_i, v_{-i}, c \cdot \mathbf{1}_N), v, \theta\big) d\Psi_c(\theta) dF_{-i}(v_{-i}) \geq 0 \\
& && \int_c \Psi_c(\cdot) d\Lambda(c) = F_0(\cdot).
\end{aligned}$$

Define $\Pi(\Psi_c)$ as the principal's expected payoff under the optimal BIC mechanism on receiving public signal c, that is,

$$\begin{aligned}
\Pi(\Psi_c) = \sup_x \quad & \int_\theta \int_v u_0\big(x(v, c \cdot \mathbf{1}_N), v, \theta\big) dF_V(v) d\Psi_c(\theta) \\
s.t. \quad & \forall i, v_i \neq v_i' : \quad (BIC_{v_i \to v_i'|c}) \quad \text{and} \quad (IIR_{v_i|c}) \quad \text{hold.}
\end{aligned}$$

Then (P_{pub}) becomes the standard Bayesian persuasion problem whose solution depends on the shape of $\Pi(\Psi_c)$ on its domain $\Delta(\Theta)$, as well as the common prior F_0. Let $\hat{\Pi}$ be the concave closure of Π, defined by $\hat{\Pi}(\Psi_c) := \sup\{z \mid (\Psi_c, z) \in Conv(\Pi)\}$, where $Conv(\Pi)$ denotes the convex hull of the graph of Π. Thus, private disclosure policy achieves a strictly higher expected payoff for the principal than public disclosure policy if and only if the value of (P^*) is higher than $\hat{\Pi}(F_0)$.

Linear environment with independent private value

Yamashita (2017) considers linear environment with independent private values and one-dimensional payoff types, where an allocation $x = (q_i, p_i)_{i=1}^N \in \mathbb{R}^{2N}$ consists of a non-monetary allocation to agent i, named q_i, and monetary transfer from agent i to the principal, named p_i. Let $Q \subseteq \mathbb{R}^N$ denote the feasible set of non-monetary allocations to the agents, which is assumed to be convex since we allow randomized allocations; while no restriction is imposed on p. Agent i's payoff is $q_i y_i(v_i, \theta) - p_i$, and the principal's payoff is $y_0(q, v, \theta) + \sum_{i=1}^N p_i$, where y_0, y_i are bounded and continuous in all their arguments, y_0 is linear with respect to q, and $\frac{\partial y_i}{\partial v_i} > 0$. Yamashita (2017) proves that $\Pi(\Psi_c)$, as a function of Ψ_c, is convex and continuous over $\Delta(\Theta)$, and thus, by Jensen's inequality, full revelation is optimal for public disclosure, and the maximum ex ante expected payoff for the principal is given by

$$(P'_{pub}) \quad \sup_{q(v,\theta)\in Q} \int_\theta \int_v \Big(y_0\big(q(v,\theta), v, \theta\big) + \sum_{i=1}^N q_i(v,\theta)\gamma_i(v_i,\theta) \Big) dF_V(v) dF_0(\theta)$$
$$s.t. \quad \mathbb{E}_{v_{-i}}\big[q_i(v_i, v_{-i}, \theta)\big] \leq \mathbb{E}_{v_{-i}}\big[q_i(v'_i, v_{-i}, \theta)\big], \quad \forall i, \theta, v_i < v'_i,$$

where $\gamma_i(v_i,\theta) = y_i(v_i,\theta) - \frac{dy_i(v_i,\theta)}{dv_i}\frac{1-F_i(v_i)}{f_i(v_i)}$ is the virtual value function, for $i = 1, \ldots, N$.

Lemma 4. In linear environment, the solution to (P_1) is given by:

$$(P'_1) \quad \sup_{\Phi, q\in Q} \int_v \int_\theta \Big[y_0\big(\chi(v,\theta), v, \theta\big) + \sum_{i=1}^N \chi_i(v,\theta)\gamma_i(v_i,\theta) \Big] dF_0(\theta) dF_V(v)$$
$$s.t. \quad \forall i, m_i, v_i < v'_i:$$
$$Mon_{v_i < v'_i} \quad \int_{\theta, m_{-i}} \Big(\mathbb{E}_{v_{-i}}[q_i(v'_i, v_{-i}, m)] - \mathbb{E}_{v_{-i}}[q_i(v_i, v_{-i}, m)] \Big) \cdot \big(y_i(v'_i,\theta) - y_i(v_i,\theta) \big) d\Psi_{m_i}(\theta, m_{-i}) \geq 0,$$

where $\chi_i(v,\theta) = \int_m q_i(v,m) d\Xi_\theta(m)$.

Proof. See Appendix B.1.2. □

Lemma 4 characterizes the optimal private disclosure mechanism with strong control. Then we provide a sufficient condition for an equivalence result between private disclosure and public disclosure.

Proposition 2. If the monotonicity constraints in (P'_{pub}) are all slack, then the optimal private disclosure mechanism achieves no better performance than the optimal public disclosure mechanism.

Proof. See Appendix B.1.3. □

Proposition 2 characterizes a class of environments where full disclosure remains optimal even if the principal can use private disclosure. In mechanism design literature we usually impose certain version of regularity conditions on payoff environments to make the monotonicity conditions irrelevant. Consider a single-object auction where the principal is purely revenue-maximizing, that is, $Q := \{q \in [0,1]^N \mid \sum_{i=1}^N q_i \leq 1\}$ and $y_0 = 0$. If we assume the single-crossing condition: for any $i, v_i < v_i', j, v_j, \theta$, we have $\gamma_i(v_i, \theta) \geq \gamma_j(v_j, \theta)$ implies $\gamma_i(v_i', \theta) > \gamma_j(v_j, \theta)$, then the solution to the unconstrained problem of (P_{pub}') automatically satisfies all the monotonicity constraints.

On the other hand, the optimality of full disclosure fails when private disclosure is strictly better than public disclosure. Notice that (P_{pub}') and (P_1') have the same form of objective function, then the necessary condition for private disclosure to strictly improve the principal's expected payoff is that, (P_1') permits a larger feasible set than (P_{pub}').

Proposition 3. Any private disclosure mechanism that achieves strictly higher payoff than the optimal public disclosure mechanism must send *noisy* signals about the states to some agent.

Proof. See Appendix B.1.4. □

The optimal public disclosure mechanism requires that the monotonicity constraints should be satisfied under all possible realizations of θ; while the optimal private disclosure mechanism only imposes an "interim" version of monotonicity constraints on the allocation rule, where the amount of violation of the monotonicity constraint under some state potentially can be offset by its slackness under other states. This serves as the main reason why it is beneficial for the principal to send (completely) noisy signals about the state of the world to each single agent.[23]

1.6.2 Finiteness assumption

So far, we have assumed that $\mathscr{A}$, $(V_i)_{i\in I}$ and Θ are all finite sets, in order to (i) establish the existence of the solution to (P), and (ii) facilitate the construction of the optimal private disclosure policy. Next, we show that the finiteness assumption is innocuous.

[23] See Appendix B.3.3 for a numerical example to illustrate this idea.

Existence of the solution to (P)

In Section 1.4 we establish the existence of the solution to (P) by showing that its relaxed problem (P^*) has a solution which is also implementable in (P). Here we provide a sufficient condition to guarantee that (P^*) admits an optimal mechanism x^*.

Lemma 5. Assume that $\mathscr{A}$, $(V_i)_{i\in I}$ and Θ are all nonempty compact metrizable spaces, u_0 and $(u_i)_{i\in I}$ are measurable and bounded functions, and are continuous on the domain $\mathscr{A}\times V\times\Theta$. Then, the solution to the relaxed problem (P^*) exists, provided that there exists one mechanism satisfying all $(BIC_{v_i\to v_i'})$ and (IIR_{v_i}) constraints.

Proof. See Appendix B.1.5. □

Lemma 5 directly follows from Theorem 13.5 in Kadan, Reny, and Swinkels (2017). We can see that none of the previous finiteness assumptions is necessary for the existence of solution to (P).

Disclosure policy for continuously distributed states

We now assume that $\Theta=[0,T]\in\mathbb{R}$, subject to a continuous cumulative distribution function $F_0(\theta)$ with full-support density function $f_0(\theta)$. Assume that conditions in Lemma 5 hold. As in the discrete case, let $\{x^*(v,\theta)\}_{(v,\theta)\in V\times\Theta}$ be the solution to (P^*). Let $M_i=\Theta=[0,T]$ for $i=1,\ldots,N$. Define

$$h(x)=x-\max\{n\in\mathbb{Z}\mid n\le\frac{x}{T}\}\cdot T$$

as a natural generalization of the remainder in Euclidean division. We construct the following information disclosure policy with strong control:

$$\Phi^S(\theta,m)=\begin{cases}\frac{f_0(\theta)}{T}, & \text{if } m_s=h\big(m_t+(s-t)\theta\big),\quad \forall s,t\in I\\ 0, & \text{otherwise.}\end{cases}$$

It is easy to check that Φ^S also satisfies all the properties in Lemma 2. To prove property (i) and (ii), we make similar arguments as in discrete case by using the fact that, the restriction of $h(x)$ to any domain $[a,b]$ such that $b-a\le T$ is a bijection. Property (iii) is satisfied because for any $\theta\in\Theta$ and $m_i\in M_i$ we have

$$F_0(\theta\mid m_i)=\frac{\int_{m_{-i},\tilde{\theta}\le\theta}d\Phi^S(\tilde{\theta},m_i,m_{-i})}{\int_{m_{-i},\tilde{\theta}\le T}d\Phi^S(\tilde{\theta},m_i,m_{-i})}=\frac{\int_{\tilde{\theta}\le\theta}\frac{f_0(\tilde{\theta})}{T}d\tilde{\theta}}{\int_{\tilde{\theta}\le T}\frac{f_0(\tilde{\theta})}{T}d\tilde{\theta}}=\frac{F_0(\theta)}{F_0(T)}=F_0(\theta).$$

Property (iv) holds because for any $A \subseteq \Theta$ we have

$$\begin{aligned}\int_{\theta\in A, m\in M} d\Phi^S(\theta,m) &= \int_{m_i\in M_i}\int_{\theta\in A}\int_{m_{-i}\in M_{-i}} d\Phi^S(m_{-i}\mid m_i,\theta)d\Phi^S(\theta\mid m_i)d\Phi^S(m_i)\\ &= \int_{m_i\in M_i}\int_{\theta\in A} dF_0(\theta\mid m_i)d\Lambda^S(m_i)\\ &= \int_{m_i\in M_i}\int_{\theta\in A} dF_0(\theta)d\Lambda^S(m_i) = F_0(A).\end{aligned}$$

Thus, we can define $\theta^+(m)$ as before, and construct the allocation rule x^S as follows:

$$x^S(v,m) = \begin{cases} x^*\big(v,\theta^+(m)\big), & \text{if } \Theta^+(m)\neq\emptyset \\ \tilde{x}, & \text{otherwise,}\end{cases}$$

where $\tilde{x}$ is arbitrarily drawn from $\Delta(\mathscr{A})$. Similar to Proposition 1, (Φ^S, x^S) is Bayesian incentive compatible and achieves the same expected payoff as x^* in (P^*). By applying the techniques we developed in Section 1.4.2, we can extend Theorem 1, 2, 3 to the case with continuously distributed states. (See Section B.2.7.)

1.6.3 Information sharing among agents

We consider the same setup as in Section 1.3. For simplicity, we only consider the case with four or more than four agents. Since $\Theta = \{\theta_1,\dots,\theta_T\}$ is a finite set, we rename θ_t as t for $t = 1,\dots,T$. Information sharing among agents is modeled in the following way: before agents (simultaneously) report their own signals to the principal, each agent i knows $\mathscr{E}_i := (m_j)_{j\in\tilde{I}_i}$, that is, the realizations of signals observed by agents in the subset $\tilde{I}_i \subseteq I$. Obviously, we have $i\in\tilde{I}_i$.

We first construct the information disclosure policy, denoted by Ξ^{IS}. As before, define $M_i = \{1,\dots,K\}$ for all $i\in I$, where $K \geq \max\{T,N\}$ is a prime number. Let $\varepsilon_1,\dots,\varepsilon_{N-3}\in\{1,\dots,K\}$ be mutually independent random variables, satisfying $\Pr(\varepsilon_\tau = k) = \frac{1}{K}$ for $\tau = 1,\dots,N-3$ and $k = 1,\dots,K$. Moreover, $(\varepsilon_1,\dots,\varepsilon_{N-3})$ is independent of θ. The signal profile $m = (m_1,\dots,m_N)$ is defined as follows:

$$\begin{pmatrix} m_1\\ \vdots\\ m_{N-3}\\ m_{N-2}\\ m_{N-1}\\ m_N\end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ 1 & 2 & \cdots & N-2\\ 1 & 2^2 & \cdots & (N-2)^2\end{pmatrix}\cdot\begin{pmatrix}\theta\\ \varepsilon_1\\ \vdots\\ \varepsilon_{N-3}\end{pmatrix} \mod K,$$

where each signal m_i is equal to the *residue* of the dot product of i-th row vector (denoted as $\boldsymbol{\zeta}_i$) and column vector $(\theta, \varepsilon_1, \ldots, \varepsilon_{N-3})$ (denoted as $\boldsymbol{\varepsilon}$), modulo K.[24]

Lemma 6. Ξ^{IS} satisfies the following properties:

(i) For any $\tilde{I} \subseteq I$ such that $|\tilde{I}| \leq N-3$, the posterior belief about θ conditional on observing truthful reports of $(m_i)_{i\in\tilde{I}}$, denoted by $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta)$, coincides with F_0.

(ii) Any truthful reports of $(N-2)$ signals uniquely pin down the realization of θ, as well as $(\varepsilon_1, \ldots, \varepsilon_{N-3})$ and the remaining two signals.

(iii) Any report of N signals reveals whether there exists unilateral deviation from truth-telling, and, if so, the identity of the agent who misreports.

Proof. See Appendix B.1.6. □

Here is the rough idea of how to prove Lemma 6. Because the congruence relation in modular arithmetics on the integers is compatible with addition, subtraction, and multiplication, we proceed in a similar way to linear algebra. It is straightforward to show that $(\boldsymbol{\zeta}_i)_{i\in\tilde{I}}$ are linearly independent for any $\tilde{I} \subseteq I$ such that $|\tilde{I}| = N-2$. Then the $(N-2)$-dimensional vector $\boldsymbol{\varepsilon}$ can be determined by $(N-2)$ equations induced by any $(N-2)$ signals. With $(N-3)$ or less signals, there will be too many degrees of freedom to determine $\boldsymbol{\varepsilon}$. While with N signals, there are more equations than variables so that the principal can perform a cross check on any unilateral misreport.

By Lemma 6, if $|\tilde{I}_i| \leq N-3$ for all $i \in I$, then Ξ^{IS} meets all the requirements for Theorem 1 to hold: the disclosure policy should be individually uninformative and aggregately revealing, and be able to identify the agent who misreports.

Theorem 5. When $N \geq 4$ and $|\tilde{I}_i| \leq N-3$ for all $i \in I$, the optimal private disclosure mechanism that is *immune to information sharing* among agents is given by (Ξ^{IS}, x^{IS}):

$$x^{IS}(v,m) = \begin{cases} x^*(v,\theta), & \text{if } \exists i \in I, \exists \theta \in \Theta, \text{ s.t. } \theta = \theta^+(m_{-ij}), \forall j \in I \setminus \{i\} \\ \tilde{x}, & \text{otherwise,} \end{cases}$$

where $\tilde{x}$ is arbitrarily drawn from $\Delta(\mathscr{A})$, and $\theta^+(\cdot)$ is defined in the proof of Lemma 6.

Proof. See Appendix A.7. □

[24]The residue is calculated in the way of standard modular arithmetic, except that when the dot product can be exactly divided by K, we write $m_i = K$ instead of $m_i = 0$.

Theorem 5 says, as long as the principal guarantees that each agent knows up to $(N-3)$ signals, the optimal disclosure mechanism won't be affected by information sharing among agents. In practice, the assumption in Theorem 5 is relatively easy to hold. For example, the principal only needs to keep three agents' signals absolutely secret from the other agents. Also, the principal may divide agents into groups each including at most $(N-3)$ agents, and forbid any inter-group communication.

Information sharing can be modeled in a more general way: each agent i knows $\mathscr{E}_i := F_{\mathscr{E}_i} \in \Delta\big(\prod_{j\in\tilde{I}_i} M_j\big)$ where $|\tilde{I}_i| \leq N-3$, and the marginal distribution of $F_{\mathscr{E}_i}$ over M_i is degenerate. Basically, each agent knows the realization of his own signal, and receives a noisy information about the others' signals. The only restriction on agents' communications is that, any "new" information received by each agent about the others' signals should involve at most $(N-4)$ signals. Since

$$\begin{aligned}\Pr(\theta \mid \mathscr{E}_i) &= \sum_{(m_j)_{j\in\tilde{I}_i}} \Pr\big(\theta \mid (m_j)_{j\in\tilde{I}_i}, \mathscr{E}_i\big) \Pr\big((m_j)_{j\in\tilde{I}_i} \mid \mathscr{E}_i\big) \\ &= \sum_{(m_j)_{j\in\tilde{I}_i}} \Psi^{IS}_{(m_j)_{j\in\tilde{I}_i}}(\theta) F_{\mathscr{E}_i}\big((m_j)_{j\in\tilde{I}_i}\big) = \sum_{(m_j)_{j\in\tilde{I}_i}} F_0(\theta) F_{\mathscr{E}_i}\big((m_j)_{j\in\tilde{I}_i}\big) = F_0(\theta),\end{aligned}$$

Theorem 5 remains valid for this general version of information sharing. For instance, agents — whose private information is assumed to only consist of the signals generated by Ξ^{IS} — may communicate through an arbitrary network. Agents send and receive various messages (e.g., cheap-talk messages and verifiable messages) in arbitrary number of rounds. In equilibrium, each agent acquires additional information about the signals observed by those who are either directly or indirectly connected to him. Theorem 5 guarantees that x^* can be implemented through Ξ^{IS} as long as the network graph is *disconnected* and each connected sub-network has at most $(N-3)$ agents.

1.6.4 Ex post optimality

We have assumed that the principal commits to implementing $x(\hat{v}, \hat{m})$ if agents report $(\hat{v}, \hat{m})$. Generally speaking, this assumption is indispensable to the implementation of the optimal private disclosure mechanism; otherwise after observing agents' truthful reports $(\hat{v}, \hat{m}) = (v, m)$, the principal would know $(v, \theta) = \big(v, \theta^+(m)\big)$ and solve $\max_{x\in\mathscr{X}(v,m)} u_0\big(x, v, \theta^+(m)\big)$, where

$$\mathscr{X}(v,m) = \bigcap_{i\in I} \{x^*(v_i, \tilde{v}_{-i}, \tilde{\theta}) \mid (\tilde{v}_{-i}, \tilde{\theta}) \in V_{-i} \times \Theta\},$$

which might contain allocations other than $x^*(v, \theta)$. However, if we restrict agents' private information to be derived from the information disclosure process only (as in Bergemann and Pesendorfer, 2007), then under certain conditions the principal has no incentive to deviate from adopting the signal profile she actually receives, which is named as the *ex post optimality* property.

Proposition 4. If $V_i = \{v_i\}$ for all $i \in I$, and all (IIR_{v_i}) are slack in (P^*), then the optimal private disclosure mechanism satisfies the ex post optimality property.

Proof. See Appendix B.1.7. □

With all interim individual rationality constraints being slack, the principal already implements the best outcome under each realization of the state by obeying the allocation rule. Thus, the optimal private disclosure mechanism is still robust even if there is limited supervision on the principal's ex-post execution. Examples satisfying the slackness condition in Proposition 4 could be that, $u_i(\cdot)$ for all $i \in I$ are non-negative.

1.7 Conclusion

We study the problem where at the first stage, the principal secretly sends personalized signals to each agent (without observing either the true state or the realized signal profile) according to a private disclosure policy which she can commit to; then at the second stage she implements the outcomes contingent on agents' reports according to the allocation rule. We find that the optimal disclosure policy is individually uninformative and aggregately revealing in the sense that: it reveals no new information about the state to each single agent, while after gathering agents' truthful reports of the signal profile, the principal can infer the exact realization of the state.

We also show that in a class of environments – including those satisfying the standard regularity conditions – the optimal public disclosure mechanism achieves the same expected payoff for the principal as the optimal private disclosure mechanism. This finding links the individually uninformative result in our paper to the optimality of full disclosure in literature, and throws light on the optimal information structure in a much more general setup. We conclude the paper by providing some possible applications of the optimal private disclosure mechanism.

1.7.1 Collection and transmission of confidential information

The spy story in the Introduction is a perfect example of applying the IUAR disclosure policy to collection of confidential information. In reality, an intelligence agency, say, CIA, can effectively control what and how precisely spies learn about the confidential information, by authorizing each spy to investigate a particular aspect. First of all, CIA should be able to infer the targeted information after synthesizing spies' reports, which implies the requirement of aggregately revealing. But more importantly, CIA must restrict spies' knowledge about the secret in order to prevent information leaks, because some spy would betray or be caught by the enemy. To meet this individually uninformative requirement in practice, spies are usually not be told why CIA commands them to collect seemingly irrelevant information.

In a similar spirit, the IUAR disclosure policy provides a novel way for confidential information transmission. Imagine that a sender wants to transmit a secret to a receiver through a communication network, but ensures that none of the nodes in the network get any useful information about the secret in order to prevent it from being leaked to some adversary (see, for example, Renault, Renou, and Tomala, 2014). Assume that the network has a simple structure, where (i) each node is directly connected to the sender and the receiver, and (ii) no pair of nodes are linked with each other. Then the optimal private disclosure mechanism defined in Subsection 1.4.2 constitutes a secure communication protocol to transmit θ, where

(1) the sender and the receiver agree on the disclosure policy (M, Ξ^W);

(2) the sender secretly sends m_i to node i according to $\Xi^W_\theta(m)$;

(3) after receiving m, the receiver deciphers the secret according to $\theta^+(m)$.

Theorem 1 shows that this protocol (i) satisfies *secrecy* since the information structure is individually uninformative, and (ii) is *reliable* because the confidential information can be aggregately revealed even if there exists unilateral deviation from properly executing the protocol (provided there are four or more nodes).

1.7.2 Differential privacy

Differential privacy, originally proposed by Dwork, McSherry, Nissim, and Smith (2006) and formalized in Dwork (2006), is a definition of privacy in computer science which says the outcomes generated by a randomized algorithm over two similar databases are approximately the same. It ensures that the same conclusion will be reached, regardless of any individual's presence in the database, so that data users could learn relatively accurate group properties but infers no personal identifiable information about any individual.[25] In contrast to confidential

[25]Differential privacy has been adopted by some major technology companies, so as to seek approval to use client data. For example, Google's Chrome Web browser has implemented Randomized Aggregatable Privacy-

information transmission, it is the input database rather than the output of the algorithm that should be kept secret.

Although our paper is not in a statistics setting, the optimal private disclosure mechanism meets the requirements of differential privacy. Particularly, we assume that there are more than four agents, and let the signal profile m be the database and the inferred state $\theta^{+}(m)$ be the released result after processing the data. By Theorem 1, the state can be uniquely inferred from any $(N-1)$ agents' reports about their signals, which means the presence of any individual signal in the database will not affect the outcome of the algorithm $\theta^{+}(m)$. Moreover, conditional on any realization of the state, individual signal is uniformly distributed, which means that any data user who learns about the query result $\theta = \theta^{+}(m)$ gets no new information about the signal observed by each single agent. Thus, our results potentially applies to various real-world contexts where differential privacy is required to address the privacy protection issues.

Preserving Ordinal Response (RAPPOR), a technology for crowdsourcing statistics subject to differential privacy, to collect data about Chrome clients, and learn statistics about unwanted or malicious hijacking of user settings (Erlingsson, Pihur, and Korolova, 2014). Another example is the adoption of differential privacy technology by Apple starting from its iOS 10, which helps Apple discover the usage patterns of a large number of users without compromising individual privacy.

Chapter 2

On the Foundations of Ex Post Incentive Compatible Mechanisms[1]

Takuro Yamashita[2] **Shuguang Zhu**[3]

July 20, 2017

In private-value auction environments, Chung and Ely (2007) establish maxmin and Bayesian foundations for dominant-strategy mechanisms. We first show that similar foundation results for ex post mechanisms hold true even with interdependent values if the interdependence is only *cardinal*. This includes, for example, the one-dimensional environments of Dasgupta and Maskin (2000) and Bergemann and Morris (2009b). Conversely, if the environment exhibits *ordinal* interdependence, which is typically the case with multi-dimensional environments (e.g., a player's private information comprises a noisy signal of the common value of the auctioned good and an idiosyncratic private-value parameter), then in general, ex post mechanisms do not have foundation. That is, there exists a non-ex-post mechanism that achieves strictly higher expected revenue than the optimal ex post mechanism, regardless of the agents' high-order beliefs.

[1]We are grateful to Thomas Mariotti, Renato Gomes, Jiangtao Li, Pierre Boyer, Takakazu Honryo, Raphaël Levy, Andras Niedermayer, Chengsi Wang, Yi-Chun Chen, Tilman Börgers, Takashi Kunimoto, and seminar participants at Toulouse School of Economics, Mannheim (Center for Doctoral Studies in Economics), Hitotsubashi University, and Decentralization Conference (Michigan). Takuro Yamashita also gratefully acknowledges financial support from the European Research Council (Starting Grant #714693).

[2]Toulouse School of Economics, University of Toulouse Capitole, France. takuro.yamashita@tse-fr.eu

[3]Toulouse School of Economics, University of Toulouse Capitole, France. shuguang.zhu@tse-fr.eu

2.1 Introduction

The recent literature on mechanism design provides a series of studies on the robustness of mechanisms, motivated by the idea that a desirable mechanism should not rely too heavily on the agents' common knowledge structure.[4] One approach taken in the literature is to adopt stronger solution concepts that are insensitive to various common knowledge assumptions. For instance, in private-value environments, Segal (2003) studies dominant-strategy incentive compatible sales mechanisms. In interdependent-value environments, Dasgupta and Maskin (2000) study efficient auction rules that are independent of the details under the concept of ex post incentive compatibility.

However, a mechanism that achieves desired outcomes without the agents' common knowledge assumption does not immediately imply dominant-strategy or ex post incentive compatibility. In revenue maximization in private-value auction (under "regularity" conditions), Chung and Ely (2007) fill in this gap by establishing the *maxmin* and *Bayesian* foundation of the optimal dominant-strategy mechanism, in the following sense. Consider a situation where the seller in an auction (principal) only knows a joint distribution of the bidders' (agents) valuation profile for the auctioned object, which may be based on data about similar auctions in the past. On the other hand, he does not have reliable information about the bidders' beliefs about each other's value. For example, the bidders may have more or less information than the seller, or may simply have a "wrong" belief (from the seller's point of view) for various reasons. Thus, the seller's objective is to find a mechanism that achieves a good amount of revenue *regardless of the bidders' (high-order) beliefs*. Note that, in a dominant-strategy mechanism, it is always an equilibrium for each bidder to report his true value, and therefore, it always guarantees the same level of expected revenue. On the other hand, in non-dominant-strategy mechanisms, expected revenue may vary with the bidders' (high-order) beliefs. In the definition of Chung and Ely (2007), there is a maxmin foundation for a dominant-strategy mechanism if, for any non-dominant-strategy mechanism, there is a possible belief of the seller with which the dominant-strategy mechanism achieves (weakly) higher expected revenue than the non-dominant-strategy mechanism.[5]

In this paper, we examine the existence of such foundations for ex post incentive compatible mechanisms in interdependent-value environments. Our main observation is that the

[4] See, for example, Wilson (1985).

[5] As a stronger concept, if the same belief can be found for any non-dominant-strategy mechanism with which a dominant-strategy mechanism achieves (weakly) higher expected revenue, then there is a *Bayesian* foundation, because, as long as the seller is Bayesian rational and has that particular belief, he finds it optimal to offer a dominant-strategy mechanism, even though he can also offer any other mechanism.

key property that guarantees such foundations is what we call the *cardinal* vs. *ordinal* interdependence. To explain these concepts, imagine an auction problem, where each bidder's willingness-to-pay depends both on his own type and the other bidders' types. If one type of each bidder always has a higher valuation for the good than another type *regardless of the other bidders' types* (even if each type's valuation itself may vary with the others' types), then we say that the environment exhibits only *cardinal* interdependence. Conversely, if the types cannot be ordered in such a uniform manner with respect to the others' types, then we say that the environment exhibits *ordinal* interdependence.[6]

We first show that, in the environments with only cardinal interdependence, (both maxmin and Bayesian) foundations exist for ex post mechanisms. This includes, for example, private-value environments (in this sense, our result is a generalization of Chung and Ely (2007)), and the one-dimensional environments of Dasgupta and Maskin (2000) and Bergemann and Morris (2009b).

Conversely, if the environment exhibits *ordinal* interdependence, which is typically the case with multi-dimensional environments (e.g., a player's private information comprises a noisy signal of the common value of the auctioned good and an idiosyncratic private-value parameter), then in general, ex post mechanisms do not have foundation. That is, there exists a non-ex-post mechanism that achieves strictly higher expected revenue than the optimal ex post mechanism, regardless of the agents' high-order beliefs.

Regarding the foundation results, Chen and Li (2016) consider a general class of private-value environments where agents have multi-dimensional payoff types, and show that if the environment satisfies the *uniform-shortest-path-tree* property, then the maxmin (and Bayesian) foundation exists for dominant-strategy mechanisms. This property simply means that, for any allocation rule the principal desires to implement, the set of binding constraints is invariant. This holds true in the single-good auction environment of Chung and Ely (2007) with regularity, and in this sense, their result generalizes that of Chung and Ely (2007), keeping the private-value assumption. Our work is a complement to Chen and Li (2016) in that we consider interdependent-value environments. For our foundation result (Theorem 6), a similar property to their uniform-shortest-path-tree property holds, which suggests that some of their

[6] These interdependence concepts are obviously related to the "size" of interdependence (e.g., private-value environments are special cases of cardinally interdependent cases). However, they are not necessarily corresponding to each other. For example, if a bidder's valuation in an auction is a sum of a function only of his own type and another function of the others' types, then however large is the second term, the environment never exhibits ordinal interdependence. In this sense, a more appropriate interpretation is that these interdependent concepts are related to the *diversity* of interdependence across types.

argument may be applicable even in interdependent-value environments.

Regarding the no-foundation results, there are several papers in the literature that provide examples or a restrictive class of environments in which (various versions of) foundations for dominant-strategy or ex post mechanisms do not exist. For example, for interdependent-value environments, Bergemann and Morris (2005) provide examples in the context of implementation of certain ("non-separable") social choice correspondences, and Jehiel, Meyer-ter Vehn, Moldovanu, and Zame (2006) provide an example for revenue maximization in sequential sales. Chen and Li (2016) also provide an instance of environment where, without their uniform-shortest-path-tree property, there might not exist a foundation for dominant-strategy mechanisms, even in private-value environments. Our work contributes to this line of research by providing a general class of environments with a no-foundation result (instead of providing examples), and the economic intuition based on the cardinal vs. ordinal interdependence.

Other closely related papers include Bergemann and Morris (2005) and Börgers (2013). In interdependent-value environments, Bergemann and Morris (2005) show that any *separable* social choice correspondence that is implementable given any (high-order) belief structure of the agents must satisfy ex post incentive compatibility. In this sense, they provide another sort of foundation for ex post incentive compatible mechanisms. Their separable social choice correspondence necessarily admits a unique non-monetary allocation for each payoff-type profile, and hence, in general, excludes revenue maximization as the principal's objective. Thus, our work is complementary to theirs in that we consider revenue maximization.

Börgers (2013) criticizes the foundation theorems by constructing a non-dominant-strategy (or more generally, a non-ex-post) mechanism that yields *weakly* higher expected revenue than the optimal dominant-strategy mechanism for any belief structure of the agents, while it yields *strictly* higher expected revenue for some belief structures. Our no-foundation result is stronger in that it provides a *strict* improvement in expected revenue for *any* (high-order) belief structure, though under stronger conditions on the environment.

2.2 Model

There is a finite set of risk-neutral agents, $1, 2, \ldots, I$. Agent i's privately-known *payoff type* is $\theta_i \in \Theta_i \subseteq \mathbb{R}^d$, where $|\Theta_i| < \infty$.[7] A payoff-type profile is written as $\theta = (\theta_1, \ldots, \theta_I) \in \Theta_1 \times \ldots \times \Theta_I = \Theta$. The principal's (subjective) prior belief for θ is given by $f \in \Delta(\Theta)$, where we assume $f(\theta) > 0$ for all $\theta \in \Theta$.

[7] Potential extensions to cases with continuous payoff type spaces are discussed in Section 2.5.

Each agent i's willingness-to-pay for $q_i \in Q_i \subseteq \mathbb{R}_+$ units of the good is denoted by $v_i(q_i, \theta)$. We assume that $0 \in Q_i$, $|Q_i| < \infty$,[8] $v_i(0, \theta) = 0$, and $v_i(\cdot, \theta)$ is increasing for all θ. Moreover, as a standard single-crossing condition, we assume that, for each $\theta_i \neq \theta_i'$, and θ_{-i}, we have either

$$v_i(q_i, \theta_i, \theta_{-i}) - v_i(q_i', \theta_i, \theta_{-i}) > v_i(q_i, \theta_i', \theta_{-i}) - v_i(q_i', \theta_i', \theta_{-i}), \quad \forall q_i > q_i';$$

or

$$v_i(q_i, \theta_i, \theta_{-i}) - v_i(q_i', \theta_i, \theta_{-i}) < v_i(q_i, \theta_i', \theta_{-i}) - v_i(q_i', \theta_i', \theta_{-i}), \quad \forall q_i > q_i'.$$

In the first (second) case, we denote $\theta_i \succ_i^{\theta_{-i}} \theta_i'$ ($\theta_i \prec_i^{\theta_{-i}} \theta_i'$, respectively). Our assumption throughout the paper is that $\prec_i^{\theta_{-i}}$ is a total ordering over Θ_i for any θ_{-i}, although $\prec_i^{\theta_{-i}}$ can be different from $\preceq_i^{\theta_{-i}'}$. Let

$$\eta = \min_{i, \theta_i \neq \theta_i', \theta_{-i}, q_i \neq q_i'} |v_i(q_i, \theta_i, \theta_{-i}) + v_i(q_i', \theta_i', \theta_{-i}) - v_i(q_i', \theta_i, \theta_{-i}) - v_i(q_i, \theta_i', \theta_{-i})| \; (> 0).$$

In particular, this implies that, by taking $q_i > 0 = q_i'$,

$$|v_i(q_i, \theta_i, \theta_{-i}) - v_i(q_i, \theta_i', \theta_{-i})| \geq \eta$$

for all $\theta_i \neq \theta_i'$ and θ_{-i}.

Paying $p_i \in \mathbb{R}$ to the principal, agent i's final payoff is $v_i(q_i, \theta) - p_i$. The principal's objective is the total revenue, $\sum_i p_i$. The feasible set of $q = (q_1, \ldots, q_I)$ is denoted by $Q \subseteq \prod_i Q_i$, where the shape of Q depends on the specific environment of interest.

For example, auctions, trading, and public-goods environments are in this class, with (or without) interdependence. In terms of interdependence, our framework includes a typical "common + private" environment studied in the auction literature: Imagine that each agent i has a unit demand for the good, his payoff-type comprises $(c_i, d_i) \in \Theta_i \subseteq \mathbb{R}^2$, where c_i may be interpreted as a "common-value" component and d_i may be interpreted as an idiosyncratic "private-value" component, and his valuation for the good is $\pi_i(c_1, \ldots, c_N) + d_i$ for some function π_i.

2.2.1 Type space

The agents' private information includes their own payoff types, their (first-order) beliefs about their payoff types, and their arbitrarily higher-order beliefs. To model this, we introduce type spaces as in Bergemann and Morris (2005).

[8] As it becomes clearer, the finiteness of Q_i is without loss of generality (though it simplifies the notation), given that Θ is finite and we only consider finite mechanisms (including ex post incentive compatible mechanisms).

A ("known-own-payoff-type") type space, denoted by $\mathscr{T} = (T_i, \widehat{\theta}_i, \widehat{\pi}_i)_{i=1}^I$, is a collection of a measurable space of types T_i for each agent i, a measurable function $\widehat{\theta}_i : T_i \to \Theta_i$ that describes the agent's payoff type, and a measurable function $\widehat{\pi}_i : T_i \to \Delta(T_{-i})$ that describes his belief about the others' types. Let $\widehat{\beta}_i(t_i)$ denote the belief hierarchy associated with type t_i (i.e., it describes t_i's first-order belief about θ_{-i}, second-order belief, and so on, up to an arbitrary high order). We say that $\mathscr{T}$ has *no redundant types* if for each i, mapping $t_i \mapsto (\widehat{\theta}_i(t_i), \widehat{\beta}_i(t_i))$ is one-to-one.

In fact, there exists a (compact) *universal type space* $\mathscr{T}^* = (T_i^*, \widehat{\theta}_i^*, \widehat{\pi}_i^*)_{i=1}^I$, such that any type space without redundant types can be embedded into it, in the following sense.[9]

Lemma 7. Let $\mathscr{T}$ be a type space with no redundant types. Then, for each i, there exist subsets $\widehat{T}_i \subset T_i^*$ and bijections $h_i : T_i \to \widehat{T}_i$ such that:

1. $\widehat{\theta}_i^*(h_i(t_i)) = \widehat{\theta}_i(t_i)$ for all $t_i \in T_i$; and

2. $\widehat{\pi}_i^*(h_i(t_i))[h_{-i}(t_{-i})] = \widehat{\pi}_i(t_i)[t_{-i}]$ for all $t_i \in T_i$ and $t_{-i} \in T_{-i}$,

where $h_{-i}(t_{-i}) = (h_1(t_1), \ldots, h_{i-1}(t_{i-1}), h_{i+1}(t_{i+1}), \ldots, h_I(t_I))$.

In what follows, we directly work with this universal type space.[10] Specifically, let $\mu \in \Delta(T^*)$ represent the principal's prior belief over T^* such that $\mu(\{t|\widehat{\theta}^*(t) = \theta\}) = f(\theta)$ for each θ, that is, the principal's (first-order) belief for θ is given by $f(\theta)$, as assumed above. The other information contained in μ captures the principal's belief over the agents' possible belief structures. Let $\mathscr{M} \subseteq \Delta(T^*)$ represent the set of all such μ.

In some contexts, it may be reasonable to assume that (the principal believes that) the agents do not have extreme (non-full-support) first-order beliefs. For example, instead of assuming that each agent's belief or knowledge is exogenous, one may be interested in a situation where each agent engages in his own information acquisition (through which his belief is updated), where the information acquisition cost is a linear function of relative entropy (Sims (2003)). Then, it is infinitely costly for each agent to know other agents' payoff types.

Formally, let $\mathscr{M}^{\text{full}} \subset \mathscr{M}$ denote the set of μ such that every agent i has a full-support first-order belief about the other agents. More precisely, for each agent i with type t_i, let $\widehat{\pi}_i^{*1}(t_i) \in \Delta(\Theta_{-i})$ denote his first-order belief, that is, $\widehat{\pi}_i^{*1}(t_i)[\theta_{-i}] = \int_{t_{-i}|\widehat{\theta}^*_{-i}(t_{-i})=\theta_{-i}} d\widehat{\pi}_i^*(t_i)[t_{-i}]$ for each θ_{-i}. Then, $\mathscr{M}^{\text{full}}$ is the set of all $\mu \in \mathscr{M}$ such that $\mu\left(\{t \mid \forall i, \theta_{-i},\ \widehat{\pi}_i^{*1}(t_i)[\theta_{-i}] > 0\}\right) = 1$.

[9] For constructions of universal type spaces, see Mertens and Zamir (1985) and Brandenburger and Dekel (1993).

[10] The results would not change even if we allow for type spaces with redundant types, but more notation would be involved.

2.2.2 Mechanism

The principal designs a mechanism, denoted by (M,q,p), where M_i represents a message set for each agent i, $M = M_1 \times \ldots \times M_I$, $q : M \to Q = [0,1]^I$ is an allocation rule, and $p : M \to \mathbb{R}^I$ is a payment function. Each agent i reports a message $m_i \in M_i$ simultaneously, and then he receives the good with probability $q_i(m)$ and pays $p_i(m)$ to the principal. We assume that M_i contains a non-participation message $\emptyset \in M_i$ such that $\big(q_i(\emptyset,m_{-i}), p_i(\emptyset,m_{-i})\big) = (0,0)$ for any $m_{-i} \in M_{-i}$.

We now introduce a class of mechanisms with ex post incentive compatibility (an EPIC mechanism for short).

Definition 2. An *EPIC mechanism* is a mechanism $\Gamma = (M,q,p)$ such that, for each i, (i) $M_i = \Theta_i$, and (ii) for each $\theta \in \Theta$ and $\theta_i \neq \theta_i' \in \Theta_i$:

$$\begin{aligned} v_i(q_i(\theta),\theta) - p_i(\theta) &\geq 0, \\ v_i(q_i(\theta),\theta) - p_i(\theta) &\geq v_i(q_i(\theta_i',\theta_{-i}),\theta) - p_i(\theta_i',\theta_{-i}). \end{aligned}$$

The expected revenue in the truth-telling (ex post) equilibrium in an EPIC mechanism is given by:

$$R_f(\Gamma) = \sum_{\theta} \sum_{i} p_i(\theta) f(\theta).$$

Note that this does not depend on μ, and in this sense, $R_f(\Gamma)$ may be interpreted as a "robustly guaranteed" expected revenue with respect to the agents' beliefs and higher-order beliefs. Let R_f^{EP} denote the maximum expected revenue among all EPIC mechanisms.

Applying the standard argument, the optimal mechanism among all EPIC mechanisms is characterized by the corresponding virtual-value maximization. To explain this, let $F_i(\theta_i,\theta_{-i}) = \sum_{\tilde{\theta}_i \preceq_i^{\theta_{-i}} \theta_i} f(\tilde{\theta}_i,\theta_{-i})$ denote the cumulative distribution function of i's payoff types given the other agents' payoff-type profile θ_{-i}.

Agent i's *virtual valuation* at payoff-type profile θ is given by:

$$\gamma_i(q_i,\theta) = v_i(q_i,\theta) - \frac{\sum_{\tilde{\theta}_i \succ_i^{\theta_{-i}} \theta_i} f(\tilde{\theta}_i,\theta_{-i})}{f(\theta)} \big(v_i^+(q_i,\theta) - v_i(q_i,\theta)\big),$$

where $v_i^+(\theta_i,\theta_{-i}) = \min_{\tilde{\theta}_i \succ_i^{\theta_{-i}} \theta_i} v_i(\tilde{\theta}_i,\theta_{-i})$ whenever the right-hand side is well-defined; otherwise $\gamma_i(q_i,\theta) = v_i(q_i,\theta)$.

The following result is standard, so we omit its proof.

Lemma 8.

$$R_f^{EP} = \max_{q:\Theta\to Q} \sum_\theta \sum_i \gamma_i(q_i(\theta),\theta) f(\theta)$$
$$\text{sub. to} \quad \forall i, \theta_i, \theta_i', \theta_{-i};$$
$$\theta_i \succ_i^{\theta_{-i}} \theta_i' \Rightarrow q_i(\theta_i,\theta_{-i}) \geq q_i(\theta_i',\theta_{-i}). \qquad \text{(M)}$$

We assume that the solution exists in this maximization problem, which we denote by $q^{EP} = (q_i^{EP}(\theta))_{i,\theta}$. The corresponding payment rule is denoted by $p^{EP} = (p_i^{EP}(\theta))_{i,\theta}$.[11]

As in Chung and Ely (2007), we further assume the following "regularity" condition throughout the paper.

Assumption 3. There exists $\varepsilon > 0$ such that, for any distribution over Θ, $\tilde{f}$, such that $\|\tilde{f} - f\| < \varepsilon$ (in a Euclidean distance), the monotonicity constraints (M) are not binding in the problem of $R_{\tilde{f}}^{EP}$. In particular, this implies

$$R_f^{EP} = \max_{q:\Theta\to Q} \sum_\theta \sum_i \gamma_i(q_i(\theta),\theta) f(\theta).$$

Of course, the conditions on the environment that imply the above assumption can vary with the environment. For example, in an auction environment with $Q = \{q \in \{0,1\}^N | \sum_i q_i \leq 1\}$, the regularity assumption is satisfied if, for each $i \neq j$, and θ,[12]

$$\gamma_i(\theta) \geq \gamma_j(\theta) \Rightarrow \forall \theta_i' > \theta_i,\ \gamma_i(\theta_i',\theta_{-i}) > \gamma_j(\theta_i',\theta_{-i}).$$

In a *digital-good* environment of Goldberg, Hartline, Karlin, Saks, and Wright (2006) with $Q = \{0,1\}^N$, the regularity assumption is satisfied under the strict monotone hazard rate condition, i.e., for each i and θ, $\frac{1-F_i(\theta)}{f(\theta)}$ is decreasing in θ_i. In a multi-unit sales environment as in Mussa and Rosen (1978), the regularity assumption is satisfied under the strict monotone hazard rate condition and concavity of each v_i with respect to q_i.

[11] p^{EP} is given as follows. For each i, θ_i and θ_{-i}, (i) if there is no $\theta_i' \prec_i^{\theta_{-i}} \theta_i$, then

$$p_i^{EP}(\theta) = v_i(q_i^{EP}(\theta),\theta);$$

(ii) otherwise, letting $\theta_i' \prec_i^{\theta_{-i}} \theta_i$ be such that no θ_i'' satisfies $\theta_i' \prec_i^{\theta_{-i}} \theta_i'' \prec_i^{\theta_{-i}} \theta_i$,

$$p_i^{EP}(\theta) = v_i(q_i^{EP}(\theta),\theta) - v_i(q_i^{EP}(\theta_i',\theta_{-i}),\theta) + p_i^{EP}(\theta_i',\theta_{-i}).$$

[12] Chung and Ely (2007) call it the single-crossing condition. A stronger sufficient condition is the combination of the strict monotone hazard rate property (i.e., for each i and θ, $\frac{1-F_i(\theta)}{f(\theta)}$ is decreasing in θ_i), and affiliation in f (which includes independent f as a special case).

The following notation is extensively used in the subsequent analysis. For each i and $q_i > 0$, define

$$\Theta_i^*(q_i, \theta_{-i}) \quad = \quad \{\theta_i \in \Theta_i | q_i^{EP}(\theta_i, \theta_{-i}) \geq q_i\}$$

as the set of i's payoff types whose allocation given θ_{-i} is greater than or equal to q_i in the optimal EPIC mechanism. Note that, by monotonicity, if $\theta_i \in \Theta_i^*(q_i, \theta_{-i})$ and $\theta_i' \succ_i^{\theta_{-i}} \theta_i$, then $\theta_i' \in \Theta_i^*(q_i, \theta_{-i})$. Let $\theta_i^*(q_i, \theta_{-i})$ be the lowest element in $\Theta_i^*(q_i, \theta_{-i})$ with respect to $\prec_i^{\theta_{-i}}$, that is, for any $\theta_i \in \Theta_i^*(q_i, \theta_{-i})$, we have $\theta_i \succ_i^{\theta_{-i}} \theta_i^*(q_i, \theta_{-i})$. This $\theta_i^*(q_i, \theta_{-i})$ is called i's threshold type with respect to q_i given θ_{-i}. Finally, let

$$\Theta_{-i}^*(q_i, \theta_i) = \{\theta_{-i} \in \Theta_{-i} | \theta_i \in \Theta_i^*(q_i, \theta_{-i})\}$$

denote the set of θ_{-i} with which θ_i is allocated greater than or equal to q_i units in the optimal EPIC mechanism.

2.2.3 Foundations

For a non-EPIC mechanism, expected revenue may vary with the agents' belief structure, and the principal—who does not know the agents' belief structure—may not want to offer a mechanism if the expected revenue is low for some possible belief structures. Following Chung and Ely (2007), we say that there is a *maxmin foundation* for EPIC mechanisms if, for any non-EPIC mechanism $\Gamma = (M, q, p)$, there exists $\mu \in \mathscr{M}$ such that, for any Bayesian equilibrium σ^*, the expected revenue obtained in the equilibrium is less than R_f^{EP}, that is:

$$\int_{t \in T} \sum_i p_i(\sigma^*(t)) d\mu \leq R_f^{EP}.$$

If there exists a single $\mu \in \mathscr{M}$ that achieves the above inequality for all Γ, then we say that there is a *Bayesian foundation* for EPIC mechanisms.[13]

[13] These definitions are consistent with the verbal explanations of the corresponding definitions in Chung and Ely (2007). However, in fact, the mathematical definitions of them in Chung and Ely (2007) are slightly different: for example, their mathematical definition of maxmin foundation says that, for any non-EPIC mechanism $\Gamma = (M, q, p)$,

$$\inf_{\mu \in \mathscr{M}} \left[\max_{\sigma^*:\text{Bayesian equilibrium}} \int_{t \in T} \sum_i p_i(\sigma^*(t)) d\mu \right] \leq R_f^{EP}.$$

To see the difference, let $R(\mu)$ denote the term inside the bracket on the left-hand side (i.e., the expected revenue given μ), and imagine a case where (i) $R(\mu) > R_f^{EP}$ for any μ, while (ii) for any $\varepsilon > 0$, there exists μ such that $R(\mu) - \varepsilon < R_f^{EP}$. That is, the non-EPIC mechanism Γ is a *strict* improvement over the optimal EPIC mechanism, In the context where (the principal believes that) the agents have full-support first-order beliefs, we replace $\mathscr{M}$ by $\mathscr{M}^{\text{full}}$ in the above definitions, and we say that there is a *strong* maxmin / Bayesian foundation for EPIC mechanisms.

2.3 Without ordinal interdependence

First, we consider the case where, for each i, θ_{-i}, and θ'_{-i}, $\prec_i^{\theta_{-i}} = \prec_i^{\theta'_{-i}}$. This includes the private-value environment (as in Chung and Ely (2007)) as a special case, but also includes some interdependent-value environments. For example, assume that $\Theta_i \subseteq \mathbb{R}$ and $v_i(q_i, \theta_i, \theta_{-i})$ is an increasing function of θ_i for each given q_i, θ_{-i}. Because i's payoff is affected by θ_{-i}, the environment exhibits interdependence, but it is only *cardinal* interdependence in the sense that a higher value of θ_i corresponds to a higher type with respect to $\prec_i^{\theta_{-i}}$ for any θ_{-i}.

On the other hand, even if v_i is increasing in θ_i, if $\Theta_i \subseteq \mathbb{R}^d$ with $d > 1$, it is possible to have $\prec_i^{\theta_{-i}} \neq \prec_i^{\theta'_{-i}}$ for some θ_{-i} and θ'_{-i}. For example, consider an auction environment in which each agent i's payoff-type comprises $(c_i, d_i) \in \Theta_i \subseteq \mathbb{R}^2$, where c_i denotes a "common-value" component and d_i denotes an idiosyncratic "private-value" component, and his valuation for the good is $\pi_i(c_1, \ldots, c_N) + d_i$ for some function π_i strictly increasing in all the arguments. Then, for $(c_i, d_i), (c'_i, d'_i) \in \Theta_i$ such that $c_i < c'_i$ and $d_i > d'_i$, it is possible that, given some c_{-i}, (c_i, d_i) has a higher valuation for the good than (c'_i, d'_i) (i.e., $\pi_i(c_i, c_{-i}) + d_i > \pi_i(c'_i, c_{-i}) + d'_i$), while given another c'_{-i}, (c_i, d_i) has a lower valuation than (c'_i, d'_i).

Such environments with *ordinal* interdependence are studied in the next section.

Definition 3. We have *ordinal interdependence* if there exists i, θ_{-i}, and θ'_{-i} such that $\prec_i^{\theta_{-i}} \neq \prec_i^{\theta'_{-i}}$.

Generalizing Chung and Ely (2007) (for private-value auction environments), we show that no ordinal interdependence implies the strong maxmin / Bayesian foundations for EPIC mechanisms.

Theorem 6. With Assumption 3 and no ordinal interdependence, EPIC mechanisms have the strong Bayesian (and hence strong maxmin) foundation.

Our proof for Theorem 6 is a direct extension of Chung and Ely (2007) in the private-value setting to the interdependent-value environment. We provide a sketch of the proof here, and the formal proof in the Appendix.

while it is not a *uniform* improvement. The verbal definition of Chung and Ely (2007) (which we follow in this paper) suggests that there is no maxmin foundation, while their mathematical definition says there is. The difference is not innocuous, because the non-EPIC mechanism we propose is indeed such a mechanism.

First, we impose the *non-singularity* condition on the payoff-type distribution f, which says that f satisfies certain full-rank conditions, and consider the Bayesian mechanism design problem with a simple type space having a particular belief structure. We show that under such a belief structure, it is without loss of generality to treat all participation constraints and all "adjacent downward" incentive constraints with equality, and ignore all the other constraints. Then we show that the total expected revenue in this Bayesian problem is maximized by the optimal EPIC mechanism.

The next step is to relax the non-singularity assumption by choosing a sequence of non-singular distributions which converge to the given payoff-type distribution. Since the optimal EPIC mechanisms achieve the highest expected revenue over the sequence of simple type spaces with the particular belief structure, by taking the limit, we show that the Bayesian foundation also exists for any arbitrary payoff-type distribution, as long as Assumption 3 is satisfied.[14]

2.4 With ordinal interdependence

In this section, we consider the environment that further satisfies the following conditions.

Assumption 4 ("Highest Payoff Type")**.** For each i, there exists $\bar{\theta}_i \in \Theta_i$ such that, for each $\theta_i \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$, we have $\bar{\theta}_i \succ_i^{\theta_{-i}} \theta_i$.

Assumption 5 ("Richness")**.** For each i, q_i, θ_i, θ_i' and θ_{-i} such that $v_i(q_i, \theta_i, \theta_{-i}) > v_i(q_i, \theta_i', \theta_{-i})$, there exists θ_{-i}' such that $\theta_i \in \Theta_i^*(q_i, \theta_{-i}')$ and $\theta_i' \notin \Theta_i^*(q_i, \theta_{-i}')$.

The highest-payoff-type assumption is satisfied if Θ is a complete sublattice in $\mathbb{R}^d$, $v_i(q_i, \theta)$ is increasing in θ. The richness assumption connects the difference among i's different payoff types and the difference among their allocations. For example, consider an auction environment where each i's payoff type is $(c_i, d_i) \in C_i \times D_i (= \Theta_i)$ where $C_i, D_i \subseteq \mathbb{R}$, and his valuation is $\pi_i(c) + d_i$. The richness assumption would be easily satisfied if each D_j is rich enough so that, if $\pi_i(c_i, c_{-i}) + d_i > \pi_i(c_i', c_{-i}) + d_i'$ for some $c_i, c_i' \in C_i$, $d_i, d_i' \in D_i$ and $c_{-i} \in C_{-i}$, then we can find some d_{-i} such that agent i's virtual value is the highest given (c_i, d_i) (and (c_{-i}, d_{-i})) but not given (c_i', d_i') (and (c_{-i}, d_{-i})).

In this environment, EPIC mechanisms have the strong (maxmin and Bayesian) foundation if and only if we do not have ordinal interdependence.

[14]In their paper, they show by example that, without the condition corresponding to Assumption 3, there may not exist a Bayesian foundation.

Theorem 7. Under Assumptions 3, 4, 5, EPIC mechanisms have the strong foundation if and only if we do not have ordinal interdependence.

Before the formal proof of the theorem, we provide the basic intuition through the following two examples.

Example 1. Assume $I = 2$, $\Theta_1 = \Theta_2 = \{1,2\}$, and $Q = \{0,1\}^2$. Table 1 collects payoff-type distribution f, agent 1's valuation and virtual value at each payoff type profile, and the corresponding optimal EPIC allocation for agent 1. For agent 2, assume that $v_2(\theta) = \theta_2 + 1$ for all θ so that the optimal EPIC allocation for him is $(q_2^{EP}(\theta), p_2^{EP}(\theta)) = (1,2)$ for all θ.

Table 2.1: Auction environment of Example 1.

$f, v_1, \gamma_1, (q_1^{EP}, p_1^{EP})$	$\theta_2 = 1$	$\theta_2 = 2$
$\theta_1 = 1$	$\frac{1}{6}$, 2, 2, $(1,1)$	$\frac{1}{6}$, 1, -1, $(0,0)$
$\theta_1 = 2$	$\frac{1}{3}$, 1, $\frac{1}{2}$, $(1,1)$	$\frac{1}{3}$, 2, 2, $(1,2)$

We have $\Theta_1^*(q_1, \theta_2) = \{1,2\}$ if $(q_1, \theta_2) = (1,1)$ and $\Theta_1^*(q_1, \theta_2) = \{2\}$ if $(q_1, \theta_2) = (1,2)$. Hence, the threshold payoff type of agent 1 given $\theta_2 = 1$ (i.e., $\theta_1 = 2$) is assigned the goods given $\theta_2 = 2$, but the non-threshold winning payoff type of agent 1 given $\theta_2 = 1$ (i.e., $\theta_1 = 1$) is unassigned given $\theta_2 = 2$. This reversal of the order over agent 1's payoff types is crucial for the no-foundation result.

Now we consider a modification of the optimal EPIC mechanism, which asks agent 1's first-order belief. More specifically, agent 1 is asked to report his payoff type θ_1 and his belief for $\theta_2 = 1$, that is:

$$y(t_1) = \int_{t_2 | \widehat{\theta}_2(t_2) = 1} d\widehat{\pi}_1(t_1)[t_2].$$

If he reports $\theta_1 = 1$ and first-order belief $y \in [0,1]$, agent 1 obtains the goods by paying $(2 - \cos\eta)$ under $\theta_2 = 1$, but fails to get the goods and still needs to pay $(1 - \sin\eta)$ under $\theta_2 = 2$, where $\eta = \arctan\frac{1-y}{y}$. We keep the optimal EPIC allocations for both agents in the other cases. It is easy to verify that the new mechanism is Bayesian incentive compatible over the universal type space.

Because we are interested in the strong foundation, assume that (the principal believes that) agent 1 always has a full-support first-order belief, that is, $y \in (0,1)$ with (μ-)probability one. Then, agent 1 with $\theta_1 = 1$ always pays strictly more than 1 regardless of his (full-support) first-order belief and agent 2's true payoff type: if $\theta_2 = 1$, agent 1 pays $2 - \cos\eta$ for some $\eta \in (0, \frac{\pi}{2})$, which is strictly greater than 1; if $\theta_2 = 2$, agent 1 pays $1 - \sin\eta$ for some $\eta \in (0, \frac{\pi}{2})$, which is strictly greater than 0.

Therefore, this new mechanism raises strictly higher expected revenue than the optimal EPIC mechanism, as long as agent 1 has a full-support first-order belief.

Example 2. Assume $I = 2$, $\Theta_1 = \{1,2,3\}$, $\Theta_2 = \{1,2\}$ and $Q = \{0,1\}^2$. We focus on agent 1 because the designer decides allocation rules for each agent separately in digital-goods auctions. Table 2.2 collects payoff-type distribution f, agent 1's valuation and virtual value at each payoff type profile, and the corresponding optimal EPIC allocation for agent 1. Clearly, agent 1's preference exhibits ordinal interdependence.

Table 2.2: Auction environment of Example 2.

$f, v_1, \gamma_1, (q_1^{EP}, p_1^{EP})$	$\theta_2 = 1$	$\theta_2 = 2$
$\theta_1 = 1$	$\frac{1}{6}$, 3, 3, (1,2)	$\frac{1}{6}$, 3, 3, (1,2)
$\theta_1 = 2$	$\frac{1}{6}$, 2, 1, (1,2)	$\frac{1}{6}$, 1, −1, (0,0)
$\theta_1 = 3$	$\frac{1}{6}$, 1, −1, (0,0)	$\frac{1}{6}$, 2, 1, (1,2)

We have $\Theta_1^*(q_1, \theta_2) = \{1,2\}$ if $(q_1, \theta_2) = (1,1)$ and $\Theta_1^*(q_1, \theta_2) = \{1,3\}$ if $(q_1, \theta_2) = (1,2)$. Hence, neither of these two sets is the subset of the other one, which never happens when we don't have ordinal interdependence. Now we construct a new detail-free mechanism as follows. When agent 1 reports $\theta_1 = a_1$ and first-order belief y, agent 1 obtains the goods by paying $(3 - \cos\eta)$ under b_1 and obtains the goods by paying $(3 - \sin\eta)$ under b_2, where $\eta = \arg\tan\frac{1-y}{y}$. We keep the optimal EPIC mechanism for both agents in the other cases. It is easy to check that the new mechanism is Bayesian incentive compatible over the universal type space. Since we assume full-support beliefs, that is, $y \in (0,1)$, then the payment from agent 1 is always strictly greater than 2, the optimal EPIC payment rule, under both b_1 and b_2. Thus the new mechanism raises strictly higher expected revenue than the optimal EPIC mechanism regardless of the designer's belief, resulting in no maxmin foundation for the EPIC mechanisms.

The two examples above identify some cases where revenue improvement is possible. Motivated by them, we define the concept of *improvability* as follows.

Definition 4 ("Improvability")**.** Revenue from i is improvable with respect to $(\theta_i, \theta_{-i}, \theta'_{-i})$ if there exists q_i and q'_i such that at least one of the following holds:

(i) $\theta_i \in \Theta_i^*(q'_i, \theta'_{-i}) \cap \Theta_i^*(q_i, \theta_{-i})$, and $\theta_i^*(q_i, \theta_{-i}) \notin \Theta_i^*(q'_i, \theta'_{-i})$, and $\theta_i^*(q'_i, \theta'_{-i}) \notin \Theta_i^*(q_i, \theta_{-i})$;

(ii) $\theta_i \in \Theta_i^*(q'_i, \theta'_{-i}) \setminus \Theta_i^*(q_i, \theta_{-i})$, and $\theta_i^*(q'_i, \theta'_{-i}) \in \Theta_i^*(q_i, \theta_{-i})$;

(iii) $\theta_i \in \Theta_i^*(q_i, \theta_{-i}) \setminus \Theta_i^*(q'_i, \theta'_{-i})$, and $\theta_i^*(q_i, \theta_{-i}) \in \Theta_i^*(q'_i, \theta'_{-i})$.

The examples essentially show that, given the optimal EPIC mechanism, if the revenue from some agent i is improvable with respect to some $(\theta_i, \theta_{-i}, \theta'_{-i})$, then the strong foundation does not exist. Thus, we complete the proof of Theorem 7 by showing that the ordinal interdependence necessarily implies the improvability. See the appendix for the formal proof.

Next, we study if EPIC mechanisms have the (not necessarily strong) foundation. The following example suggests that the same mechanism as above does not generally work, if the agents have non-full-support first-order beliefs.

Example 3. In the new mechanism proposed in Example 1, if we allow for non-full-support beliefs, there exists a situation where agent 1 always correctly predicts agent 2's payoff types. Formally, let $C = \{t \in T | \widehat{\theta}(t) = (1,1), \widehat{\pi}_1(t_1)[1] = 1\}, C' = \{t \in T | \widehat{\theta}(t) = (1,2), \widehat{\pi}_1(t_1)[2] = 1\}$, and consider μ such that $\mu(C) = f(1,1)$ and $\mu(C') = f(1,2)$. Because the optimal choice for agent 1 is $\eta^* = 0$ (or reporting $y = 1$ as his belief for $\theta_2 = 1$) if $t \in C$, and $\eta^* = \frac{\pi}{2}$ (or reporting $y = 0$) if $t \in C'$, the equilibrium payments in the new mechanism coincide with those in the optimal EPIC mechanism. Thus, without the full-support belief assumption, the new mechanism in Example 1 only *weakly* improves the expected revenue.

Now we further modify the mechanism as follows. Unless agent 1 reports $\theta_1 = 1$ and $y = 0$, the allocation is the same as the previous mechanism proposed in Example 1. If agent 1 reports $\theta_1 = 1$ and $y = 0$, then the following events happen: agent 1 does not buy the good for any θ_2, he pays $M(> 3)$ if $\theta_2 = 1$ (i.e., when his belief turns out to be "wrong"), and the principal offers price 3 for agent 2 (so that agent 2 buys only if $\theta_2 = 2$, i.e., when agent 1's belief turns out to be "right"), instead of price 2. It is easy to verify that the new mechanism is Bayesian incentive compatible on the universal type space $\mathscr{T}^*$.

This new mechanism achieves a weakly higher expected revenue than in the optimal EPIC mechanism. First, this weak improvement is obvious unless $\theta_1 = 1$ and $y = 0$. If $\theta_1 = 1$ and $y = 0$, the principal earns $M > 3$ from agent 1 if $\theta_2 = 1$ (while the optimal EPIC mechanism yields total revenue 3), and earns 3 from agent 2 if $\theta_2 = 2$ (while the optimal EPIC mechanism yields total revenue 2).

To show a strict improvement in expected revenue for any $\mu \in \mathscr{M}$, consider the case where $\theta_1 = 1$ and $\theta_2 = 2$. Because $f(1,2) > 0$, it suffices to show that, for any $y \in [0,1]$ reported by agent 1, the new mechanism achieves a strictly higher revenue than 2, the revenue in the optimal EPIC mechanism. First, as we see above, if $y = 0$ is reported, then the new mechanism yields 3 (from agent 2), and hence there is a strict improvement. If $y > 0$, then agent 2 pays 2, and agent 1 pays $1 - \sin(\arctan \frac{1-y}{y}) > 0$, and hence, there is again a strict improvement.

Notice that the key for strict improvement is to use agent 1's belief to modify the price for agent 2. If agent 1 is correct, such modification is profitable for the principal. Otherwise, the principal collects a "fine" from agent 1, which is also profitable.

As suggested in the example, it seems impossible to raise any additional revenue from an agent if he always correctly predicts the other agents' payoff types.[15] Instead, in such a case, a natural alternative idea may be to use this agent's prediction to raise additional revenue from the other agents (and to fine him if his prediction turns out to be wrong in order for the principal to "hedge", as in the example above). Because this means that we need to be able to change an agent's allocation without changing the others', we assume that the feasible allocation set Q is a product set, $Q = \prod_i Q_i$.

In addition, even if an agent correctly predicts the occurrence of some θ_{-i} (or its non-occurrence), such information does not necessarily make the principal earn strictly more revenue from the other agents (for example, imagine that any $j(\neq i)$'s virtual valuation is negative given θ_{-i}). Thus, we need a stronger version of the improvability.

Definition 5. We have the *strong improvability* if there exist $i, \theta_i, \theta_j, q_j, \theta_{-ij}$ such that $\theta_j \in \Theta_j^*(q_j, \theta_i, \theta_{-ij})$, and that revenue from i is improvable with respect to $\left(\theta_i, (\theta_j, \theta_{-ij}), (\theta_j^*(q_j, \theta_i, \theta_{-ij}), \theta_{-ij})\right)$.

Roughly, the strong improvability implies that, if agent i with θ_i correctly predicts that $-i$'s payoff types are not θ'_{-i}, then (given θ_{-ij}) the principal can know that j's type is not a threshold type for some q_j. Such information enables the principal to earn higher expected revenue from j.

Proposition 5. Under Assumptions 4, 5 and $Q = \prod_i Q_i$, strong improvability implies no foundation of EPIC mechanisms.

A natural question is, under which additional conditions, the ordinal interdependence implies the strong improvability, so that EPIC mechanisms do not have the foundation if and only if we do not have the ordinal interdependence. A sufficient condition is the following richness condition on Q.

Assumption 6. For each i, θ_i, and θ_{-i}, we have $q_i^{EP}(\theta_i, \theta_{-i}) > 0$, and for each $\theta'_i \neq \theta_i$, we have $q_i^{EP}(\theta_i, \theta_{-i}) \neq q_i^{EP}(\theta'_i, \theta_{-i})$.

A representative example is a monopoly problem with multiple buyers and multiple units of trading.[16] Note that this excludes some situations where the lowest payoff type of an agent (given the other agents' payoff types) is "excluded" from trading.

[15] See Yamashita.

[16] See Mussa and Rosen (1978) and Segal (2003) (or their straightforward generalizations) for such environments, although they focus on private-value environments.

Theorem 8. Under Assumptions 3, 6 and $Q = \prod_i Q_i$, EPIC mechanisms have the foundation if and only if we do not have ordinal interdependence.

2.5 Discussion: Continuous payoff-type space

In the previous finite payoff-type setup, given the others' payoff-type profiles, the difference in valuations between any two payoff types is strictly positive and bounded away from 0, which enables us to exploit the "gaps" in valuations and increase the payments. However, if we have infinitely many payoff types, such "gaps" may not exist, since η defined in Section 2.2 could be equal to 0, and then the previous construction no longer works.

In this sense, the continuous payoff-type space case is more complicated than the finite case, and hence the analysis of the continuous case is beyond the scope of the current paper.[17] Nevertheless, it may be interesting to note how our approach may be useful (with appropriate modifications), even in the continuous case. The following example explains this.

We assume $I = 2$, $\Theta_1 = \{0,1\} \times [0,2] (\ni (c_1,d_1))$, $\Theta_2 = \{0,1\} (\ni c_2)$, and $Q = \{0,1\}^2$. Agent 1's valuation for $q_1 = 1$ is $v_1(c_1,d_1,c_2) = c_1 c_2 + d_1$, and agent 2's valuation for $q_2 = 1$ is $v_2(c_2) = 1 + \frac{8}{7} c_2$.[18] One may interpret c_i as a (binary) common-value component, and d_1 as a private-value component for agent 1. Essentially, the only difference from the previous sections is that agent 1 now has a continuous payoff-type space. The other specifications are for simplicity.

For the principal's prior for θ, assume that each c_i takes 0 or 1 equally likely, independently from c_{-i}. Independently from c_2, the density of d_1 given c_1 is:

$$
\begin{aligned}
f(d_1 | c_1 = 0) &= \begin{cases} \frac{3}{4} & \text{if } \; d_1 \in [0,1], \\ \frac{1}{4} & \text{if } \; d_1 \in [1,2], \end{cases} \\
f(d_1 | c_1 = 1) &= \begin{cases} \frac{1}{4} & \text{if } \; d_1 \in [0,1], \\ \frac{3}{4} & \text{if } \; d_1 \in [1,2]. \end{cases}
\end{aligned}
$$

[17] Generalizing Theorem 6 to the continuous case (even in the private-value environment as in Chung and Ely (2007)) may also be non-trivial.

[18] We omit q_1, q_2 in the arguments of v_1, v_2 for brevity.

We can show that the optimal EPIC mechanism $(q_i^*, p_i^*)_{i=1,2}$ is given as follows:

$$
\begin{aligned}
q_1^*(c_1,d_1,c_2) &= \begin{cases} 1 & \text{if} \quad c_1c_2+d_1 \geq \frac{3}{4}c_2+1, \\ 0 & \qquad\text{otherwise,} \end{cases} \\
p_1^*(c_1,d_1,c_2) &= (\frac{3}{4}c_2+1)q_1^*(c_1,d_1,c_2), \\
q_2^*(c_1,d_1,c_2) &= \begin{cases} 1 & \text{if} \quad c_2=1, \\ 0 & \text{if} \quad c_2=0, \end{cases} \\
p_2^*(c_1,d_1,c_2) &= \frac{15}{7}q_2^*(c_1,d_1,c_2).
\end{aligned}
$$

This mechanism can be interpreted as a posted-price mechanism, where the price for agent 2 is always $\frac{15}{7}$ (so that only high-value type of agent 2 buys), and the price for agent 1 is $\frac{3}{4}c_2+1$, varying with c_2.

Our basic idea for improvement is very similar to the finite case in the previous section. However, to explain the basic incentive issues, we first consider the following "bundling" interpretation. Imagine that the seller is selling to agent 1 a right to obtain the good when $c_2=0$, and another right to obtain the good when $c_2=1$. To buy the bundle, agent 1 pays 1 when $c_2=0$ and $\frac{7}{4}$ when $c_2=1$, as in the optimal EPIC mechanism. Similarly, to buy only when $c_2=1$ (but not when $c_2=0$), agent 1 pays $\frac{7}{4}$ when $c_2=1$. However, to buy only when $c_2=0$ (but not when $c_2=1$), agent 1 pays $1-\varepsilon$ for a small $\varepsilon>0$.

If agent 1's purchase behavior is the same as in the optimal EPIC mechanism (in particular, if agent 1 buys both when $c_2=0$ and $c_2=1$ as long as $d_1 \geq 1$ and $c_1+d_1 \geq \frac{7}{4}$), then this new mechanism achieves a strictly higher expected revenue. However, such a behavior may not be incentive compatible. For example, if agent 1 believes that $c_2=0$ with probability very close to 1, then no payoff type of agent 1 would buy the bundle: even for the highest payoff-type (i.e., $(c_1,d_1)=(1,2)$), it would be better to buy the good only when $c_2=0$ (with price $1-\varepsilon$) than to buy the good for both $c_2 \in \{0,1\}$. Such a deviation makes the expected revenue much smaller than under the optimal EPIC mechanism.

To avoid this, we introduce the following side bet: if (and only if) agent 1 buys the bundle, he can further buy a lottery that yields to the agent ε if $c_2=0$, and $-b\varepsilon$ if $c_2=1$ (for some $b \in (\frac{8}{7},\frac{9}{7})$). Furthermore, if agent 1 buys this lottery, then with probability ε, the principal offers price 1 instead of $\frac{15}{7}$ to agent 2 (and the principal continues to offer price $\frac{15}{7}$ to agent 2 with the other probability $1-\varepsilon$).

Then, as long as $d_1 \geq 1$ and $c_1+d_1 \geq \frac{7}{4}+b\varepsilon$, agent 1 prefers to "buying the bundle and the lottery" to "buying only when $c_2=0$", regardless of his belief. Whether or not he actually buys the lottery depends on his belief. However, observe that regardless of the true state, the expected revenue of the principal from the lottery is always non-negative: if $c_2 = 0$, the principal pays ε to agent 1 while he earns additional revenue 1 from agent 2 with probability ε (hence, the expected revenue gain is non-negative); if $c_2 = 1$, the principal receives $b\varepsilon$ from agent 1 while he loses revenue $\frac{8}{7}$ from agent 2 with probability ε (hence, the expected revenue gain is non-negative for $b > \frac{8}{7}$). Therefore, the worst-case scenario for the principal is that agent 1 never buys the lottery.

On the other hand, if $d_1 > 1$ and $c_1 + d_1 \in (\frac{7}{4}, \frac{7}{4} + b\varepsilon)$, the worst-case scenario is that agent 1 buys only when $c_2 = 0$, even though he buys both for $c_2 \in \{0,1\}$ in the optimal EPIC mechanism. This revenue loss occurs regardless of agent 1's belief, and this is one of the fundamental differences from the (generic) finite case where only the gain exists as long as the agents have full-support first-order beliefs.

Nevertheless, the overall expected revenue change is strictly positive, at least for sufficiently small ε, which is approximately:

$$\begin{aligned}
& (1-\varepsilon)\Pr(d_1 \in (1-\varepsilon, 1),\ c_1 + d_1 < \frac{7}{4})\Pr(c_2 = 0) \\
& -\varepsilon \Pr(d_1 > 1,\ c_1 + d_1 < \frac{7}{4})\Pr(c_2 = 0) \\
& -\Pr(d_1 > 1, c_1 + d_1 \in (\frac{7}{4}, \frac{7}{4} + b\varepsilon))\Pr(c_2 = 1) \\
\simeq\ & \frac{3\varepsilon}{16} - \frac{3\varepsilon}{64} - \frac{7b\varepsilon}{64},
\end{aligned}$$

which is positive if $b < \frac{9}{7}$. The first term (on the left-hand side) is because agent 1 whose payoff type satisfies $d_1 \in (1-\varepsilon, 1)$ and $c_1 + d_1 < \frac{7}{4}$ does not buy in any state in the optimal EPIC mechanism, while he buys when $c_2 = 0$ in the modified mechanism. The second term is because, for agent 1 whose payoff type satisfies $d_1 > 1$ and $c_1 + d_1 < \frac{7}{4}$, the price he pays in the modified mechanism (when $c_2 = 0$) is smaller by ε. The third term is because agent 1 whose payoff type satisfies $d_1 > 1$ and $c_1 + d_1 \in (\frac{7}{4}, \frac{7}{4} + b\varepsilon)$ buys both for $c_2 \in \{0,1\}$ in the optimal EPIC mechanism, while he buys only when $c_2 = 0$ in the modified mechanism.[19]

This is, of course, just one example, and whether a similar approach works more generally is left to be determined. However, we believe that, as demonstrated in this example, our basic idea of modifying mechanisms carries over even to some continuous environments.

[19] There are other changes in agent 1's behavior, but their effects on the expected revenue are $o(\varepsilon)$, and hence omitted.

2.6 Conclusion

If the environment exhibits only cardinal interdependence (and certain regularity conditions), then there exist the maxmin and Bayesian foundations for EPIC mechanisms, in the sense of Chung and Ely (2007). If the environment exhibits ordinal interdependence, (and certain additional conditions), then such a foundation may not exist.

In interdependent-value environments, Yamashita (2015) provides an alternative solution concept (that is, *incentive compatibility in value revelation*), which is also robust to the agents' belief structure in a related sense and useful in the implementation of social choice correspondences in undominated strategies. It may be interesting to investigate similar sorts of foundation results for this alternative solution concept.

Chapter 3

Dynamic Inconsistency In Collective Decision[1]

Shuguang Zhu[2]

October 30, 2017

We show that dynamic inconsistency in collective decision-making can derive from heterogeneity in group members' outside options, even if individuals share the same exponentially discounting time preference. This model of endogenous dynamic inconsistency facilitates the analysis of welfare consequences, since time-consistent individual preferences allow for a well-defined measurement of social welfare. We further characterize the optimal Bayesian-persuasion information disclosure policy, which takes the form of upper revealing rules, to alleviate the welfare distortion caused by inconsistent collective decisions. Our framework proves to be highly adaptable to various contexts, including provision of public facilities and assignment on team work.

[1]I am indebted to Takuro Yamashita for his inspiration and guidance at every stage of my research work. I also thank Alessandro Bonatti, Tilman Börgers, Alexandre de Cornière, João Correia-da-Silva, Jacques Crémer, Christian Gollier, Renato Gomes, Johannes Hörner, and seminar participants at Asian Meeting of the Econometric Society (2017) and China Meeting of the Econometric Society (2017) for extremely valuable comments and suggestions on this project.

[2]Toulouse School of Economics, University of Toulouse Capitole, France. shuguang.zhu@tse-fr.eu

3.1 Introduction

Collective decisions have drawn increasing scholarly attention because of the sophisticated processes of aggregating individual preferences. Many economic decisions, such as household savings and consumption, board decisions and public goods provision, are essentially intertemporal choices of consumption streams made by a group of decision makers. If group members are heterogeneous in their *outside options* — opportunity costs that individuals have to pay in order to join the group — then collective decision-making can exhibit *dynamic preference reversal*[3] under certain conditions. We establish a framework to explain this phenomenon, and explore how dynamic inconsistency is linked to the distribution of group members' outside options. We also study the related welfare consequences, and characterize the optimal Bayesian-persuasion mechanism to restore efficiency.

Classical economics usually treats firms and other organizations as time-consistent agents who discount the future reward in an exponential manner; however, this way of modeling has been questioned since economists recognize that heterogeneity of individuals could induce time varying discount rates (e.g., Marglin, 1963; Feldstein, 1964; Becker, 1992). In the recent two decades, time inconsistency has been studied by a fast-growing literature of behavioral economics. From (quasi-)hyperbolic discounting (e.g., Loewenstein and Prelec, 1992; Laibson, 1997) and non-hyperbolic discounting (e.g., Bleichrodt, Rohde, and Wakker, 2009), to multi-selves assumption (e.g., Fudenberg and Levine, 2006), behavioral economics successfully proves itself to be a powerful angle to test and explain time inconsistency of individual preference.[4] However, it is lacking in foundations to model inconsistent collective decisions by using behavioral assumptions which are mainly suitable for individual preferences. Thus, novel frameworks have to be developed for this direction of research.

A natural way is to enrich the structure of collective decision-making by introducing the group leader (i.e., the Principal) whose objective represents the collective utility function, and the group members (i.e., the Agents) with the standard exponential time preference. By checking the group leader's ranking of different temporal rewards, one can identify which

[3] That is, as the evaluation period moves forward, the decision maker would have an incentive to deviate from her plan made in the early periods. A relative concept is *static preference reversal*, which violates the stationarity studied in decision literature (Koopmans, 1960; Fishburn and Rubinstein, 1982). Particularly, the ranking of two temporal payments is not pinned down by the difference in size and the relative delay, but also depends on the distance from the evaluation period. Halevy (2015) has studied the relationship between these two types of preference reversal.

[4] Also, its applications to contract theory, mechanism design and public policy seem to be fruitful (e.g., O'Donoghue and Rabin, 1999; Galperti, 2015; Bisin, Lizzeri, and Yariv, 2015).

factors give rise to time inconsistency. Gollier and Zeckhauser (2005), followed by Jackson and Yariv (2014, 2015), have showed that heterogeneity in individual discount rates makes collective decisions exhibit *present bias*.[5] In this paper, we follow their two-layer setup of collective decision with common consumption; while instead of assuming different discount rates among group members, we highlight the heterogeneity of opportunity costs of joining the group. We prove that if the principal lacks commitment devices, collective decisions exhibit *dynamic present bias* if and only if agents have heterogeneous outside options (Theorem 9). More precisely, as long as the principal initially prefers a later consumption, she would also choose it even if the decision time is postponed; we can find a pair of consumptions such that the principal initially prefers the later one to the earlier one, but when the decision time is postponed, her preference gets reversed.

To illustrate this dynamic preference reversal, imagine that a utilitarian social planner wants to construct the road system for a new community, which is expected to take two years. While she can save one year of construction by cutting some branches and auxiliary facilities. Due to pre-construction community planning, at the beginning households have to decide whether to live in this area or to move out to get their yearly outside options[6] elsewhere. Participation is voluntary but reentry is not allowed.

Cutting the construction period does reduce the value of the road system, however, it saves households' waiting costs before the construction ends. Moreover, high-outside-option households prefer the one-year plan because it is inefficient to wait that long in the two-year plan; while low-outside-option households prefer the latter since it generates enough surplus to compensate for their opportunity costs. Under certain conditions the group leader should select the two-year plan which excludes high-outside-option households, in order to maximize

[5] Also known as decreasing rate of impatience. That is, decision makers prefer smaller immediate reward to larger delayed reward; while the ranking is reversed when both alternatives are equally shifted into the future. Gollier and Zeckhauser (2005) consider allocations of private consumptions across agents in each period, and point out that the variation of each individual's share of resources over time, which equalizes individual intertemporal marginal rates of substitution, is the driving force of inconsistency. While Jackson and Yariv (2014, 2015) work on common consumption streams, where the planner allocate public goods over time, and find that the difference in individuals' exponentially discounting rates changes the relative effect of individuals' preferences on the weighted average utility, and thus induces present bias. A common feature of these works is that they focus on the aggregate preferences over temporal rewards and exclude any dynamic interactions between the group leader and group members. In other words, they study static preference reversal.

[6] That is, the highest possible utility each household can obtain by living outside of this community each year. It is natural to think that outside options are different among households because of the variety of income levels and tastes for environment.

social welfare[7]. However, since the social planner cannot make any credible announcement about her future decision, high-outside-option households won't move out of the community, and their presence forces the social planner to take their interests into consideration and end up with insufficient public facilities.

The above example explains how heterogeneous outside options generate dynamic inconsistency. The difference in outside options not only creates tensions of individual preferences as heterogeneity of discounting factors does, but also entitles group members to nontrivial participation decisions.[8] Due to lack of commitment power, agents can affect the principal's future preference through their current participation decisions, resulting in an inefficient outcome from the group leader's initial perspective.[9]

We prove that such dynamic preference reversal arises when the *liquidation rate* — the ratio of earlier consumption to later consumption — is at the intermediate level (Theorem 10). Clearly, no dynamic preference reversal would occur in extreme cases. If the difference in valuations between two consumptions is negligible, then by no means should the principal take more time to wait for the later consumption. On the other hand, if the earlier consumption generates too little surplus compared with the later one, then the earlier consumption is always an inferior alternative. We also show that the range of liquidation rates that induce dynamic preference reversals is determined by the distribution of agents' outside options and the size of consumption (Proposition 6, 7).

Our framework is quite suitable for analyzing the welfare consequences of dynamic time inconsistency, because group members share the same exponential discounting preference, which allows for a well-defined and comparable measurement of social surplus. Since efficient outcomes are not implemented due to dynamic preference reversals for intermediate liquidation rates, we're interested in whether the principal can improve the total welfare by manipulating how agents learn about liquidation rates. We base our methodology on the Bayesian persuasion literature (e.g., Rayo and Segal, 2010; Kamenica and Gentzkow, 2011), and assume that the group leader can commit to a *public disclosure* rule on the liquidation rate

[7] Since the social planner's objective is to maximize the total utility of the all households, including those who move out, it is equivalent to maximize the net social surplus generated by the road system.

[8] Unlike Jackson and Yariv (2014), where agents always get positive surplus in the group, in our paper group members may optimally choose to quit the group under voluntary participation.

[9] The key insight that inconsistency arises in multi-period decisions when decision makers have limited commitment power, is closely related to Kydland and Prescott (1977), that is, there is no mechanism to induce the future decision maker to take into consideration the influence of her action on current decisions of group members with rational expectations. See Thomas and Worrall (1988) and Kocherlakota (1996) for more discussions on the commitment issues.

which is not directly observable by agents. We prove that the optimal information structure is a *censorship* policy (e.g., Kolotilin, Li, Mylovanov, and Zapechelnyuk, 2015; Yamashita, 2016), which completely reveals any realization of the liquidation rate higher than a threshold, while conceals it otherwise.

The economic intuition that induces the optimality of upper revealing rule is basically to synthesize both commitment and flexibility. Particularly, by pooling some intermediate liquidation rates with low realizations, on receiving the pooling message group members will form a posterior about the liquidation rate which secures the implementation of default option. This censorship policy plays the similar role as commitment power does, to deter the participation of high-outside-option group members in order to achieve the efficient outcome under these intermediate liquidation rates. On the other hand, for the remaining intermediate liquidation rates that cannot be included in the pooling message, the disclosure rule completely reveals these realizations so that group members can adjust their participating decisions to achieve the second-best outcome.

The suboptimality of complete information revelation from the sender's perspective is aligned with the literature on information disclosure (e.g., Morris and Shin, 2002; Angeletos and Pavan, 2007; Ganuza and Penalva, 2010; Kamenica and Gentzkow, 2011). In these papers, the optimality of partial disclosure rule mainly derives from either preference conflicts between the sender and receivers, or the frictions in games. Our paper differs from the literature by providing another incentive for the utilitarian sender to conceal certain information, that is, partial disclosure policy makes up for the lack of commitment devices which would otherwise induce dynamic inconsistency issues.

The remainder of our paper is organized as follows. Section 3.2 introduces the setup. Section 3.3 shows how dynamic inconsistency derives from the heterogeneity of agents' outside options. Section 3.4 characterizes when dynamic preference reversals occur, and studies the resource size effect and group composition effect. Section 3.5 characterizes the optimal Bayesian-persuasion information disclosure rule to restore efficiency. Section 3.6 rationalizes the equilibrium refinement adopted in our model, and provides two applications to public facility provision and team work. Section 3.7 concludes with a brief discussion on combining our framework with mechanism design. All omitted proofs are collected in the appendices.

3.2 The Model

We consider a discrete-time model, with $t \in \{0, 1, 2, \ldots\}$. Let C_t^x represent a common consumption (i.e. a public good) of x at time t. A group of risk-neural agents, with mass of 1 labelled by $i \in I = [0,1]$, make a collective decision on which common consumption to implement in the group. We formalize the collective decision-making process by introducing a utilitarian group leader (i.e. the principal, denoted by "She"; while "He" denotes the agent), who aims to maximize the total utility of all agents. The principal's objective is called the *collective utility function*.

3.2.1 Agents' actions and payoffs

We assume that agents share the same exponential discounting time preference with discount factor normalized to be 1. The instantaneous utility function at time t for each agent i is $u_{i,t}(x) = x$. At the beginning of each period t, agents simultaneously choose their actions $a_t = (a_{i,t})_{i \in I} \in \prod_{i \in I} A_{i,t} \subseteq \{0,1\}^I$, where 1 means that the agent participates in the group, while 0 means that the agent stays out of the group. Assume that all agents are initially in the group; and that once an agent quits the group, he has no option for reentry, that is, $A_{i,t} \equiv \{0\}$ for all $t > \bar{t}$ if $a_{i,\bar{t}} = 0$.

Each agent i has a deterministic time-invariant per-period outside option $w_{i,t} = w(i) : [0,1] \longmapsto [0,W]$, which serves as the opportunity cost of participating in the group. Thus, the cumulative distribution function of agents' outside options is given by

$$F(w) = \int_{w(i) \leq w} di, \text{ for any } w \in [0,W].$$

Define $\alpha := \sup\{x \mid F(x) = 0\}$ and $\beta := \inf\{x \mid F(x) = 1\}$. If $\alpha < \beta$, we can find two subsets $I_1, I_2 \subseteq I$ with strictly positive measure, such that $w(i_1) \neq w(i_2)$ for any $i_1 \in I_1$ and any $i_2 \in I_2$. If $\alpha = \beta$, all agents share the same outside option, except for those who belong to a measure-zero subset of I. We say that agents have *heterogeneous* outside options if and only if $\alpha < \beta$.

From the time-t_0 perspective, if an agent with outside option w expects to have common consumption C_t^x (with $t > t_0$), by quitting at time $\bar{t}$, his net payoff is given by

$$U_{i|t_0}(C_t^x, w \mid \bar{t}) = \begin{cases} -w(\bar{t} - t_0), & \text{if } t_0 \leq \bar{t} \leq t \\ x - w(\bar{t} - t_0), & \text{if } \bar{t} \geq t + 1. \end{cases}$$

Obviously, the agent's optimal quitting time is given by

$$\bar{t}^*(C_t^x, w, t_0) = \begin{cases} t_0, & \text{if } x < w(t - t_0 + 1) \\ t + 1, & \text{if } x \geq w(t - t_0 + 1), \end{cases}$$

subject to a tie-breaking rule that when agents get the same net payoff between quitting immediately and staying until the public good is consumed, they choose the latter. Then the highest outside option of agents who don't quit at time t_0 is given by $\bar{w}(C_t^x, t_0) = \frac{x}{t-t_0+1}$. The time-$t_0$ valuation of C_t^x for an agent with outside option w is

$$U_{i|t_0}(C_t^x, w) = \begin{cases} 0, & \text{if } x < w(t-t_0+1) \\ x - w(t-t_0+1), & \text{if } x \geq w(t-t_0+1). \end{cases}$$

Obviously, individual agent's decision is dynamic consistent.

Lemma 9. For any $0 \leq t_0 < t_0' \leq t < t'$, any w, and any x, x', we have

$$U_{i|t_0}(C_t^x, w) \left\{ \begin{array}{c} > \\ < \end{array} \right\} U_{i|t_0}(C_{t'}^{x'}, w) \quad \Longrightarrow \quad U_{i|t_0'}(C_t^x, w) \left\{ \begin{array}{c} > \\ < \end{array} \right\} U_{i|t_0'}(C_{t'}^{x'}, w).$$

Proof. Equivalently, we write $U_{i|t_0}(C_t^x, w) = \max\{0, x - w(t-t_0+1)\}$. If $U_{i|t_0}(C_t^x, w) > U_{i|t_0}(C_{t'}^{x'}, w)$, then we must have $x - w(t-t_0+1) > \max\{0, x' - w(t'-t_0+1)\}$. Notice that $x - w(t-t_0'+1) > x - w(t-t_0+1) > 0$ and

$$[x - w(t-t_0'+1)] - [x' - w(t'-t_0'+1)] = [x - w(t-t_0+1)] - [x' - w(t'-t_0+1)] > 0,$$

then $U_{i|t_0'}(C_t^x, w) = x - w(t-t_0'+1) > \max\{0, x' - w(t'-t_0'+1)\} = U_{i|t_0'}(C_{t'}^{x'}, w)$. □

3.2.2 Principal's problem

There are two common consumptions available for the group, denoted by $\{C_{t_1}^x, C_{t_2}^y\}$ with $t_1 < t_2$. Let $I_t^* = \{i \in I \mid a_{i,t-1} = 1\}$ denote the subset of *active* agents who have not quit the group before time t. The valuation of common consumption C_t^x at time t_0 for the principal is given by

$$U_{P|t_0}(C_t^x \mid I_{t_0}^*) = \int_{i \in I_{t_0}^*} U_{i|t_0}\big(C_t^x, w(i)\big) di. \tag{3.1}$$

In other words, the collective decision is made to maximize the total net payoff of the active agents. We assume that the principal has *no commitment device* to convince the agent of her future choice *until* time t_d such that $0 < t_d \leq t_1$. The time between t_d and t_1 might be interpreted as the necessary period of preparation for implementing the common consumption $C_{t_1}^x$. The principal essentially commits to implementing $C_{t_2}^y$ by not starting preparing for $C_{t_1}^x$ at time t_d. It is critical to assume $t_d > 0$; otherwise the principal could commit herself to her time-0 decision and be free from dynamic inconsistency issues.

3.2.3 Strategies and solution concept

We assume that agents' outside options and all above settings are common knowledge within the group. Moreover, all past actions are observable by agents and the principal. The timing is as follows:

(i) At the beginning of each period $0,1,\ldots,t_d-1$, active agents simultaneously choose whether to stay in the group or not.

(ii) At the beginning of period t_d (before agents make their participation decisions), the principal chooses $C\in\{C^x_{t_1},C^y_{t_2}\}$ after observing $I^*_{t_d}$.

(iii) After observing C, agents decide their optimal quitting time.

We can see that the game essentially ends after the principal chooses C, because in stage (iii) each active agent i's optimal quitting time is simply given by $\bar{t}^*\big(C,w(i),t_d\big)$. Thus, we only need to study the finite-period game from time 0 to time t_d.

A pure strategy for agent i is a mapping $s_i: H\longmapsto\{0,1\}$, where $H=\bigcup_{t=0}^{t_d-1}\left(\prod_{i\in I}H^t_i\right)$ and $H^t_i=\{0,1\}^t$. An element of H^t_i is given by $h^t_i=(a_{i,0},a_{i,1},\ldots,a_{i,t-1})$, which denotes agent i's past actions before time t. An element of H is given by $h^t=(a_0,a_1,\ldots,a_{t-1})$, which denotes agents' past action profiles before time t. Let S_i collect agent i's all feasible strategies that satisfy the no-reentry condition.

A pure strategy for the principal is a mapping $s_P:\{0,1\}^{t_d\times I}\longmapsto\{C^x_{t_1},C^y_{t_2}\}$. From (3.1) we can see that the principal's optimal choice only depends on the subset of active agents at the end of period t_d-1, i.e. $I^*_{t_d}$. Thus, the principal's strategy can also be written as $s_P: 2^I\longmapsto\{C^x_{t_1},C^y_{t_2}\}$. Let S_P collect the principal's all possible strategies.

The solution concept we applied here is *pure-strategy subgame perfect equilibrium without coordination failure among agents*, which is abbreviated to SPE-NCF hereafter. Define $U_{i|0}\big((s_i)_{i\in I},s_P,w(i)\big)=U_{i|0}\big(s_P(h^{t_d}[(s_i)_{i\in I}]),w(i)\big)$, where $h^{t_d}[(s_i)_{i\in I}]$ denotes the past actions before time t_d induced by strategy profile $(s_i)_{i\in I}$.

Definition 6. We say that $\big((s^*_i)_{i\in I},s^*_P\big)$ constitute a SPE-NCF strategy profile if:

1. $\big((s^*_i)_{i\in I},s^*_P\big)$ is a subgame perfect equilibrium;

2. For any $\tilde{I}\subseteq I$ and any $(\tilde{s}_i)_{i\in\tilde{I}}\in\prod_{i\in\tilde{I}}S_i$, we have for any $i\in\tilde{I}$,

$$U_{i|0}\big((s^*_i)_{i\in I},s^*_P,w(i)\big)\geq U_{i|0}\big((\tilde{s}_i)_{i\in\tilde{I}},(s^*_i)_{i\in I\setminus\tilde{I}},s^*_P,w(i)\big).$$

The only difference between our solution concept and the standard subgame perfect equilibrium is that, we allow agents to form coalitions, so that they can jointly decide whether to stay in the group or not. This equilibrium refinement is critical for our analysis, because it induces a unique outcome by ruling out the "uninteresting" equilibriums with coordination failure among agents.

To see this, if we have $x < y$, then agents with higher outside options would prefer the earlier consumption $C^x_{t_1}$, while agents with lower outside options would prefer the later consumption $C^y_{t_2}$. Since the principal maximizes the group's total utility, the presence of agents with higher (or lower) outside options will dispose the principal to choose earlier (or later) consumption at time t_d. Roughly speaking, coordination failure among agents refers to the situation where the joint participation of a subset of agents is able to drive the principal to choose their preferred outcome, while the participation of only a fraction of them will end up with a less-preferred outcome for them. This is often the case when individual agents have limited effect on the equilibrium outcome, as is in our model with a continuum of agents.[10]

3.3 Dynamic Inconsistency: Present Bias

Jackson and Yariv (2014) define a static version of present-biased collective utility functions, where the collective-decision maker evaluates future consumptions from a fixed-time perspective. Then present bias in their definition refers to a situation where the collective-decision maker becomes more patient as the candidate pair of consumptions are both postponed. Motivated by this, we define the dynamic present bias:

Definition 7 ("Dynamic Present Bias"). Let $V_{t_0}(C^x_t) : \{0,1,\dots\}^2 \times \mathbb{R} \longmapsto \mathbb{R}$ be a collective utility function which defines the time-t_0 valuation of C^x_t for the principal. Then V exhibits dynamic present bias if for all $0 < t_0 \leq t_1 < t_2$:

1. For all x and y, $V_0(C^x_{t_1}) \geq V_0(C^y_{t_2})$ implies $V_{t_0}(C^x_{t_1}) \geq V_{t_0}(C^y_{t_2})$;

2. There exist x and y such that $V_0(C^x_{t_1}) < V_0(C^y_{t_2})$ and $V_{t_0}(C^x_{t_1}) > V_{t_0}(C^y_{t_2})$.

A notable feature that distinguishes our definition from the one in Jackson and Yariv (2014) is that the principal (i.e. the collective-decision maker) no longer stands at a fixed time. The first part of the definition says that if the principal initially prefers an earlier consumption to a later one, then as the decision time moves on, her preference doesn't change. The second part

[10]See Subsection 3.6.1 for more discussions on the solution concept.

states that, we can find a pair of consumptions such that initially the principal prefers the later consumption, but as the decision time approaches the earlier consumption, her preference gets reversed. Thus, dynamic present bias indicates increasing impatience as the decision time is postponed.

We consider two scenarios. Scenario 1 is a hypothetical situation where the principal makes a decision at the beginning of time 0 as if she could commit to it during the whole game. After observing the principal's time-0 decision, each agent chooses his own quitting time. Because initially agents are all in the group, i.e $I_0^* = I$, the principal's optimal time-0 choice is given by

$$\max_{C \in \{C_{t_1}^x, C_{t_2}^y\}} U_{P|0}(C) = \int_{i \in I} U_{i|0}\big(C, w(i)\big) di,$$

which is actually the efficient outcome. Let C^{eff} be the solution to this problem.

Scenario 2 is the actual game defined in Section 3.2, where the principal's decision time is postponed to time t_d. Given agents' equilibrium strategy profile $s^* = (s_i^*)_{i \in I}$, which results in the subset of active agents, denoted by $I_{t_d}^*(s^*)$, the principal's optimal time-t_d choice is given by

$$\max_{C \in \{C_{t_1}^x, C_{t_2}^y\}} U_{P|t_d}\big(C \mid I_{t_d}^*(s^*)\big) = \int_{i \in I_{t_d}^*(s^*)} U_{i|t_d}\big(C, w(i)\big) di,$$

which is named the equilibrium outcome C^{equ}. A collective decision rule is said to be *dynamic inconsistent* if there exists discrepancy between time-0 choice (C^{eff}) and time-t_d choice (C^{equ}). The following theorem proves that (utilitarian) collective decisions exhibit dynamic present bias if and only if agents have heterogeneous outside options.

Theorem 9. Fixed any $F \in \Delta([0,W])$ and any $0 < t_d \le t_1 < t_2$:

1. For all x and y, $C^{\text{eff}} = C_{t_1}^x$ implies $C^{\text{equ}} = C_{t_1}^x$;

2. There exist x and y such that $C^{\text{eff}} = C_{t_2}^y$ and $C^{\text{equ}} = C_{t_1}^x$, if and only if $\alpha < \beta$.

We provide the formal proof in Appendix D.1.1. Here we illustrate how heterogeneity in outside options could result in dynamic present-biased collective decisions by a three-period binary-agent model. Time is denoted by $\{0,1,2\}$, where 3 means the end of period 2. The principal's decision time is $t_d = 1$, with two candidate common consumptions $\{C_1^{0.85}, C_2^1\}$. Let $I = \{1,2\}$. Agents' outside options are given by $w(1) = 0$ and $w(2) = 0.4$. Obviously, Agent 1 prefers C_2^1 to $C_1^{0.85}$; while Agent 2 prefers consuming $C_1^{0.85}$ to quitting at time 0 (if C_2^1 is implemented). The principal maximizes the sum of both agents' payoffs. On the one hand, because

$$
\begin{aligned}
U_{P|0}(C_1^{0.85}) &= \underbrace{U_{1|0}(C_1^{0.85} \mid 2)}_{\text{Agent 1 quits at time 2}} + \underbrace{U_{2|0}(C_1^{0.85} \mid 2)}_{\text{Agent 2 quits at time 2}} \\
&= (0.85 - 0 \times 2) + (0.85 - 0.4 \times 2) = 0.9, \\
U_{P|0}(C_2^1) &= \underbrace{U_{1|0}(C_2^1 \mid 3)}_{\text{Agent 1 quits at time 3}} + \underbrace{U_{2|0}(C_2^1 \mid 0)}_{\text{Agent 2 quits at time 0}} \\
&= (1 - 0 \times 3) + 0 = 1,
\end{aligned}
$$

we get $C^{\text{eff}} = C_2^1$. On the other hand, if both agents do not quit at time 0, then at the principal's decision time $t_d = 1$, we have

$$
\begin{aligned}
U_{P|1}\big(C_1^{0.85} \mid I_1^* = \{1,2\}\big) &= \underbrace{U_{1|1}(C_1^{0.85} \mid 2)}_{\text{Agent 1 quits at time 2}} + \underbrace{U_{2|1}(C_1^{0.85} \mid 2)}_{\text{Agent 2 quits at time 2}} \\
&= (0.85 - 0 \times 1) + (0.85 - 0.4 \times 1) = 1.3, \\
U_{P|1}\big(C_2^1 \mid I_1^* = \{1,2\}\big) &= \underbrace{U_{1|1}(C_2^1 \mid 3)}_{\text{Agent 1 quits at time 3}} + \underbrace{U_{2|1}(C_2^1 \mid 3)}_{\text{Agent 2 quits at time 3}} \\
&= (1 - 0 \times 2) + (1 - 0.4 \times 2) = 1.2.
\end{aligned}
$$

Thus, in the subgame induced by $I_1^* = \{1,2\}$, the principal chooses $C_1^{0.85}$. Notice that both agents get strictly positive payoffs in this subgame, then we get a unique SPE (and thus a unique SPE-NCF) where both agents stay in the group at time 0 and $C_1^{0.85}$ is implemented at time 1, i.e. $C^{\text{equ}} = C_1^{0.85}$.

Intuitively, it is too costly to include Agent 2 (i.e. the high-outside-option agent) in the group. If the principal could credibly announce her future choice at time 0, she would lead Agent 2 to quit at the very beginning by committing to C_2^1. However, due to the lack of commitment power at time 0, the presence of Agent 2 in the group compels the principal to implement $C_1^{0.85}$. We can see that heterogeneity in outside options affects the principal's decision in two ways: (i) it creates conflicts of individual preferences; (ii) it induces different quitting decisions among agents. This enables agents to alter the principal's choice by their non-trivial participation decisions. Also, agents' voluntary participation is indispensable. Suppose instead the principal could force Agent 2 to quit at time 0, then with Agent 1 to be the only active agent, the principal's equilibrium choice at time 1 would be C_2^1, which coincides with the efficient choice.

Theorem 9 contributes to the study on inconsistency in collective decision-making, by providing a foundation for dynamic present bias based on time-consistent individual preferences.

In behavioral economics, time inconsistency is usually modeled in an exogenous way, such as quasi-hyperbolic discounting. In the above three-period example, assume that the collective utility function at time t_0 is given by

$$V_{t_0}(c_{t_0}, c_{t_0+1}, c_{t_0+2}) = c_{t_0} + r_1\left(r_2 c_{t_0+1} + r_2^2 c_{t_0+2}\right),$$

where c_t is the consumption at time t, and $r_1, r_2 \in (0,1)$. One can easily check that if $r_2 > 0.85$ and $r_1 r_2 < 0.85$, then we have $V_0(C_1^{0.85}) < V_0(C_2^1)$ and $V_1(C_1^{0.85}) > V_1(C_2^1)$. However, our model reproduces the pattern of dynamic present bias that can be explained by behavioral models, without imposing "non-standard" assumptions on individual preferences. Moreover, we show that such dynamic inconsistency in collective decision-making is resulted from the heterogeneity in agents' outside options.

It is worth noting that, dynamic present bias guarantees that preference reversals between efficient choice and equilibrium choice only occur for *some* pair of candidate consumptions. For example, if the later consumption is smaller than the earlier one, both efficient and equilibrium choices will be the earlier larger consumption. Thus, it is the relative size of candidate consumptions that matters. In the next section, we characterize when such preference reversal happens by using a three-period model where the principal reallocates common consumptions over time.

3.4 Dynamic Preference Reversal

We consider the same setup as in Section 3.2, except that time is restricted to $\{0,1,2\}$ and the principal's decision time is $t_d = 1$. The utilitarian principal has resource y at time 2, which is public goods. At time 1, the principal has access to a debt plan with borrowing rate $\delta \in [0,1]$, which measures the effectiveness of reallocating resources from later periods to earlier periods. If the principal decides to reallocate a fraction $\gamma \in [0,1]$ of the total resource from time 2 to time 1, then she can provide agents with a common consumption stream $(c_1, c_2) = \big(\gamma\delta y,\, (1-\gamma)y\big)$. We assume that at time 0, the principal has no commitment device to convince the agents of her time-1 choice.

We assume that agents have heterogeneous outside options, so that the collective decision is present-biased. For simplicity, we assume that $F(w) \in \Delta([0,W])$ has full-support continuous density. We first characterize when dynamic preference reversal arises in the benchmark case where δ is common knowledge and realized at time 0. The timing is the same as in Subsection 3.2.3, except that at stage (ii) the principal chooses from $\left\{\big(\gamma\delta y,\, (1-\gamma)y\big)\right\}_{\gamma\in[0,1]}$

instead of binary choices. Lemma 10 proves that at stage (ii), the utilitarian principal actually chooses between $\gamma = 0$ and $\gamma = 1$, i.e. the default option C_2^y or the liquidation option $C_1^{\delta y}$.[11]

Lemma 10. Given any $I_1^* \subseteq I$, principal's optimal choice at time 1 is either $\gamma = 0$ or $\gamma = 1$.

Next we study the condition on δ to generate discrepancy between principal's efficient choice at time 0 and equilibrium choice at time 1. Let $x = \delta y (\leq y)$. Anticipating $C \in \{C_1^x, C_2^y\}$ at time 1, agents participate at time 0 only if $U_{i|0}(C, w(i)) \geq 0$. It follows that the highest participating outside option at time 0 is $\bar{w}(C_1^x, 0) = \frac{x}{2}$ (or $\bar{w}(C_2^y, 0) = \frac{y}{3}$) if principal chooses C_1^x (or C_2^y) at time 1.

Case 1. $\frac{x}{2} \leq \frac{y}{3}$. Immediately, we have $\sup_{i \in I_1^*} w(i) \leq \frac{y}{3}$. From (D.1) in Appendix D.1.1 we have that all agents who stay in the group until the public good is consumed always prefer C_2^y to C_1^x. Thus, $C^{\text{eff}} = C^{\text{equ}} = C_2^y$. In other words, there is no dynamic preference reversal when $\delta \leq \frac{2}{3}$.

Case 2. $\frac{x}{2} > \frac{y}{3}$. Define $I(w) = \{i \in I \mid w(i) \leq w\}$, then we have $I(\frac{y}{3}) \subseteq I_1^* \subseteq I(\frac{x}{2})$ in any SPE. Agents with any $w \leq \frac{y}{3}$ get net gains under both C_1^x and C_2^y. As for $\frac{y}{3} < w \leq \frac{x}{2}$, they quit at time 0 if principal chooses C_2^y, but quit at time 2 if principal chooses C_1^x. We first characterize the condition under which C_1^x is chosen in the subgame induced by $I_1^* = I(\frac{x}{2})$. Define $\varphi_{|1}(x)$ as the difference between $U_{P|1}(C_1^x \mid I(\frac{x}{2}))$ and $U_{P|1}(C_2^y \mid I(\frac{x}{2}))$:

$$\varphi_{|1}(x) = \int_0^{\frac{x}{2}} (x - w) dF(w) - \int_0^{\frac{x}{2}} (y - 2w) dF(w).$$

Since we have

$$\begin{cases} \varphi_{|1}(\frac{2y}{3}) = \int_0^{\frac{y}{3}} (w - \frac{y}{3}) dF(w) < 0, \quad \varphi_{|1}(y) = \int_0^{\frac{y}{2}} w dF(w) > 0, \\ \frac{d}{dx} \varphi_{|1}(x) = \frac{1}{2}(\frac{3x}{2} - y) f(\frac{x}{2}) + F(\frac{x}{2}) > 0, \end{cases}$$

by continuity of $\varphi_{|1}(x)$, there exists a unique $\hat{x} \in (\frac{2y}{3}, y)$ such that $\varphi_{|1}(\hat{x}) = 0$. $\hat{x}$ is the threshold which determines principal's equilibrium choice at time 1. If and only if $x > \hat{x}$, the joint participation of agents from subset $I(\frac{x}{2})$ will make high-outside-option agents' interest outweigh low-outside-option agents' interest, resulting in the SPE outcome C_1^x. Moreover, it is the unique SPE-NCF outcome. To see this, suppose C_2^y is also a SPE-NCF outcome, where the subset of active agents before the principal make a decision is $\tilde{I} \in I$. First we must have $\tilde{I} \subseteq I(\frac{y}{3})$, because any agent with $w > \frac{y}{3}$ gets negative payoff by consuming C_2^y. Then agents with $w \in I(\frac{x}{2}) \setminus \tilde{I}$ may jointly deviate from quitting at time 0 to staying in the group, which

[11] This property is derived from the linearity of agents' utility function. Allowing for more general utility functions seems not to add much more insights, but makes the model intractable.

makes the principal choose C_1^x and benefits all these agents, contradicting the definition of SPE-NCF. Thus, $C^{\text{equ}} = C_1^x$ if and only if $x > \hat{x}$.

Next, we characterize the condition that makes C_1^x suboptimal from principal's time-0 perspective. Define $\varphi_{|0}(x)$ as the difference between $U_{P|0}(C_1^x)$ and $U_{P|0}(C_2^y)$:

$$\varphi_{|0}(x) = \int_0^{\frac{x}{2}} (x-2w)dF(w) - \int_0^{\frac{y}{3}} (y-3w)dF(w).$$

Since we have

$$\begin{cases} \varphi_{|0}(\hat{x}) = \varphi_{|1}(\hat{x}) + \int_{\frac{y}{3}}^{\frac{\hat{x}}{2}} (y-3w)dF(w) < 0, \\ \varphi_{|0}(y) = \int_0^{\frac{y}{3}} wdF(w) + \int_{\frac{y}{3}}^{\frac{y}{2}} (y-2w)dF(w) > 0, \\ \frac{d}{dx}\varphi_{|0}(x) = F(\frac{x}{2}) > 0, \end{cases}$$

by continuity of $\varphi_{|0}(x)$, there exists a unique $\tilde{x} \in (\hat{x}, y)$ such that $\varphi_{|0}(\tilde{x}) = 0$. $\tilde{x}$ is the threshold which determines principal's optimal choice if she has commitment power at time 0. Thus, $C^{\text{eff}} = C_1^x$ if and only if $x > \tilde{x}$.

We assume the tie-breaking rule that, when the principal is indifferent between the two options, she implements the default option (i.e. C_2^y). Define $\hat{\delta} = \frac{\hat{x}}{y}$ and $\tilde{\delta} = \frac{\tilde{x}}{y}$. Since $\hat{x} < \tilde{x}$, we have the following results: (i) $C^{\text{eff}} = C^{\text{equ}} = C_2^y$ when $0 \leq \delta \leq \hat{\delta}$, (ii) $C^{\text{eff}} = C_2^y$ and $C^{\text{equ}} = C_1^x$ when $\hat{\delta} < \delta \leq \tilde{\delta}$, (iii) and $C^{\text{eff}} = C^{\text{equ}} = C_1^x$ when $\tilde{\delta} < \delta \leq 1$. Figure 3.1 illustrates C^{eff}, C^{equ} and the principal's time-0 payoff at different values of δ.

We can see that there is no collision between the efficient choice and the equilibrium choice for extreme values of δ. Particularly, when δ is too low, the liquidation cost becomes so large that even the presence of high-outside-option agents in the group could not make the principal choose $C_1^{\delta y}$. While if δ is high enough, $C_1^{\delta y}$ not only can be induced by the participation of high-outside-option agents, but also becomes the efficient choice because the loss in consumption due to reallocation is compensated by the gain from saving agents' opportunity costs. However, for intermediate δ, it is ex-ante efficient to choose C_2^y and exclude the agents with high outside options; however, those agents will crowd into the group and force the principal to liquidate the future consumption in the final decision. In short, dynamic preference reversal only occurs for intermediate level of $\delta \in (\hat{\delta}, \tilde{\delta}]$. We summarize this result in the following theorem.

Theorem 10. Dynamic preference reversals arise when the effectiveness of reallocating resources from later periods to earlier periods is at the intermediate level.

Because individual preferences are time consistent, the total welfare is well-defined by the principal's time-0 payoff. We clearly see from Figure 3.1 that dynamic preference reversals

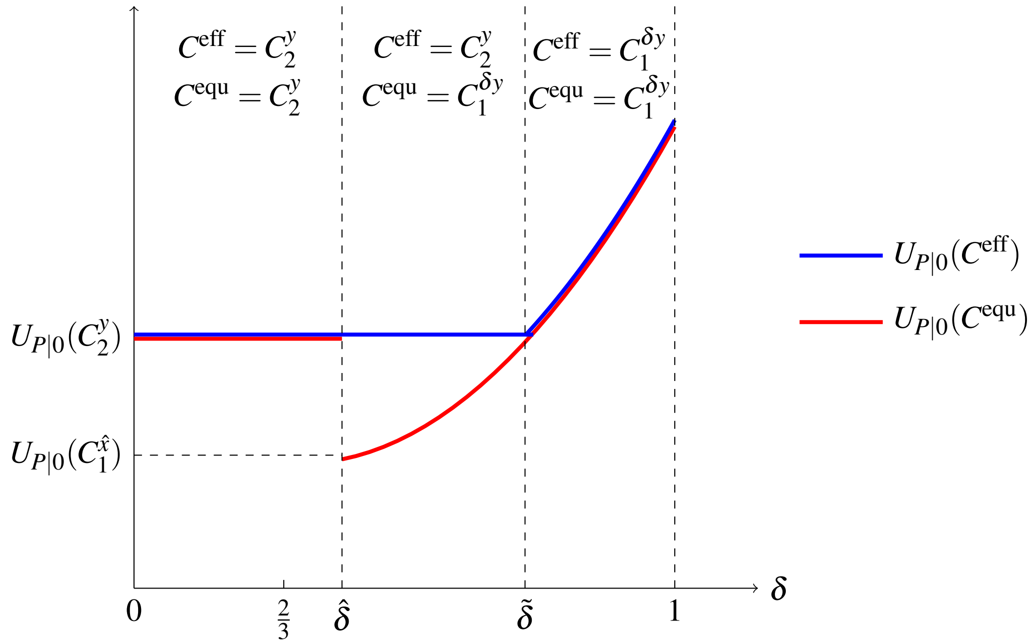

Figure 3.1: The principal's time-0 payoff at C^{eff} and C^{equ}, respectively.

cause welfare loss, which is measured by difference between $U_{P|0}(C_2^y)$ and $U_{P|0}(C_2^{\delta y})$, i.e. the distance between the blue line and the red line at each δ. Moreover, the jump point of the red line means that welfare loss reaches the severest level in the neighborhood of $\hat{\delta}$, and gradually decreases as δ increases. In Section 3.5 we consider a situation where δ is no longer common knowledge and the principal has control over how agents learn about δ through public messages. We exploit the discontinuity of $U_{P|0}(C^{\text{equ}})$ to construct the optimal Bayesian-persuasion information structure.

3.4.1 Resource size effect

We have characterized the range of effectiveness of reallocation where the efficient outcome is not achieved due to dynamic preference reversals. A natural question is whether this range depends on the size of resources (y). If not, then the amount of resource has limited influence on the principal's decision; otherwise one has to be more vigilant against the size effect of resources, because increasing total resources does not necessarily improve welfare, but could give rise to detrimental inconsistency problems under certain circumstances.

We write the two thresholds $\tilde{\delta}$ and $\hat{\delta}$ as functions of y by means of implicit function.

Specifically, $\tilde{\delta}(y) = \frac{\tilde{x}(y)}{y}$, where $\tilde{x}(y)$ is given by

$$\int_0^{\frac{y}{3}} (y - 3w)dF(w) = \int_0^{\frac{\tilde{x}(y)}{2}} (\tilde{x}(y) - 2w)dF(w);$$

and $\hat{\delta}(y) = \frac{\hat{x}(y)}{y}$, where $\hat{x}(y)$ is given by

$$\int_0^{\frac{\hat{x}(y)}{2}} (\hat{x}(y) - w)dF(w) = \int_0^{\frac{\hat{x}(y)}{2}} (y - 2w)dF(w).$$

Let $e_F(w) := \frac{f(w)w}{F(w)}$ be the elasticity of $F(w)$. By definition $F(w)$ is the distribution of agents' outside options, which can be viewed as the proportion of accepting agents when consumption w is proposed by the principal. Then $e_F(w)$ measures the percentage change in agents' acceptance proportion induced by unit of percentage change in the group's provision of public goods. It turns out that $e_F(w)$ plays a decisive role in determining the properties of $\tilde{\delta}(y)$ and $\hat{\delta}(y)$.

Proposition 6. If $e_F(w)$ is monotonically increasing (or decreasing) in w, then both $\tilde{\delta}(y)$ and $\hat{\delta}(y)$ are monotonically decreasing (or increasing) in y.

The proposition states that, if the elasticity of $F(w)$ is monotone, then the intermediate range of δ where dynamic preference reversal arises will monotonically shift as the size of resources varies. Intuitively, suppose that $e_F(w)$ is increasing, then if we raise y by one percent, the relative measure of new comers (whose outside options are relatively high) compared with the original active agents (whose outside options are relatively low) is increasing with the current level of y. This has two effects on the principal's decision. First, the utilitarian principal caters more to the high-outside-option agents, which makes $C_1^{\delta y}$ more likely to be the efficient outcome, and thus $\tilde{\delta}(y)$ decreases. Second, the proportion of high-outside-option agents in all active agents goes up, then it becomes easier for them to make the principal choose the suboptimal choice. So $\hat{\delta}(y)$ also decreases.

An immediate corollary is that, if $e_F(w)$ is constant, then both $\tilde{\delta}(y)$ and $\hat{\delta}(y)$ are invariant with respect to y. It is equivalent to require that the distribution function $F(w)$ take the form of power functions, including the uniform distribution. In this case the range of dynamic preference reversal is independent of y, thus the amount of resources is positively correlated with the total welfare. Otherwise, the changes in the size of y may alter the principal's decision and cause discontinuous variation of total welfare. For instance, C_2^y is both the efficient and equilibrium outcome at some $\delta = \hat{\delta}(y_0) - \varepsilon$ where $\varepsilon > 0$ is sufficiently small. Suppose $F(w)$ has an increasing elasticity, then we have that $\hat{\delta}(y)$ is decreasing in y. Now there is some "good" news that principal has access to a bit more resources, so that the resource increases from y_0 to y_1. However, more resources can induce a lower threshold $\hat{\delta}(y_1) < \delta < \tilde{\delta}(y_1)$, resulting in dynamic preference reversal of the principal. Notice that the welfare loss is significant near the threshold, exceeding by far the potential gain brought by the increase in resources, thus the "good" news could end up with an adverse outcome.

3.4.2 Group composition effect

Another interesting question is how dynamic inconsistency is related to the group composition of agents' outside options. It is striking that the elasticity function of the distribution of agents' outside options again proves to be a prime indicator.

Proposition 7. For any two distributions $F_1(w)$ and $F_2(w)$, if $e_{F_1}(w) \leq e_{F_2}(w)$ for all w, then given any y we have $\tilde{\delta}_{F_1}(y) \geq \tilde{\delta}_{F_2}(y)$ and $\hat{\delta}_{F_1}(y) \geq \hat{\delta}_{F_2}(y)$.

The above proposition establishes the link between the range of liquidation rate (δ) which leads to dynamic preference reversals, and the distribution of agents' opportunity cost of participation. Take any two groups, if the distribution of one group's outside options has an elasticity uniformly higher than the other group, then the former group is more likely to suffer from dynamic inconsistency under lower liquidation rates compared with the latter one.

To better understand this result, we restrict to distributions with constant elasticity, which take the form of power functions as we discussed before. Thus, the relative size of resource plays no role in determining the two thresholds, and we can focus on the group composition effect. We further normalize the upper bound of outside options to be 1, i.e. $W = 1$. Then the possible set of distribution functions is given by $\mathbb{C} := \{F : [0,1] \mapsto [0,1] \mid F(w) = w^r, r > 0\}$, and the elasticity e_F is given by r. Pick any pair $r_1 < r_2$, and then we have $F_1(w) = w^{r_1} \geq w^{r_2} = F_2(w)$ for all $w \in [0,1]$, which means F_2 dominates F_1 in the sense of first-order stochastic dominance. Immediately, we have the following corollary.

Corollary 1. Given any pair of distributions $F_1, F_2 \in \mathbb{C}$, dynamic preference reversal for Group 1 occurs at (weakly) higher liquidation rates than Group 2 if and only if F_2 first-order stochastically dominates F_1.

Roughly speaking, the group with outside options concentrated in the lower level will more likely exhibit dynamic preference reversal under higher level of liquidation rates, and vise versa. This result throws light on collective decisions in two aspects. First, it point out that policy makers should pay close attention to the organizations they are faced with. When manipulating the liquidation rate through various policy instruments such as debt plans, tax rates and subsidies, policy makers should be aware of the fact that the policy could impose quite different effects on different entities. Second, it provide some guideline for compositional optimization of the group. Heterogeneity of group members' outside option is inevitable in reality, however, there always exists some room for the group leader to adjust the composition. Given certain reallocation device, the group leader can make appropriate adjustments to the distribution of agents' outside options through, for instance, changing the initial members and segmentation, so that the group's decision is able to circumvent the inconsistency issues.

3.5 Welfare Improvement

3.5.1 Optimal Bayesian persuasion mechanism

In the benchmark case we assume that δ is common knowledge and realized at $t = 0$, and show that social welfare is distorted under intermediate level of liquidation rates due to dynamic preference reversals. In reality, the group leader usually has an information advantage over the group members in assessing the effectiveness of reallocating resources. Thus it is meaningful to ask whether the benevolent group leader can manipulate the information disclosure rule to alleviate welfare loss. In this section we apply the Bayesian persuasion approach to characterize the optimal disclosure rule.

Consider the previous three-period model, where agents' outside options and the principal's time-2 resource y are common knowledge, except that the liquidation rate δ is the state of world subject to a common prior $G[0,1]$ and is privately observed by principal at time 0. Imagine that the principal is aimed at maximizing the expected total welfare by choosing an message structure M_0 before δ is realized at time 0. As is always assumed in Bayesian persuasion model, the principal can commit to that revealing rule in the whole time horizon.[12] After observing δ, according to the disclosure rule, principal sends a public message $m_0 \in M_0$ which induces a time-0 posterior belief $\mu_0 \in \Delta([0,1])$ shared among the agents. Then agents decide whether to stay in the group, resulting in the participation subset I_1^*. At time 1, principal observes I_1^* and chooses the fraction α of time-2 resource to be liquidated. Then the common consumption stream $(c_1, c_2) = (\alpha\delta y, (1-\alpha)y)$ is implemented and each agent who

[12]For instance, before observing the state of world, the principal can employ a third-party to certify the information due to lack of the ability to generate hard evidence herself. The principal can also design a machine to automatically reveal the realized state according to some preset program.

participated at time 0 decides his optimal quitting time.[13]

First, since we consider pure-strategy SPE-NCF, I_1^* satisfies $\{i \in I \mid w(i) \leq w^*\}$, where $w^* \leq \frac{y}{2}$ is the upper bound of participating outside options. By Lemma 10, at time 1 the group leader actually chooses between $C_1^{\delta y}$ and C_2^y. Given any δ, define $\bar{w}(\delta)$ as the minimum w^* that make the group leader choose $C_1^{\delta y}$ at time 1 if δ is realized. Suppose $\bar{w}(\delta) > \delta y$, since $\bar{w}(\delta)$ equalizes the group leader's time-1 objective function between choosing $C_1^{\delta y}$ and C_2^y, that is,

$$\int_0^{\delta y} (\delta y - w) dF(w) + \int_{\delta y}^{\bar{w}(\delta)} 0 dF(w) = \int_0^{\bar{w}(\delta)} (y - 2w) dF(w)$$
$$\iff \quad \int_0^{\delta y} (\delta y + w - y) dF(w) = \int_{\delta y}^{\bar{w}(\delta)} (y - 2w) dF(w) > 0,$$

which means $\delta > \frac{1}{2}$, contradicting the fact that the highest possible value of w^* is $\frac{y}{2}$ when agents know $\delta = 1$ at time 0. Thus, we only need to consider $\bar{w}(\delta) \leq \delta y$. Let

$$\varphi_{|1}(\delta, w^*) = \int_0^{w^*} (\delta y - w) dF(w) - \int_0^{w^*} (y - 2w) dF(w) = \int_0^{w^*} (\delta y + w - y) dF(w),$$

then we have $\varphi_{|1}(\delta, \bar{w}(\delta)) = 0$, $\delta y + \bar{w}(\delta) - y > 0$, $\frac{d\bar{w}(\delta)}{d\delta} = -\frac{F(\bar{w}(\delta))y}{f(\bar{w}(\delta))(\delta y + \bar{w}(\delta) - y)} < 0$, and $\exists \frac{1}{2} < \underline{\delta} < \hat{\delta}$ such that $\bar{w}(\underline{\delta}) = \frac{y}{2}$.[14] Moreover, we have $\bar{w}(\delta) > \frac{\hat{\delta} y}{2}$ if $\delta < \hat{\delta}$, $\bar{w}(\delta) < \frac{\hat{\delta} y}{2}$ if $\delta > \hat{\delta}$, and $\bar{w}(\hat{\delta}) = \frac{\hat{\delta} y}{2}$. It is also easy to check that $\bar{w}(1) = 0$ and $\bar{w}(\delta)$ is continuous.

After receiving the principal's message m_0, agents form a posterior belief μ_0 over all possible δ. Since $\bar{w}(\delta)$ is a decreasing function, and for any δ the group leader will choose $C_1^{\delta y}$ at time 1 if and only if $w^* > \bar{w}(\delta)$, then agents' participation decision sets a threshold δ such that group leader will choose the default option if $\delta' \in [0, \delta]$ and choose the liquidation option if $\delta' \in (\delta, 1]$. Define $w_{\mu_0}(\delta)$ as the maximum participating outside option if the threshold is δ and agents' posterior belief is μ_0, i.e.,

$$\int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0}(\delta)\big) d\mu_0(\delta') + \int_{\delta' \in (\delta,1]} \big(\delta' y - 2w_{\mu_0}(\delta)\big) d\mu_0(\delta') = 0$$
$$\implies \quad w_{\mu_0}(\delta) = y \frac{\mu_0(\delta) + \int_{\delta' \in (\delta,1]} \delta' d\mu_0(\delta')}{2 + \mu_0(\delta)},$$

[13] At this stage δ becomes common knowledge because agents can infer the value of δ from (c_1, c_2).

[14] Since we have $\frac{d\varphi_{|1}(\delta, w^*)}{dw^*} = (\delta y + w^* - y) f(w^*)$, then starting from $\varphi_{|1}(\delta, 0) = 0$, $\varphi_{|1}(\delta, w^*)$ first decreases over $(0, y - \delta y)$, then increases over $(y - \delta y, \frac{y}{2})$. Notice that $\varphi_{|1}(\hat{\delta}, \frac{\hat{\delta} y}{2}) = 0$ and $\frac{d\varphi_{|1}(\delta, w^*)}{dw^*} = yF(w^*) > 0$, then there exists $\frac{1}{2} < \underline{\delta} < \hat{\delta}$ such that as δ decreases, $\bar{w}(\delta)$ gradually increase until it reaches the upper bound $\frac{y}{2}$. For any $\delta < \underline{\delta}$ the leader will always chooses C_2^y.

where the integral is defined in the sense of Lebesgue because we allow the posterior belief μ to be arbitrary cumulative distribution. Immediately, we have that $w_{\mu_0}(\delta) \leq \frac{y}{2}$ for all δ, $w_{\mu_0}(1) = \frac{y}{3}$ and $w_{\mu_0}(\delta)$ is right-continuous with left limits, since μ_0 is right-continuous with left limits. The following proposition establishes the existence and uniqueness of pure-strategy SPE-NCF given any μ_0.

Proposition 8. After principal sending any message $m_0 \in M_0$, the following game has a unique pure-strategy SPE-NCF, satisfying

(1) $I_1^* = \{i \in I \mid w(i) \leq w^*(\mu_0) := w_{\mu_0}(\delta^*)\}$, where $\delta^* = \min\{\delta \mid w_{\mu_0}(\delta) = \bar{w}(\delta)\}$;

(2) The principal chooses $C_1^{\delta y}$ if $\delta > \delta^*$, and C_2^y if $\delta \leq \delta^*$.

We define the *upper revealing truncation* of μ_0 at δ^T as follows: replace m_0 by a collection of new messages $\{m_0'\} \cup \{m_{0|\delta}\}_{\delta > \delta^T}$ following the distribution $\pi(m_0') = \int_0^{\delta^T} d\mu_0(\delta)$ and $\pi(\{m_{0|\delta}\}_{\delta \in \Lambda}) = \mu_0(\{\delta\}_{\delta \in \Lambda})$ for all $\Lambda \in 2^{(\delta^T,1]}$. The posterior belief induced by m_0' is $\mu_0'(\delta) := \mu_0[m_0'](\delta) = \frac{\mu_0(\delta)}{\mu_0(\delta^T)}$ for all $\delta \leq \delta^T$, and the posterior belief induced by $m_{0|\delta}$ for any $\delta > \delta^T$ is $\mathbf{Pr}(\delta \mid m_{0|\delta}) = 1$. Let $w_{\mu_0}(\delta, \delta^T) := w_{\mu_0'}(\delta)$, then we have that, if $\delta < \delta^T$,

$$
\begin{aligned}
& \int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0}(\delta,\delta^T)\big) d\mu_0'(\delta') + \int_{\delta' \in (\delta,1]} \big(\delta' y - 2w_{\mu_0}(\delta,\delta^T)\big) d\mu_0'(\delta') = 0 \\
\Leftrightarrow \quad & \int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0}(\delta,\delta^T)\big) d\mu_0(\delta') + \int_{\delta' \in (\delta,\delta^T]} \big(\delta' y - 2w_{\mu_0}(\delta,\delta^T)\big) d\mu_0(\delta') = 0 \\
\Leftrightarrow \quad & w_{\mu_0}(\delta,\delta^T) = y \frac{\mu_0(\delta) + \int_{\delta' \in (\delta,\delta^T]} \delta' d\mu_0(\delta')}{2\mu_0(\delta^T) + \mu_0(\delta)},
\end{aligned}
$$

and if $\delta \geq \delta^T$, $w_{\mu_0}(\delta, \delta^T) = \frac{y}{3}$. Now we provide the optimal information structure in the following theorem.

Theorem 11. The optimal information disclosure rule is an upper revealing truncation of the common prior $G[0,1]$ at $\delta^T = \delta^{SB}$, satisfying

$$
\begin{aligned}
\delta^{SB} &:= \max\big\{\delta^T \leq \tilde{\delta} \mid \max_{\check{\delta} \leq \delta^T} \chi_G(\check{\delta}, \delta^T) \leq 0\big\}, \\
\chi_G(\check{\delta}, \delta^T) &:= \int_{\delta' \in [0,\check{\delta}]} \big(y - 3\bar{w}(\check{\delta})\big) dG(\delta') + \int_{\delta' \in (\check{\delta},\delta^T]} \big(\delta' y - 2\bar{w}(\check{\delta})\big) dG(\delta').
\end{aligned}
$$

The optimal information disclosure rule recaptures the so-called *censorship policy* as in Kolotilin, Li, Mylovanov, and Zapechelnyuk (2015) and Yamashita (2016). The driving force, which we mentioned in the introduction and will be discussed later, is a synthetical consideration of providing substitute for commitment power and adjusting to the state of world.

The technique we developed to characterize the optimal disclosure rule contributes to solving more general Bayesian persuasion problem in two aspects. First, we work on continuous state space rather than the finite state space which is often assumed in standard Bayesian persuasion literature. Due to the infinite dimensional environment, the mature graphical approach, which directly explores the concavity or convexity of Sender's payoff as a function of Receiver's beliefs, is no longer applicable. Second, the standard recommendation approach (e.g., Kamenica and Gentzkow, 2011), which sends straightforward signals to receivers to guide their actions, fails to simplify the solving process. Instead of having only one receiver or multiple receivers who are essentially separate from each other, in our model there is a continuum of receivers whose final payoffs are interdependent. In other words, for any posterior belief induced by some message, we're looking for a fixed point of a subgame rather than just a maximizer of a single receiver's payoff. In our approach, we proceed in a direct way: we first characterize the equilibrium of the subgame induced by arbitrary posterior belief, then explore several properties possessed by the optimal rule, and finally pin down principal's optimal sending strategy.

3.5.2 Welfare consequences

Dynamic inconsistency harms social welfare for intermediate liquidation $\delta \in (\hat{\delta}, \tilde{\delta})$, where the welfare loss becomes severer as δ shifts from $\tilde{\delta}$ to $\hat{\delta}$. This is because near the right limit of $\hat{\delta}$, the change of principal's time-1 decision causes a sudden drop in total utility, while the potential gain from increasing the liquidated resource is gradual. To improve the social welfare, the optimal disclosure rule should maximize the implementation of default option for $\delta \in (\hat{\delta}, \tilde{\delta})$, and meanwhile keep the other δ unaffected. The reason why we pool $\delta \in (\hat{\delta}, \delta^{SB}]$ with $\delta \leq \hat{\delta}$ is as follows: (1) on receiving the pooling message, agents form a pessimistic expectation about the actual liquidation rate, which will deter the participation of the high-outside-option agents, resulting in implementing the default option; (2) welfare gains is larger from amending the outcome under smaller $\delta \in (\hat{\delta}, \tilde{\delta})$ than the larger ones, and pooling smaller $\delta \in (\hat{\delta}, \tilde{\delta})$ will provide less incentive for the high-outside-option agents to crowd in the group. From Theorem 11 we immediately get the following corollary which assesses the performance of the Bayesian persuasion mechanism.

Corollary 2. The optimal information disclosure rule fully restores efficiency if $\delta^{SB} = \tilde{\delta}$, and partially alleviates the welfare loss due to dynamic inconsistency if $\delta^{SB} < \tilde{\delta}$.

We can see that Bayesian persuasion approach works pretty well in the sense that, without further instruments such as payment rules, it always manages to restore some efficiency only by controlling the revealed information, and under certain condition it can thoroughly eliminate the welfare distortion. The implication of this result is that, full transparency of information within a group is suboptimal for collective decision processes. When suffering from dynamic inconsistency issues, the group leader who maximizes the social welfare should conceal the unfavorable news and completely reveal only the auspicious news.

3.6 Discussions and Applications

3.6.1 Solution concept

The solution concept we applied guarantees the determinacy of equilibrium outcome; otherwise we would have to adopt alternative criteria such as the maxmin rule, where the collective utility is evaluated according to the worst-case scenario, which would make the problem intractable. In this subsection we provide some foundations for this refinement of equilibriums.

From the definition of SPE-NCF, one straightforward way is to allow agents to play a *cooperative game without transfers* at stage (i). Take Proposition 8 for example, given the posterior belief induced by the public message, every subgame perfect equilibriums can be labeled by a threshold δ_n above which the principal chooses $C_1^{\delta y}$. List all SPE in an ascending order, i.e. $\delta_1 < \delta_2 < \ldots < \delta_n < \ldots$, and the subset of active agents $I^*_{1(n)} = \{i \in I \mid w(i) \leq w_{\mu_0}(\delta_n)\}$ satisfies $I^*_{1(1)} \supseteq I^*_{1(2)} \supseteq \ldots \supseteq I^*_{1(n)} \supseteq \ldots$. We can easily check that agents with $w \in I^*_{1(1)}$ participate at time 0 is the unique *core* of the cooperative game, which exactly corresponds to SPE-NCF. To prove this, pick any $\tilde{I} \neq I^*_{1(1)}$, first we must have $\tilde{I} \subseteq I^*_{1(1)}$ because any $w \in \tilde{I} \setminus I^*_{1(1)}$ will get negative utility and can be better off by quitting at time 0. Then agents with $w \in \{i \in I \mid \frac{y}{3} \leq w(i) \leq w_{\mu_0}(\delta_1)\} \setminus \tilde{I}$ form a coalition and participate at time 0, which can alter the principal's time-1 action and strictly make these agents better off. Thus, the only participating subset that will not be blocked by any coalition of agents is $I^*_{1(1)}$.

Another way to rationalize this equilibrium selection is to extend the simultaneous-move game at stage (i) to a *sequential game*, where (1) each agent has one opportunity to decide whether to quit the group before the principal's decision time, (2) the decision order of agents is common knowledge and deterministic, but can be arbitrary, and (3) agents can observe the entire history of past decisions. In the three-period model in Section 3.4, the SPE-NCF is actually the unique subgame perfect Nash equilibrium of the sequential game (Lemma 19 in Appendix D.2.1).

As for the Bayesian persuasion model in Section 3.5, the SPE-NCF can be uniquely approximated by a sequence of subgame perfect Nash equilibriums of the corresponding sequential games (Lemma 20 in Appendix D.2.1).[15] The key point is to approximate the original game with a continuum of agents by a sequence of sequential games with finitely many agents. Immediately, in each finite sequential game, there exists a unique SPE; moreover, that SPE is free from coordination failure among agents.[16] Apply the upper hemi-continuity of Nash correspondence, which has been extensively studied in Green (1984) and Fudenberg and Levine (1988), we conclude that the sequence of subgame perfect equilibriums has a limit, which is also a SPE in the original game. Because the payoff structures of the finite game and the original game are almost the same when they get sufficiently close to each other, the SPE approximated by the finite games is indeed SPE-NCF; otherwise in those finite games we can find some agent who would deviate since the unique SPE is the one without coordination failure.

3.6.2 Provision of public facilities

Our model is highly adaptable to the study of public facility provisions, and has strong explanatory power for interpreting how dynamic inconsistency derives from the interactions between the social planner and households (or firms). A noticeable feature of public facilities is that the construction usually proceeds in multiple stages and depreciation should be taken into consideration. Assume that the public facility can generate value of 1 at the final stage which takes T years to build. If the social planner terminates the construction at stage $t_0 \leq T$, then its value at $t \geq t_0$ after depreciation is given by $\delta(t_0,t)$, satisfying (i) $\delta(t_0,t)$ is increasing on t_0 and decreasing on t, and (ii) $\delta(t_0,t) \geq \delta(t_0-k,t-k)$ for all $t \geq t_0 \geq k > 0$.[17] Obviously, our benchmark model is a special case where $T=2$ and $\delta(t_0,t) \equiv 0$ for all $t > t_0$. Agents are

[15]We cannot directly apply Lemma 19 to the Bayesian persuasion model; instead, we get a weaker version of it. This is because the principal's action is to choose a subset of δ to implement $C_1^{\delta y}$, which has infinitely many cases; while when δ is common knowledge, the principal's choice is binary.

[16]With finite agents, the distribution of outside options doesn't have full-support density, then $\bar{w}(\delta)$ is decreasing and right-continuous with left limits. Define a correspondence $\bar{w}^0(\delta) = \{w \mid \lim_{\delta' \downarrow \delta} \bar{w}(\delta') \leq w \leq \lim_{\delta' \uparrow \delta} \bar{w}(\delta')\}$, which is obviously continuous. Through exactly the same argument we can find the minimum δ^* such that $w_{\mu_0}(\delta^*) = \bar{w}^0(\delta^*)$. It follows that δ^* satisfies: (1) $w_{\mu_0}(\delta^*) \geq \bar{w}(\delta^*)$, and (2) $w_{\mu_0}(\delta) < \bar{w}(\delta)$ for all $\delta < \delta^*$. Thus, agents with $w \leq w_{\mu_0}(\delta^*)$ participate at time 0 and principal implements liquidation option if and only if $\delta > \delta^*$ constitute a subgame perfect equilibrium. There is no coordination failure because the participation set is maximized. On the other hand, from backward induction we know that the finite extensive-form game has a unique SPE, as long as we assume that agents prefer participation when indifference occurs. Thus, uniqueness is also proved.

[17]It is natural to require that the value of more comprehensive facilities should exceed the value of less comprehensive ones after depreciating for equal length of periods.

households who decide whether to settle down in this community or to move elsewhere.

Unlike the binary choices in the benchmark case, the planner has to determine the duration of construction to maximize the social welfare, which is essentially an optimal stopping game; however, the driving force of dynamic inconsistency is preserved. We assume no information asymmetry for simplicity. Anticipating the planner stopping at t_0, household with w who lives in the community at time 0 will move out at $\bar{t}(w,t_0)$ such that $w = \delta(t_0, \bar{t}(w,t_0))$; and the highest outside option, denoted by $\bar{w}(t_0)$, that participates at time 0 satisfies

$$U_{A|0}(\bar{w}(t_0),t_0) = \int_{t=t_0}^{\bar{t}(\bar{w}(t_0),t_0)} \delta(t_0,t)dt - \bar{t}(\bar{w}(t_0),t_0)\cdot \bar{w}(t_0) = 0.$$

The efficient duration of construction, denoted by t_0^{FB}, solves

$$\max_{t_0\in[0,T]} U_{P|0}(t_0) = \int_0^{\bar{w}(t_0)} \Big(\int_{t=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t)dt - \bar{t}(w,t_0)\cdot w \Big) dF(w).$$

We can prove that the pure-strategy SPE-NCF is the one with the largest subset of active households at time 0; and the associated construction duration is t_0^{SB}. Then we have the following proposition.

Proposition 9. Dynamic inconsistency occurs in the process of public facility provision if and only if $\bar{w}(t_0^{SB}) > \bar{w}(t_0^{FB})$.

Basically, the absence of planner's commitment power leaves room for the high-outside-option households to jointly stay in the community, so as to induce social planner to choose the duration which is favorable to them but is ex ante inefficient.

3.6.3 Team work

Our framework is also suitable for modeling dynamic inconsistency in team work. For example, the director of a laboratory (i.e. the principal) wants to conduct a research project and decide the assignment among lab members. The research achievement will win honor for the group, which, in essence, is a public good. The ultimate goal of the project requires long periods of research, while the interim results are also promising, which enable the research team to publish their achievements on an early date. In this context, agents represent units of research time of the lab members. The heterogeneity of outside options comes from the facts that lab members may have various expertise, and potentially they can work on other suitable research projects.

We consider the three-period model in Section 3.4, where δ is common knowledge. An interesting feature brought by the team work context is the *network effect*. We assume that the principal's resource at time 2 depends on the subset of active agents who stay in the group until the common consumption arrives, i.e. $y(I_1^*) : 2^{[0,1]} \mapsto \mathbb{R}$. Intuitively, research quality can be affected by the total engaged time, composition of expertise and cooperation efficiency. Since the subset of active agents in any SPE takes the form of $I_1^* = \{i \in I \mid w(i) \leq w^*\}$, we write $y(I_1^*) = y(w^*)$. For simplicity, we assume that $y(\cdot)$ satisfies the Inada conditions: $y(\cdot)$ is continuously differentiable, $y(0) = 0$, $y'(\cdot) > 0$, $y''(\cdot) < 0$, $\lim_{w^* \to 0} y'(\cdot) = +\infty$ and $\lim_{w^* \to +\infty} y'(\cdot) = 0$. Thus, anticipating that C_2^y will be implemented, the highest outside option among active agents, denoted as $\bar{w}(C_2^y)$, is well-defined by $\bar{w}(C_2^y) = \frac{1}{3} y\big(\bar{w}(C_2^y)\big)$. Similarly, we define $\bar{w}(C_1^{\delta y})$ such that $\bar{w}(C_1^{\delta y}) = \frac{\delta}{2} y\big(\bar{w}(C_1^{\delta y})\big)$. Let $\kappa = y\big(\bar{w}(C_1^{\delta y})\big) / y\big(\bar{w}(C_2^y)\big)$.

To focus on the network effect on dynamic inconsistency, we assume that $F(w)$ has constant elasticity so that we can get rid of the resource size effect. In the benchmark case, the two thresholds $\hat{\delta}$ and $\tilde{\delta}$ are independent of y. Let $\hat{\delta}^t$ and $\tilde{\delta}^t$ be the corresponding thresholds in the team work model, then we provide the following result.

Proposition 10. $\hat{\delta}^t \equiv \hat{\delta}$, and $\tilde{\delta}^t = \tilde{\delta} / \kappa$. The collective decision is present-biased if $\tilde{\delta} > \kappa \hat{\delta}$, future-biased if $\tilde{\delta} < \kappa \hat{\delta}$, and dynamic consistent if $\tilde{\delta} = \kappa \hat{\delta}$.

As is proved in Theorem 9, $\bar{w}(C_1^{\delta y}) \geq \bar{w}(C_2^y)$ is necessary to induce dynamic inconsistency. Due to network effect, $C_1^{\delta y}$ attracts more team members who in return boost the value of $C_1^{\delta y}$, and thus outperforms C_2^y more frequently than in the benchmark case. Roughly speaking, higher κ means stronger network effect. As a result, the intermediate range of liquidation rates (δ) inducing dynamic preference reversals will shrink as κ increases, which means that moderate network effect mitigates dynamic inconsistency. However, too strong network effect can also be noxious. Particularly, $C_1^{\delta y}$ may become ex ante optimal even for some relatively small δ which is not large enough to support the implementation of $C_1^{\delta y}$ in SPE-NCF, giving rise to future bias.

3.7 Conclusions

This paper establishes a framework where collective decisions made by a group of consistent agents with exponential discounting time preferences could exhibit dynamic inconsistency. We point out that the driving force is the heterogeneity of opportunity cost that agents have to forgo in order to participate in the group. We further characterize the optimal information disclosure rule, which takes the form of censorship policy, to restore the welfare distortion caused by inconsistent collective decisions. Our framework proves to be quite adaptable to more general environment, and has considerable explanatory power in many applications such as public facility provision and team work.

Application of behavioral economics in mechanism design has received considerable attention in recent literature, where agents' individual preferences are assumed to exhibit irrational properties such as present bias, bounded rationality and overconfidence. Instead of assuming nonstandard preferences, it is interesting to explore the new properties brought by the endogenous dynamic inconsistency established in this paper. The standard setup will be extended to three layers, where the grand principal designs the mechanism according to certain criterion, and the group leader makes collective decisions on the optimal signaling strategy subject to group members' actions.[18] We expect to draw novel insights from this direction of future research.

[18] See Appendix D.2.2 for an example.

Appendix A

Appendix for Chapter 1: Omitted Proofs

A.1 Proof of Lemma 1

It is easy to check that $u_i(x,v,\theta)$ is linear with respect to x, for $i = 0,1,\ldots,N$. Let (Ξ^+,x^+) be the solution to (P_1). Define $\tilde{x}(v,\theta) := \int_m x^+(v,m)d\Xi^+_\theta(m)$ for each pair $(v,\theta) \in V \times \Theta$. For any $v_i \neq v_i'$,

$$
\begin{aligned}
&\int_\theta \int_{v_{-i}} u_i\big(\tilde{x}(v_i,v_{-i},\theta),v,\theta\big)dF_{-i}(v_{-i})dF_0(\theta)\\
&= \int_\theta \int_{v_{-i}} u_i\Big(\int_m x^+(v_i,v_{-i},m)d\Xi^+_\theta(m),v,\theta\Big)dF_{-i}(v_{-i})dF_0(\theta)\\
&\overset{(1)}{=} \int_\theta \int_{v_{-i}} \int_m u_i\big(x^+(v_i,v_{-i},m_i,m_{-i}),v,\theta\big)d\Xi^+_\theta(m)dF_{-i}(v_{-i})dF_0(\theta)\\
&= \int_{m_i} \int_{v_{-i}} \int_{\theta,m_{-i}} u_i\big(x^+(v_i,v_{-i},m_i,m_{-i}),v,\theta\big)d\Psi^+_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i})d\Lambda^+_i(m_i)\\
&\overset{(2)}{\geq} \int_{m_i} \int_{v_{-i}} \int_{\theta,m_{-i}} u_i\big(x^+(v_i',v_{-i},m_i,m_{-i}),v,\theta\big)d\Psi^+_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i})d\Lambda^+_i(m_i)\\
&= \int_\theta \int_{v_{-i}} u_i\big(\tilde{x}(v_i',v_{-i},\theta),v,\theta\big)dF_{-i}(v_{-i})dF_0(\theta),
\end{aligned}
$$

where (1) is due to the linearity of $u_i(x,v,\theta)$ with respect to x, and (2) comes from the fact that (Ξ^+,x^+) satisfies $BIC_{v_i \to v_i'|m_i}$ in (P_1) for any m_i. Thus, $\tilde{x}$ satisfies $BIC_{v_i \to v_i'}$ in (P^*). Similarly, we can check that $\tilde{x}$ satisfies IIR_{v_i} in (P^*).

Let x^* be the solution to (P^*), then we have

$$\begin{aligned}\int_\theta \int_v u_0\big(x^*(v,\theta),v,\theta\big)dF_V(v)dF_0(\theta) &\geq \int_\theta \int_v u_0\big(\tilde{x}(v,\theta),v,\theta\big)dF_V(v)dF_0(\theta)\\ &= \int_\theta \int_v u_0\Big(\int_m x^+(v,m)d\Xi^+_\theta(m),v,\theta\Big)dF_V(v)dF_0(\theta)\\ &= \int_\theta \int_m \int_v u_0\big(x^+(v,m),v,\theta\big)dF_V(v)d\Xi^+_\theta(m)dF_0(\theta).\end{aligned}$$

Thus, (P^*) is a relaxed problem of (P_1), and is also a relaxed problem of (P).

A.2 Proof of Lemma 2

(i) Suppose there exist some $m \in M$ and $t_1 \neq t_2$ such that $\Phi^S(\theta_{t_1},m) > 0$ and $\Phi^S(\theta_{t_2},m) > 0$, then pick any $i \leq N-1$, and we have

$$\begin{cases} m_{i+1} \equiv m_i + t_1 \pmod T \\ m_{i+1} \equiv m_i + t_2 \pmod T. \end{cases}$$

It follows that $t_1 \equiv t_2 \pmod T$. Since $1 \leq t_1, t_2 \leq T$, then $t_1 = t_2$, contradicting that t_1 and t_2 are distinct numbers.

(ii) Without loss of generality, let $\theta = \theta_t$. For any $j \neq i$, we can define $m_j := \big(m_i + (j-i)t\big) \bmod T$, since $m_j \in \{1,2,\dots,T\}$. Thus, $m_j \equiv m_i + (j-i)t \pmod T$. For any $l_1, l_2 \in I \setminus \{i\}$, we have

$$\begin{cases} m_{l_1} \equiv m_i + (l_1 - i)t \pmod T \\ m_{l_2} \equiv m_i + (l_2 - i)t \pmod T. \end{cases}$$

Thus, $m_{l_1} - m_{l_2} \equiv (l_1 - l_2)t \pmod T$. By definition, $\Phi^S(\theta_t, m_i, m_{-i}) = \frac{\alpha_t}{T} > 0$. Suppose there exist another $m'_{-i} \neq m_{-i}$ such that $\Phi^S(\theta_t, m_i, m'_{-i}) > 0$, then we must have $m'_\tau \neq m_\tau$ for some $\tau \neq i$, satisfying

$$\begin{cases} m_\tau \equiv m_i + (\tau - i)t \pmod T \\ m'_\tau \equiv m_i + (\tau - i)t \pmod T. \end{cases}$$

It follows that $m_\tau \equiv m'_\tau \pmod T$. Since $1 \leq m_\tau, m'_\tau \leq T$, we have $m_\tau = m'_\tau$, which induces a contradiction. Thus, such m_{-i} is unique.

(iii) From property (ii) we immediately have that, agent i's posterior belief about the probability to have θ_t conditional on receiving m_i is given by

$$\Pr(\theta_t \mid m_i) = \frac{\int_{m_{-i}} d\Phi^S(\theta_t, m_i, m_{-i})}{\int_{m_{-i},\tilde{\theta}} d\Phi^S(\tilde{\theta}, m_i, m_{-i})} = \frac{\frac{\alpha_t}{T}}{\sum_{t=1}^T \frac{\alpha_t}{T}} = \frac{\alpha_t}{\sum_{t=1}^T \alpha_t} = \alpha_t, \quad \forall \theta_t \in \Theta.$$

(iv) Pick any $\theta_t \in \Theta$, from property (ii) we have

$$\Pr(\theta_t) = \sum_{m\in M} \Phi^S(\theta_t, m) = \sum_{m_i=1}^{T} \sum_{m_{-i}\in M_{-i}} \Phi^S(\theta_t, m_i, m_{-i}) = \sum_{m_i=1}^{T} \frac{\alpha_t}{T} = \alpha_t.$$

A.3 Proof of Theorem 1

Because Φ^W is defined in the same way as Φ^S, obviously Φ^W preserves all the properties in Lemma 2. To guarantee that we can construct such (Φ^W, x^W), it suffices to show that $x^*\big(v, \theta^+_{-i}(m)\big)$ is well-defined when $\Theta^+(m) = \emptyset$ and $\Theta^+_{-i}(m) \neq \emptyset$ for some i.

Lemma 11. Fixed any $m \in M$, if $\Theta^+(m) = \emptyset$ and there exists i such that $\Theta^+_{-i}(m) \neq \emptyset$, then $|\Theta^+_{-i}(m)| = 1$, and $\Theta^+_{-j}(m) = \emptyset$ for any $j \neq i$.

Proof of Lemma 11. (i) Pick any $l_1, l_2 \neq i$ such that $l_2 > l_1$, and suppose that there exists distinct $\theta_{t_1}, \theta_{t_2} \in \Theta^+_{-i}(m)$, then by definition we have

$$\begin{cases} m_{l_2} \equiv m_{l_1} + (l_2 - l_1)t_1 \pmod K \\ m_{l_2} \equiv m_{l_1} + (l_2 - l_1)t_2 \pmod K. \end{cases}$$

It follows that $0 \equiv (l_2 - l_1)(t_1 - t_2) \pmod K$, which means K exactly divides $(l_2 - l_1)(t_1 - t_2) \neq 0$. Because K is a prime number and $0 < l_2 - l_1 < N \leq K$, we must have K exactly divides $(t_1 - t_2)$. Notice that $1 \leq t_1, t_2 \leq T \leq K$, then we have $t_1 = t_2$, contradicting how we select t_1 and t_2. Thus, we have $|\Theta^+_{-i}(m)| = 1$.
(ii) Suppose there exists $j \neq i$ such that $\Theta^+_{-j}(m) \neq \emptyset$, then from (i) we have $|\Theta^+_{-j}(m)| = 1$, that is, there exists a unique $\theta_{t'} \in \Theta$ such that $m_{l_2} \equiv m_{l_1} + (l_2 - l_1)t' \pmod K$ for any $l_1, l_2 \in I \setminus \{j\}$. Denote $\theta^+_{-i}(m)$ as θ_t. Because $N \geq 4$, we can find $l_1, l_2 \neq i, j$ such that $l_2 > l_1$ satisfying

$$\begin{cases} m_{l_2} \equiv m_{l_1} + (l_2 - l_1)t \pmod K \\ m_{l_2} \equiv m_{l_1} + (l_2 - l_1)t' \pmod K. \end{cases}$$

Then we have $0 \equiv (l_2 - l_1)(t - t') \pmod K$, and through exactly the same argument as in (i) we have $t = t'$. Now, by definition of $\theta^+_{-j}(m)$, we have $m_i \equiv m_{l_1} + (i - l_1)t' \pmod K$ for any $l_1 \neq i, j$. Notice that by definition of $\theta^+_{-i}(m)$, we have $m_{l_1} \equiv m_\tau + (l_1 - \tau)t' \pmod K$ for any $\tau \neq i$, then we have $m_i \equiv m_\tau + (i - \tau)t' \pmod K$. It follows that $\theta^+(m) = \theta_{t'}$, contradicting the assumption that $\Theta^+(m) = \emptyset$. □

Next, we check whether (Φ^W, x^W) satisfies all the constraints in (P). The interim expected utility of agent i with v_i observing m_i by reporting $(\hat{v}_i, \hat{m}_i)$ is

$$\begin{aligned}
U_i(\hat{v}_i, \hat{m}_i; v_i, m_i) &= \int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v, \theta\big) d\Psi^W_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\
&= \int_{v_{-i}} \int_{\theta} u_i\big(x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m^+_{-i}(\theta, m_i)), v, \theta\big) d\Psi^W_{m_i}(\theta) dF_{-i}(v_{-i}) \\
&= \int_{v_{-i}} \int_{\theta} u_i\big(x^*(\hat{v}_i, v_{-i}, \theta), v, \theta\big) dF_0(\theta) dF_{-i}(v_{-i}),
\end{aligned}$$

which is independent of $\hat{m}_i$, regardless of what $\hat{v}_i$ agent i reports to the principal. Thus, no agent has incentive to misreport his signal, and we get $(BIC_{m_i, v_i \to m'_i, v'_i})$ and (IIR_{m_i, v_i}) immediately from $(BIC_{v_i \to v'_i})$ and (IIR_{v_i}). Notice that the principal's expected payoff is

$$\begin{aligned}
&\int_v \int_\theta \int_m u_0\big(x^W(v, m), v, \theta\big) d\Xi^W_\theta(m) dF_0(\theta) dF_V(v) \\
&= \int_v \int_\theta u_0\Big(\int_m x^W(v, m) d\Xi^W_\theta(m), v, \theta\Big) dF_0(\theta) dF_V(v) \\
&= \int_v \int_\theta u_0\Big(x^*(v, \theta), v, \theta\Big) dF_0(\theta) dF_V(v),
\end{aligned}$$

which achieves the upper bound defined by the relaxed problem (P^*), then we conclude that (Φ^W, x^W) is the optimal private disclosure mechanism when $N \geq 4$.

A.4 Proof of Theorem 2

First, we prove the following lemma to show that any unilateral misreport by some agent will be detected by the principal, and thus the punishment can be implemented.

Lemma 12. Under information disclosure policy Φ^W, for any m such that $\Theta^+(m) \neq \emptyset$, fixed any agent i, there exists no other $m'_i \neq m_i$ such that $\Theta^+(m'_i, m_{-i}) \neq \emptyset$.

Proof of Lemma 12. Suppose there exist agent i, signal profile m such that $\Theta^+(m) \neq \emptyset$, and $m'_i \neq m_i$ satisfying $\Theta^+(m'_i, m_{-i}) \neq \emptyset$. Let $\theta^+(m) = \theta_t$ and $\theta^+(m'_i, m_{-i}) = \theta_{t'}$, then from $N = 3$ we can find $j \neq k \neq i$ such that

$$\begin{cases} m_j \equiv m_k + (j-k)t \pmod{K} \\ m_j \equiv m_k + (j-k)t' \pmod{K}, \end{cases}$$

which means $t = t'$. On the other hand, notice that

$$\begin{cases} m_i \equiv m_j + (i-j)t \pmod{K} \\ m'_i \equiv m_j + (i-j)t' \pmod{K}, \end{cases}$$

then we have $m_i - m'_i \equiv (i-j)(t-t') \equiv 0 \pmod{K}$, contradicting $m'_i \neq m_i$. □

Compared with (Φ^W, x^W), we only modify the off-equilibrium-path allocations so as to elicit truthful report from all agents, which means all (IIR_{m_i,v_i}) are satisfied and $(\Phi^W, x^{W|3})$ achieves the same ex ante expected utility as (Φ^W, x^W). Thus the remaining thing is to check $(BIC_{m_i,v_i\to m_i',v_i'})$ for any $(m_i,v_i)\neq(m_i',v_i')$. If $m_i=m_i'$, then $(BIC_{m_i,v_i\to m_i',v_i'})$ holds from the previous result in Section 1.4.1. If $m_i\neq m_i'$, we have

$$\begin{aligned}
&\int_{v_{-i}}\int_{\theta,m_{-i}} u_i\big(x^{W|3}(v_i',v_{-i},m_i',m_{-i}),v,\theta\big)d\Psi^W_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i})\\
&=\int_{v_{-i}}\int_{\theta} u_i\big(\underline{a},v,\theta\big)dF_0(\theta)dF_{-i}(v_{-i})\\
&=\int_{v_{-i}}\int_{\theta}\int_{a\in\mathscr{A}} u_i\big(\underline{a},v,\theta\big)dx^*(v,\theta)(a)dF_0(\theta)dF_{-i}(v_{-i})\\
&\leq\int_{v_{-i}}\int_{\theta}\int_{a\in\mathscr{A}} u_i\big(a,v,\theta\big)dx^*(v,\theta)(a)dF_0(\theta)dF_{-i}(v_{-i})\\
&=\int_{v_{-i}}\int_{\theta} u_i\big(x^*(v_i,v_{-i},\theta),v,\theta\big)dF_0(\theta)dF_{-i}(v_{-i})\\
&=\int_{v_{-i}}\int_{\theta,m_{-i}} u_i\big(x^{W|3}(v_i,v_{-i},m_i,m_{-i}),v,\theta\big)d\Psi^W_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i}).
\end{aligned}$$

As a result, $(\Phi^W, x^{W|3})$ indeed constitutes the solution to (P) for $N=3$.

A.5 Proof of Theorem 3

The disclosure policy, Φ^W, is defined as before. Let $\vec{\Psi}_{m_i}=\big(\Psi_{m_i}(m_{-i})\big)_{m_{-i}\in M_{-i}}$ be the row vector representing agent i's posterior belief about m_{-i} after observing m_i. Let $\Psi:=\big(\vec{\Psi}_{m_i}\big)^{\intercal}_{m_i\in M_i}$ be the $K\times K$ belief matrix induced by the information disclosure policy Φ^W. First, we consider the case where Ψ has full rank. Under the allocation rule $x^{W|2}(v,m)=\tilde{x}^*\big(v,\theta^+(m)\big)$, agent i observing m_i may find it profitable to report $m_i'\neq m_i$, and the amount of violation of $(BIC_{m_i,v_i\to m_i',v_i'})$ also depends on which v_i' agent i sends to the principal. Denote $\beta_i^{m_i\to m_i'}$ as the maximum gain from misreporting $m_i'\neq m_i$ when agent i observes m_i, that is,

$$\begin{aligned}
\beta_i^{m_i\to m_i'}=\max_{(v_i,v_i')\in V_i^2}\Big\{&-\int_{v_{-i}}\int_{\theta,m_{-i}}\tilde{u}_i\big(x^{W|2}(v_i,v_{-i},m_i,m_{-i}),v,\theta\big)d\Psi^W_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i})\\
&+\int_{v_{-i}}\int_{\theta,m_{-i}}\tilde{u}_i\big(x^{W|2}(v_i',v_{-i},m_i',m_{-i}),v,\theta\big)d\Psi^W_{m_i}(\theta,m_{-i})dF_{-i}(v_{-i}),0\Big\}
\end{aligned}$$

which is finite by Assumption 2. Set $\beta_i^{m_i\to m_i}$ equal to 0. Let $\vec{t}_i(m_i)=\big(t_i(m_i,m_{-i})\big)_{m_{-i}\in M_{-i}}$ be the column vector representing agent i's transfer by reporting m_i. Since Ψ has full rank and thus is invertible, we can find $\mathbf{t}_i := \left(\vec{t}_i(m_i)\right)_{m_i \in M_i}$ solving

$$\begin{pmatrix} \Psi_1(1) & \cdots & \Psi_1(K) \\ \vdots & \ddots & \vdots \\ \Psi_K(1) & \cdots & \Psi_K(K) \end{pmatrix} \begin{pmatrix} t_i(1,1) & \cdots & t_i(K,1) \\ \vdots & \ddots & \vdots \\ t_i(1,K) & \cdots & t_i(K,K) \end{pmatrix} = \begin{pmatrix} \beta_i^{1\to 1} & \cdots & \beta_i^{1\to K} \\ \vdots & \ddots & \vdots \\ \beta_i^{K\to 1} & \cdots & \beta_i^{K\to K} \end{pmatrix}.$$

Next, we show that $(\Phi^W, x^{W|2}, \mathbf{t})$ satisfies the constraints in (P). It is feasible, that is, $(x^{W|2}, \mathbf{t}) \in \Delta(\mathscr{A}) \times \mathbb{R}^N$, because we can define $p_i' = p_i - t_i$ for each agent. Pick any (m_i, v_i) and (m_i', v_i') such that $m_i \neq m_i'$, we have $\tilde{U}_i(v_i, m_i; v_i, m_i) - \tilde{U}_i(v_i', m_i'; v_i, m_i) =$

$$\begin{aligned}
&\int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^{W|2}(v_i, v_{-i}, m_i, m_{-i}), v_i, v_{-i}, \theta\big) d\Psi_{m_i}^W(\theta, m_{-i}) dF_{-i}(v_{-i}) - \vec{\Psi}_{m_i} \cdot \vec{t}_i(m_i) \\
&- \int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^{W|2}(v_i', v_{-i}, m_i', m_{-i}), v_i, v_{-i}, \theta\big) d\Psi_{m_i}^W(\theta, m_{-i}) dF_{-i}(v_{-i}) + \vec{\Psi}_{m_i} \cdot \vec{t}_i(m_i') \\
&\geq -\beta_i^{m_i \to m_i'} - \beta_i^{m_i \to m_i} + \beta_i^{m_i \to m_i'} = -\beta_i^{m_i \to m_i} = 0,
\end{aligned}$$

which means agent i will never misreport his signal, regardless of his report about his type. On the other hand, if agent i reports his true signal, he will find it optimal to also truthfully report his private type. Thus, $(\Phi^W, x^{W|2}, \mathbf{t})$ satisfies all $(BIC_{m_i, v_i \to m_i', v_i'})$. Notice that for any (m_i, v_i) we have $\tilde{U}_i(v_i, m_i; v_i, m_i) =$

$$\begin{aligned}
&\int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^{W|2}(v_i, v_{-i}, m_i, m_{-i}), v_i, v_{-i}, \theta\big) d\Psi_{m_i}^W(\theta, m_{-i}) dF_{-i}(v_{-i}) - \vec{\Psi}_{m_i} \cdot \vec{t}_i(m_i) \\
&= \int_{v_{-i}} \int_{\theta} \tilde{u}_i\big(\tilde{x}^*(v_i, v_{-i}, \theta), v, \theta\big) dF_0(\theta) dF_{-i}(v_{-i}) - 0 \geq 0,
\end{aligned}$$

then all (IIR_{m_i, v_i}) are also satisfied. Because the expected transfer with respect to $\mathbf{t}$ is 0 and $x^{W|2}$ achieves the same outcome as x^W on the equilibrium path, we conclude that $(\Phi^W, x^{W|2}, \mathbf{t})$ is the solution to (P). The remaining thing is to extend our result to the case where Ψ does not have full rank. The following lemma guarantees that we can always have a full-rank belief matrix by enlarging the signal set.

Lemma 13. If Ψ is not invertible, then by adding finitely many additional signals to each agent's signal set, we can construct a new information disclosure policy which induces a full-rank belief matrix.

Proof of Lemma 13. See Appendix B.1.8. □

A.6 Proof of Theorem 4

Let $\Upsilon := \big\{\{\theta', \theta''\} \mid \theta' \neq \theta''\big\}$ be the collection of all unordered pairs in Θ, where each element is labelled υ. Since there exists a rich $\mathbb{S}$, we can separate any $\{\theta_\upsilon', \theta_\upsilon''\} \in \Upsilon$ by a sample-product procedure $S(\upsilon)$, where without loss of generality we assume that

$$s'_\upsilon \prec^{\theta'_\upsilon}_{i_\upsilon} s''_\upsilon \quad s'_\upsilon \prec^{\theta'_\upsilon}_{j_\upsilon} s''_\upsilon$$
$$s'_\upsilon \prec^{\theta''_\upsilon}_{i_\upsilon} s''_\upsilon \quad s'_\upsilon \succ^{\theta''_\upsilon}_{j_\upsilon} s''_\upsilon.$$

That is to say, the two buyers share the same preference under θ'_υ, but have different rankings under θ''_υ; and buyer i_υ's preferred sample product is the same under both states, but buyer j_υ's preference gets reversed as state varies. The Choice Pair (A_υ, B_υ) in sample-product procedure $S(\upsilon)$ takes the value from $\{(s'_\upsilon, s''_\upsilon), (s''_\upsilon, s'_\upsilon)\}$ equally likely. Let $\{(A_\upsilon, B_\upsilon)\}_{\upsilon\in\Upsilon}$ be mutually independent.

We first check that individual buyer will not gain any new information about the state in this approach. Buyer $i \notin \bigcup_{\upsilon\in\Upsilon}\{i_\upsilon, j_\upsilon\}$ does not take any trial, and thus has no means to update his belief. Buyers, who take one or more trials, will get an independent testing result in each trial which is either "$A_\upsilon \prec B_\upsilon$" or "$A_\upsilon \succ B_\upsilon$" with equal probabilities regardless of the realization of the state, and thus will not update the prior F_0, either.

Next, we show that the seller can infer the true state after collecting all the testing results. The outcome of sample-product procedure $S(\upsilon)$ is denoted by $o_\upsilon \in \{o'_\upsilon, o''_\upsilon\}$, where o'_υ means the two buyers share the same feedback, that is, $m_{i_\upsilon} = m_{j_\upsilon}$; and o''_υ means they have opposite feedbacks, that is, $m_{i_\upsilon} \neq m_{j_\upsilon}$. On observing o_υ, the subset of states that will occur with strictly positive probabilities in the seller's posterior belief, denoted by $\mathscr{P}(o_\upsilon)$, satisfies: $\mathscr{P}(o'_\upsilon) = I_{S(\upsilon)}(\theta'_\upsilon)$ and $\mathscr{P}(o''_\upsilon) = I_{S(\upsilon)}(\theta''_\upsilon)$. Let $o = (o_\upsilon)_{\upsilon\in\Upsilon}$ be the realized testing result profile, then to meet the aggregately revealing requirement, $\mathscr{P}(o) = \bigcap_{\upsilon\in\Upsilon}\mathscr{P}(o_\upsilon)$ must have a unique element. Suppose there exist $\theta_1 \neq \theta_2$ both belonging to $\mathscr{P}(o)$, then $\{\theta_1, \theta_2\} \subseteq \mathscr{P}(o_\upsilon)$ for all υ; particularly, this is true for $\tilde{\upsilon}$ corresponding to sample-product procedure $S(\theta_1, \theta_2)$. But then we get a contradiction, since $S(\theta_1, \theta_2)$ separates θ_1 from θ_2 which means we can never have $\{\theta_1, \theta_2\} \subseteq \mathscr{P}(o_{\tilde{\upsilon}})$.

A.7 Proof of Theorem 5

We prove that (Ξ^{IS}, x^{IS}) satisfies all the constraints in (P). Since $|\tilde{I}_i| \leq N-3$, by Lemma 6, agent i knowing $\mathscr{E}_i$ cannot update his belief about θ. Moreover, the principal can identify the agent who misreports and infer the true state from the other agents' reports. Property (iii) guarantees that it is feasible to make x^{IS} unchanged by unilateral deviation. Thus, given the other agents telling the truth, the interim expected utility of agent i with v_i knowing $\mathscr{E}_i$ by reporting $(\hat{v}_i, \hat{m}_i)$ is

$$\int_{v_{-i}} \int_{\theta, m_{-i}} u_i\big(x^{IS}(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v, \theta\big) d\Psi^{IS}_{\mathscr{E}_i}(\theta, m_{-i}) dF_{-i}(v_{-i})$$
$$= \int_{v_{-i}} \int_{\theta} u_i\big(x^*(\hat{v}_i, v_{-i}, \theta), v, \theta\big) dF_0(\theta) dF_{-i}(v_{-i}),$$

which is independent of $\hat{m}_i$, regardless of what $\hat{v}_i$ agent i reports to the principal. Thus, no agent has incentive to misreport his signal, and we get $(BIC_{m_i, v_i \to m'_i, v'_i})$ and (IIR_{m_i, v_i}) immediately from $(BIC_{v_i \to v'_i})$ and (IIR_{v_i}). Notice that the objective function is

$$\int_{v,\theta} \int_m u_0\big(x^{IS}(v,m), v, \theta\big) d\Xi^{IS}_\theta(m) dF_0(\theta) dF_V(v) = \int_{v,\theta} u_0\Big(x^*(v,\theta), v, \theta\Big) dF_0(\theta) dF_V(v),$$

which achieves the upper bound defined by (P^*), then (Ξ^{IS}, x^{IS}) is optimal.

Appendix B

Supplemental Material to Chapter 1

B.1 Omitted Proofs

B.1.1 Proof of Lemma 3

First, we consider the case $N = 3$, and let $i, j, k \in \{1,2,3\}$. Suppose there exist agent i and $m_i \in M_i$ such that $\Xi_{\tilde{\theta}}(m_i, \tilde{m}_{-i}) > 0$ and $\Xi_{\hat{\theta}}(m_i, \hat{m}_{-i}) > 0$ for some $\tilde{\theta} \neq \hat{\theta}$ and $\tilde{m}_{-i}, \hat{m}_{-i} \in M_{-i}$. Immediately, we have $\tilde{m}_{-i} \neq \hat{m}_{-i}$; otherwise aggregately revealing property is violated. Suppose $\tilde{m}_j = \hat{m}_j$ for some $j \neq i$, then for $k \neq i, j$ we have $\tilde{m}_k \neq \hat{m}_k$ and $\Xi_{\tilde{\theta}}(\tilde{m}_k, m_i, \tilde{m}_j) > 0$ and $\Xi_{\hat{\theta}}(\hat{m}_k, m_i, \hat{m}_j) > 0$, which contradicts (1) of innocuous unilateral deviation property. Thus, $\tilde{m}_j \neq \hat{m}_j$ and $\tilde{m}_k \neq \hat{m}_k$. Consider the signal profile $(m_i, \hat{m}_j, \tilde{m}_k)$, and we must have $\Xi_{\hat{\theta}}(m_i, \hat{m}_j, \tilde{m}_k) = 0$; otherwise we have $\Xi_{\tilde{\theta}}(m_i, \hat{m}_j, \tilde{m}_k) > 0$ and $\Xi_{\hat{\theta}}(m_i, \hat{m}_j, \hat{m}_k) > 0$, contradicting (1). It follows that unilateral deviation is detected for $(m_i, \hat{m}_j, \tilde{m}_k)$, since $\Xi_{\theta}(m_i, \hat{m}_j, \tilde{m}_k) = 0, \forall \theta \in \Theta$. Then by (2) of innocuous unilateral deviation property, we must have $\Xi_{\theta}(m_i, \hat{m}_j, m_k) = 0, \forall \theta \neq \tilde{\theta}$ and $\forall m_k \in M_k$, contradicting that $\Xi_{\hat{\theta}}(m_i, \hat{m}_j, \hat{m}_k) > 0$. Thus, for any agent i and any $m_i \in M_i$, there exists at most one θ such that $\Xi_{\theta}(m_i, m_{-i}) > 0$ for some $m_{-i} \in M_{-i}$.

Second, we assume that $N = 2$. Suppose there exist agent i and $m_i \in M_i$ such that $\Xi_{\tilde{\theta}}(m_i, \tilde{m}_j) > 0$ and $\Xi_{\hat{\theta}}(m_i, \hat{m}_j) > 0$ for some $\tilde{\theta} \neq \hat{\theta}$ and $\tilde{m}_j, \hat{m}_j \in M_j$. Through the same argument as in case $N = 3$, we have $\tilde{m}_j \neq \hat{m}_j$, which contradicts (1).

B.1.2 Proof of Lemma 4

On receiving m_i, the interim expected payoff of agent i with type v_i by reporting $\hat{v}_i$ is given by

$$U_i(v_i, \hat{v}_i \mid m_i) = \int_{v_{-i}} \int_{\theta, m_{-i}} q_i(\hat{v}_i, v_{-i}, m) y_i(v_i, \theta) - p_i(\hat{v}_i, v_{-i}, m) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}).$$

Then fixed any $v_i < v_i'$, from ($BIC_{v_i \to v_i' | m_i}$) and ($BIC_{v_i' \to v_i | m_i}$) we can derive the monotonicity condition ($Mon_{v_i < v_i'}$). On the other hand, Bayesian incentive compatibility requires that truthful report of one's type is the best response for each agent with any type if the other agents tell the truth, that is,

$$v_i \in \arg\max_{\hat{v}_i} U_i(v_i, \hat{v}_i \mid m_i).$$

Let $U_i(v_i \mid m_i) = \max_{\hat{v}_i} U_i(v_i, \hat{v}_i \mid m_i) = U_i(v_i, v_i \mid m_i)$. By Envelope Theorem we have

$$\frac{dU_i(v_i \mid m_i)}{dv_i} = \int_{v_{-i}} \int_{\theta, m_{-i}} q_i(v_i, v_{-i}, m) \frac{dy_i(v_i, \theta)}{dv_i} d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}).$$

Then we have

$$U_i(v_i \mid m_i) = U_i(\underline{v}_i \mid m_i) + \int_{\underline{v}_i}^{v_i} \int_{v_{-i}} \int_{\theta, m_{-i}} q_i(\tilde{v}_i, v_{-i}, m) \frac{dy_i(\tilde{v}_i, \theta)}{dv_i} d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) d\tilde{v}_i,$$

where $\underline{v}_i = \min V_i$. Based on integration by parts, the principal's ex ante total expected payments can be written as follows:

$$\begin{aligned}
&\int_\theta \int_m \int_v \sum_{i=1}^N p_i(v, m) dF_V(v) d\Xi_\theta(m) dF_0(\theta) = \int_m \int_\theta \int_v \sum_{i=1}^N p_i(v, m) dF_V(v) d\Psi_m(\theta) d\Lambda(m) \\
=& \sum_{i=1}^N \int_{m_i} \int_{v_i} \Big(\int_{v_{-i}} \int_{\theta, m_{-i}} q_i(v, m) y_i(v_i, \theta) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\
&- \int_{\underline{v}_i}^{v_i} \int_{v_{-i}} \int_{\theta, m_{-i}} q_i(\tilde{v}_i, v_{-i}, m) \frac{dy_i(\tilde{v}_i, \theta)}{dv_i} d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) d\tilde{v}_i \Big) dF_i(v_i) d\Lambda_i(m_i) \\
=& \sum_{i=1}^N \int_{m_i} \Big(\int_v \int_{\theta, m_{-i}} q_i(v, m) y_i(v_i, \theta) d\Psi_{m_i}(\theta, m_{-i}) dF_V(v) \\
&- \int_{v_i} \int_{v_{-i}} \int_{\theta, m_{-i}} q_i(v_i, v_{-i}, m) \frac{dy_i(v_i, \theta)}{dv_i} \big(1 - F_i(v_i)\big) d\Psi_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) dv_i \Big) d\Lambda_i(m_i) \\
=& \sum_{i=1}^N \int_{m_i} \Big(\int_v \int_{\theta, m_{-i}} q_i(v, m) y_i(v_i, \theta) d\Psi_{m_i}(\theta, m_{-i}) dF_V(v) \\
&- \int_v \int_{\theta, m_{-i}} q_i(v, m) \frac{dy_i(v_i, \theta)}{dv_i} \frac{1 - F_i(v_i)}{f_i(v_i)} d\Psi_{m_i}(\theta, m_{-i}) dF_V(v) \Big) d\Lambda_i(m_i) \\
=& \sum_{i=1}^N \int_{m_i} \int_v \int_{\theta, m_{-i}} q_i(v, m) \Big(y_i(v_i, \theta) - \frac{dy_i(v_i, \theta)}{dv_i} \frac{1 - F_i(v_i)}{f_i(v_i)} \Big) d\Psi_{m_i}(\theta, m_{-i}) dF_V(v) d\Lambda_i(m_i) \\
=& \int_{\theta, m} \int_v \sum_{i=1}^N q_i(v, m) \gamma_i(v_i, \theta) dF_V(v) d\Phi(\theta, m).
\end{aligned}$$

Thus, the principal's objective function becomes

$$\int_{\theta,m}\int_v \Big(y_0\big(q(v,m),v,\theta\big)+\sum_{i=1}^N q_i(v,m)\gamma_i(v_i,\theta)\Big)dF_V(v)d\Phi(\theta,m)$$
$$=\int_v\int_\theta \Big[y_0\big(\chi(v,\theta),v,\theta\big)+\sum_{i=1}^N \chi_i(v,\theta)\gamma_i(v_i,\theta)\Big]dF_0(\theta)dF_V(v),$$

where $\chi_i(v,\theta)=\int_m q_i(v,m)d\Xi_\theta(m)$.

B.1.3 Proof of Proposition 2

Since none of the monotonicity constraints in (P'_{pub}) is binding, the principal's maximum ex ante expected payoff through public disclosure is equal to

$$\Pi_{pub}=\sup_{q(v,\theta)\in Q}\int_\theta\int_v \Big(y_0\big(q(v,\theta),v,\theta\big)+\sum_{i=1}^N q_i(v,\theta)\gamma_i(v_i,\theta)\Big)dF_V(v)dF_0(\theta).$$

Let (Φ,q) be any candidate solution to problem (P'_1). Since for any $(v,m)\in V\times M$, we have $q(v,m)\in Q$, then from the convexity of Q we get

$$\chi(v,\theta)=\int_m q(v,m)d\Xi_\theta(m)\in Q,$$

which means

$$\int_v\int_\theta \Big[y_0\big(\chi(v,\theta),v,\theta\big)+\sum_{i=1}^N \chi_i(v,\theta)\gamma_i(v_i,\theta)\Big]dF_0(\theta)dF_V(v)\le \Pi_{pub}.$$

Notice that the optimal public disclosure mechanism is also a feasible private disclosure mechanism, then we conclude that the principal gets the same maximum expected payoff between private disclosure and public disclosure.

B.1.4 Proof of Proposition 3

Take any private disclosure mechanism that exhibits full disclosure, then for any i,m_i, the marginal of $\Psi_{m_i}(\theta,m_{-i})$ over Θ is degenerated, which means we can rewrite the constraints in (P'_1) as follows: for any i, m_i, $v_i<v'_i$, we have

$$\int_{\theta,m_{-i}}\Big(\mathbb{E}_{v_{-i}}[q_i(v'_i,v_{-i},m)]-\mathbb{E}_{v_{-i}}[q_i(v_i,v_{-i},m)]\Big)\big(y_i(v'_i,\theta)-y_i(v_i,\theta)\big)d\Psi_{m_i}(\theta,m_{-i})\ge 0$$
$$\iff \int_{m_{-i}}\mathbb{E}_{v_{-i}}[q_i(v'_i,v_{-i},m)]-\mathbb{E}_{v_{-i}}[q_i(v_i,v_{-i},m)]d\Psi_{m_i}(m_{-i})\cdot\big(y_i(v'_i,\theta)-y_i(v_i,\theta)\big)\ge 0$$
$$\iff \mathbb{E}_{v_{-i},m_{-i}}[q_i(v'_i,v_{-i},m)\mid m_i]\ge \mathbb{E}_{v_{-i},m_{-i}}[q_i(v_i,v_{-i},m)\mid m_i].$$

Let $\big(q_i^*(v,m)\big)_{i,v,m}$ be the solution to (P_1'), and define $q_i^+(v;\theta) = \int_m q_i^*(v,m)d\Xi_\theta(m)$, then we have for any $i, \theta_i, v_i < v_i'$,

$$\mathbb{E}_{v_{-i}}[q_i^+(v_i', v_{-i}, \theta)] = \int_{m_i} \mathbb{E}_{v_{-i},m_{-i}}[q_i^*(v_i', v_{-i}, m) \mid m_i] d\Lambda_i(m_i \mid \theta)$$
$$\geq \int_{m_i} \mathbb{E}_{v_{-i},m_{-i}}[q_i^*(v_i, v_{-i}, m) \mid m_i] d\Lambda_i(m_i \mid \theta) = \mathbb{E}_{v_{-i}}[q_i^+(v_i, v_{-i}, \theta)],$$

which means $\big(q_i^+(v,\theta)\big)_{i,v,\theta}$ is feasible in the optimal public disclosure problem (P_{pub}'). Since $\big(q_i^+(v,\theta)\big)_{i,v,\theta}$ induces the same expected payoff as $\big(q_i^*(v,m)\big)_{i,v,m}$, we conclude that any private information mechanism that exhibits full disclosure achieves at most the same payoff as in the optimal public disclosure mechanism. Thus, creating some uncertainty over θ for some agent is necessary for private disclosure mechanism to strictly outperform public disclosure mechanism.

B.1.5 Proof of Lemma 5

We prove the lemma by checking the assumptions of Theorem 13.5 in Kadan, Reny, and Swinkels (2017), hereafter KRS. Notice that (P^*) is pure adverse selection problem, then we don't have agents' action spaces and signal spaces as in KRS, which means their Assumptions 5.6-5.9 are irrelevant. As in KRS, for each agent $i \in I$, we define the single-agent-i model, where the "reward space" is $R_i = \mathscr{A} \times V_{-i} \times \Theta$ and the type space is $T_i = V_i$. Since $(v_i)_{i\in I}$ and θ are mutually independent in our setup, we have that for any $(r_i, t_i) = \big((a, v_{-i}, \theta), v_i\big) \in R_i \times T_i$, agent i's utility function is $\tilde{u}_i(r_i, t_i) = u_i(a, v, \theta)$ and the principal's utility function is $\tilde{u}_0(r_i, t_i) = u_0(a, v, \theta)$. A mechanism for this single-agent-i model is a probability measure $x_i : V_i \to \Delta(R_i)$.

First we show that Assumptions 5.1-5.5 in KRS are satisfied in the single-agent-i model. Under the conditions of Lemma 5, we only need to check Assumption 5.4, that is, for any $t_i \in V_i$, for any $c \in \mathbb{R}$, the closure of $L(t_i, c) := \{r_i \in R_i \mid \tilde{u}_0(r_i, t_i) \geq c\}$ is compact. Since u_0 is continuous, we have $L(t_i, c)$ is a closed subset of R_i. Notice that $\mathscr{A}$, $(V_j)_{j\neq i}$ and Θ are all compact sets, then their Cartesian product R_i is also compact. Then $L(t_i, c)$ is compact due to the fact that closed subsets of a compact set are also compact.[1]

[1] Let X be the topological space containing R_i. Let O be an indexing set and $\mathscr{F} = \{P_o \mid o \in O\}$ be an arbitrary open cover for $L(t_i, c)$. Since $X \setminus L(t_i, c)$ is open, it follows that $\mathscr{F}$ together with $X \setminus L(t_i, c)$ is an open cover for R_i. Thus, R_i can be covered by a finite number of sets, denoted by $P_1, \ldots, P_K$, drawn from $\mathscr{F}$ together with possibly $X \setminus L(t_i, c)$. Since $L(t_i, c) \in R_i$, we have that $\{P_1, \ldots, P_K\}$ cover $L(t_i, c)$, and it follows that $L(t_i, c)$ is compact.

Next, we check the consequence of interim individual rationality constraints. As in Example 11.7 in KRS, the fact that agents can opt out after learning their types imposes the following restriction on the feasible mechanisms:

$$\int_{R_i} \tilde{u}_i(r_i,t_i) dx_i(r_i \mid t_i) \geq 0,$$

which satisfies the conditions of Corollary 11.1 in KRS. It follows that the single-agent-i model possesses a solution as long as the feasible set is nonempty. Thus, such restriction on the mechanism imposed by the (IIR_{v_i}) constraints does no harm.

Finally, we apply Theorem 13.5 in KRS and complete the proof since the above argument holds for any single-agent-i model.

B.1.6 Proof of Lemma 6

To prove Lemma 6, we first prove the following lemma:

Lemma 14. Define θ and K as in Subsection 1.6.3. Let $\omega_1 \in \{1,\dots,K\}$ be a random variable which is independent of θ and follows the discrete uniform distribution. Define $\omega_2 = A\cdot\theta + B\cdot\omega_1 \mod K$, where A,B are two non-zero integers such that K cannot exactly divide B. Then for any event $\mathscr{E}$ which induces a conditional distribution of ω_2, denoted by $F_{\mathscr{E}} \in \Delta(\{1,\dots,K\})$, we have $F_0(\theta \mid \mathscr{E}) = F_0(\theta)$ for any θ.

Proof of Lemma 14. For any $\theta \in \{1,\dots,T\}$ and $\omega_2 \in \{1,\dots,K\}$, we can uniquely pin down ω_1. Suppose not, then there exist $\omega_1 \neq \omega_1'$ satisfying

$$\begin{cases} \omega_2 = A\cdot\theta + B\cdot\omega_1 \mod K \\ \omega_2 = A\cdot\theta + B\cdot\omega_1' \mod K, \end{cases}$$

which means $B\cdot(\omega_1 - \omega_1') \equiv 0 \pmod K$. Notice that the prime number K cannot exactly divide B, then $(\omega_1 - \omega_1') \in [1-K, K-1]$ must be exactly divided by K, which is impossible. Then we have

$$\Pr(\omega_2 \mid \theta) = \Pr\big(\omega_1 \text{ s.t. } A\cdot\theta + B\cdot\omega_1 \equiv \omega_2 \pmod K \mid \theta\big) = \Pr(\omega_1 \mid \theta) = \Pr(\omega_1) = \frac{1}{K}.$$

It follows that

$$\Pr(\theta \mid \omega_2) = \frac{\Pr(\theta,\omega_2)}{\Pr(\omega_2)} = \frac{\Pr(\omega_2 \mid \theta)\Pr(\theta)}{\sum_\theta \Pr(\omega_2 \mid \theta)\Pr(\theta)} = \frac{\frac{1}{K}\Pr(\theta)}{\sum_\theta \frac{1}{K}\Pr(\theta)} = \Pr(\theta).$$

Then we have

$$\begin{aligned} F_0(\theta \mid \mathscr{E}) &= \sum_{\omega_2} \Pr(\theta \mid \omega_2, \mathscr{E}) \Pr(\omega_2 \mid \mathscr{E}) = \sum_{\omega_2} \Pr(\theta \mid \omega_2) F_{\mathscr{E}}(\omega_2) \\ &= \sum_{\omega_2} \Pr(\theta) F_{\mathscr{E}}(\omega_2) = \Pr(\theta) \sum_{\omega_2} F_{\mathscr{E}}(\omega_2) = F_0(\theta). \end{aligned}$$

□

Next, we prove Lemma 6. We distinguish four cases depending on how many signals in $(m_i)_{i\in\tilde{I}}$ that are drawn from $\{m_{N-2}, m_{N-1}, m_N\}$. We begin with properties (i) and (ii):

Case 1. If $\tilde{I} \subseteq \{1,\dots,N-3\}$, then we always have $|\tilde{I}| \leq N-3$. Since $m_i = \varepsilon_i$ for $i = 1,\dots,N-3$, by mutual independence of $(\theta, \varepsilon_1, \dots, \varepsilon_{N-3})$, we have $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta) = F_0(\theta)$ for all $\theta \in \Theta$.

Case 2. $\tilde{I} = \mathring{I} \cup \{N-2+n\}$, where $\mathring{I} \subseteq \{1,\dots,N-3\}$ and $n \in \{0,1,2\}$. Notice that

$$m_{N-2+n} = \theta + 2^n \cdot \varepsilon_1 + \cdots + (\tau+1)^n \cdot \varepsilon_\tau + \cdots + (N-2)^n \cdot \varepsilon_{N-3} \mod K,$$

then if $|\mathring{I}| \leq N-4$ (and thus $|\tilde{I}| \leq N-3$), there exists $\tau \in \{1,\dots,N-3\} \setminus \mathring{I}$ such that

$$\theta + (\tau+1)^n \cdot \varepsilon_\tau \equiv m_{N-2+n} - \sum_{i\in\mathring{I}} (i+1)^n \cdot m_i - \sum_{\substack{i\in\{1,\dots,N-3\} \\ i\notin\mathring{I},\, i\neq\tau}} (i+1)^n \cdot \varepsilon_i \pmod{K}.$$

Let ω_1 be ε_τ, and ω_2 be the residue of the right hand side, modulo K. The event $\mathscr{E}$ corresponds to the realization of $(m_i)_{i\in\tilde{I}}$. Then we have $\theta + (\tau+1)^n \cdot \omega_1 \equiv \omega_2 \pmod{K}$. Since K cannot exactly divide $(\tau+1)^n$, from Lemma 14 we get $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta) = F_0(\theta)$. So we establish property (i) for this case. If $|\mathring{I}| = N-3$ (and thus $|\tilde{I}| = N-2$), we have

$$\theta \equiv m_{N-2+n} - \sum_{i\in\mathring{I}} (i+1)^n \cdot m_i \pmod{K}. \tag{R1}$$

Thus, θ is uniquely pinned down by $(m_i)_{i\in\tilde{I}}$, which establishes property (ii).

Case 3. $\tilde{I} = \mathring{I} \cup \{N-2+n_1, N-2+n_2\}$, where $\mathring{I} \subseteq \{1,\dots,N-3\}$ and $n_1, n_2 \in \{0,1,2\}$. We further distinguish three cases depending on the value of (n_1, n_2).

(Case 3.1) If $(n_1, n_2) = (0,1)$, then we have

$$\begin{cases} m_{N-2} = \theta + \sum_{i\in\{1,\dots,N-3\}\setminus\mathring{I}} \varepsilon_i + \sum_{i\in\mathring{I}} m_i \mod K \\ m_{N-1} = \theta + \sum_{i\in\{1,\dots,N-3\}\setminus\mathring{I}} (i+1)\varepsilon_i + \sum_{i\in\mathring{I}} (i+1) m_i \mod K. \end{cases}$$

If $|\mathring{I}| \leq N-5$ (and thus $|\tilde{I}| \leq N-3$), there exist $\tau_1, \tau_2 \in \{1,\ldots,N-3\} \setminus \mathring{I}$ such that

$$\begin{cases} \theta + \varepsilon_{\tau_1} + \varepsilon_{\tau_2} \equiv m_{N-2} - \sum_{i\in\mathring{I}} m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} \varepsilon_i \pmod K \\ \theta + (\tau_1+1)\varepsilon_{\tau_1} + (\tau_2+1)\varepsilon_{\tau_2} \equiv m_{N-1} - \sum_{i\in\mathring{I}} (i+1)m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} (i+1)\varepsilon_i \pmod K. \end{cases}$$

Then we have

$$\tau_1\theta + (\tau_1-\tau_2)\varepsilon_{\tau_2} \equiv (\tau_1+1)m_{N-2} - m_{N-1} - \sum_{i\in\mathring{I}} (\tau_1-i)m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} (\tau_1-i)\varepsilon_i \pmod K.$$

Since K cannot exactly divide $(\tau_1-\tau_2)$, then from Lemma 14 we get $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta) = F_0(\theta)$. If $|\mathring{I}| = N-4$ (and thus $|\tilde{I}| = N-2$), there exists only one $\tau_1 \in \{1,\ldots,N-3\} \setminus \mathring{I}$, then

$$\tau_1\theta \equiv (\tau_1+1)m_{N-2} - m_{N-1} - \sum_{i\in\mathring{I}} (\tau_1-i)m_i \pmod K. \tag{R2-1}$$

Since τ_1 cannot be exactly divided by K, then θ is uniquely pinned down by $(m_i)_{i\in\tilde{I}}$.
(Case 3.2) If $(n_1,n_2) = (0,2)$, then we have

$$\begin{cases} m_{N-2} = \theta + \sum_{i\in\{1,\ldots,N-3\}\setminus\mathring{I}} \varepsilon_i + \sum_{i\in\mathring{I}} m_i \mod K \\ m_N = \theta + \sum_{i\in\{1,\ldots,N-3\}\setminus\mathring{I}} (i+1)^2\varepsilon_i + \sum_{i\in\mathring{I}} (i+1)^2 m_i \mod K. \end{cases}$$

If $|\mathring{I}| \leq N-5$ (and thus $|\tilde{I}| \leq N-3$), there exist $\tau_1, \tau_2 \in \{1,\ldots,N-3\} \setminus \mathring{I}$ such that

$$\begin{cases} \theta + \varepsilon_{\tau_1} + \varepsilon_{\tau_2} \equiv m_{N-2} - \sum_{i\in\mathring{I}} m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} \varepsilon_i \pmod K \\ \theta + (\tau_1+1)^2\varepsilon_{\tau_1} + (\tau_2+1)^2\varepsilon_{\tau_2} \equiv m_N - \sum_{i\in\mathring{I}} (i+1)^2 m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} (i+1)^2\varepsilon_i \pmod K. \end{cases}$$

Then we have

$$\tau_1(\tau_1+2)\theta + (\tau_1-\tau_2)(\tau_1+\tau_2+2)\varepsilon_{\tau_2} \equiv \mathscr{R}\Big((m_i)_{i\in\tilde{I}}, (\varepsilon_i)_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}}\Big) \pmod K,$$

where

$$\mathscr{R} = (\tau_1+1)^2 m_{N-2} - m_N - \sum_{i\in\mathring{I}} (\tau_1-i)(\tau_1+i+2)m_i - \sum_{\substack{i\in\{1,\ldots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}} (\tau_1-i)(\tau_1+i+2)\varepsilon_i.$$

Since $(\tau_1-\tau_2)(\tau_1+\tau_2+2)$ cannot be exactly divided by K, then from Lemma 14 we get $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta)=F_0(\theta)$. If $|\mathring{I}|=N-4$ (and thus $|\tilde{I}|=N-2$), then we have

$$\tau_1(\tau_1+2)\theta\equiv(\tau_1+1)^2m_{N-2}-m_N-\sum_{i\in\mathring{I}}(\tau_1-i)(\tau_1+i+2)m_i \mod K. \tag{R2-2}$$

Since K cannot exactly divide $\tau_1(\tau_1+2)$, then θ is uniquely pinned down by $(m_i)_{i\in\tilde{I}}$.

(Case 3.3) If $(n_1,n_2)=(1,2)$, then we have

$$\begin{cases} m_{N-1}=\theta+\sum\limits_{i\in\{1,\dots,N-3\}\setminus\mathring{I}}(i+1)\varepsilon_i+\sum\limits_{i\in\mathring{I}}(i+1)m_i \mod K\\ m_N=\theta+\sum\limits_{i\in\{1,\dots,N-3\}\setminus\mathring{I}}(i+1)^2\varepsilon_i+\sum\limits_{i\in\mathring{I}}(i+1)^2m_i \mod K.\end{cases}$$

If $|\mathring{I}|\le N-5$ (and thus $|\tilde{I}|\le N-3$), there exist $\tau_1,\tau_2\in\{1,\dots,N-3\}\setminus\mathring{I}$ such that

$$\begin{cases} \theta+(\tau_1+1)\varepsilon_{\tau_1}+(\tau_2+1)\varepsilon_{\tau_2}\equiv m_{N-1}-\sum\limits_{i\in\mathring{I}}(i+1)m_i-\sum\limits_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}}(i+1)\varepsilon_i \pmod K\\ \theta+(\tau_1+1)^2\varepsilon_{\tau_1}+(\tau_2+1)^2\varepsilon_{\tau_2}\equiv m_N-\sum\limits_{i\in\mathring{I}}(i+1)^2m_i-\sum\limits_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}}(i+1)^2\varepsilon_i \pmod K.\end{cases}$$

Then we have

$$\tau_1\theta+(\tau_1-\tau_2)(\tau_2+1)\varepsilon_{\tau_2}\equiv\mathscr{R}'\Big((m_i)_{i\in\tilde{I}},(\varepsilon_i)_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}}\Big) \pmod K,$$

where

$$\mathscr{R}'=(\tau_1+1)m_{N-1}-m_N-\sum_{i\in\mathring{I}}(\tau_1-i)(i+1)m_i-\sum_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2}}(\tau_1-i)(i+1)\varepsilon_i.$$

Since K cannot exactly divide $(\tau_1-\tau_2)(\tau_2+1)$, by Lemma 14 we get $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta)=F_0(\theta)$. If $|\mathring{I}|=N-4$ (and thus $|\tilde{I}|=N-2$), then we have

$$\tau_1\theta\equiv(\tau_1+1)m_{N-1}-m_N-\sum_{i\in\mathring{I}}(\tau_1-i)(i+1)m_i \pmod K. \tag{R2-3}$$

Since τ_1 cannot be exactly divided by K, then θ is uniquely pinned down by $(m_i)_{i\in\tilde{I}}$.

Case 4. $\tilde{I}=\mathring{I}\cup\{N-2,N-1,N\}$, where $\mathring{I}\subseteq\{1,\dots,N-3\}$. We have

$$\begin{cases} m_{N-2}=\theta+\sum\limits_{i\in\{1,\dots,N-3\}\setminus\mathring{I}}\varepsilon_i+\sum\limits_{i\in\mathring{I}}m_i \mod K\\ m_{N-1}=\theta+\sum\limits_{i\in\{1,\dots,N-3\}\setminus\mathring{I}}(i+1)\varepsilon_i+\sum\limits_{i\in\mathring{I}}(i+1)m_i \mod K\\ m_N=\theta+\sum\limits_{i\in\{1,\dots,N-3\}\setminus\mathring{I}}(i+1)^2\varepsilon_i+\sum\limits_{i\in\mathring{I}}(i+1)^2m_i \mod K.\end{cases}$$

If $|\mathring{I}| \leq N-6$ (and thus $|\tilde{I}| \leq N-3$), there exist $\tau_1, \tau_2, \tau_3 \in \{1,\dots,N-3\} \setminus \mathring{I}$ such that

$$\begin{cases} \theta + \sum\limits_{i=\tau_1,\tau_2,\tau_3} \varepsilon_i \equiv m_{N-2} - \sum\limits_{i\in\mathring{I}} m_i - \sum\limits_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2,\tau_3}} \varepsilon_i \pmod K \\ \theta + \sum\limits_{i=\tau_1,\tau_2,\tau_3} (i+1)\varepsilon_i \equiv m_{N-1} - \sum\limits_{i\in\mathring{I}} (i+1) m_i - \sum\limits_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2,\tau_3}} (i+1)\varepsilon_i \pmod K \\ \theta + \sum\limits_{i=\tau_1,\tau_2,\tau_3} (i+1)^2\varepsilon_i \equiv m_{N} - \sum\limits_{i\in\mathring{I}} (i+1)^2 m_i - \sum\limits_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2,\tau_3}} (i+1)^2\varepsilon_i \pmod K. \end{cases}$$

Then we have

$$\tau_1\tau_2(\tau_1+2)\theta + (\tau_1+2)(\tau_1-\tau_3)(\tau_2-\tau_3)\varepsilon_{\tau_3} \equiv \mathscr{R}''\Big((m_i)_{i\in\tilde{I}}, (\varepsilon_i)_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2,\tau_3}}\Big) \pmod K,$$

where

$$\begin{aligned} \mathscr{R}'' = {} & (\tau_1+2)m_N - (\tau_1+2)(\tau_1+\tau_2+2)m_{N-1} + (\tau_1+1)(\tau_1+2)(\tau_2+1)m_{N-2} \\ & - \sum_{i\in\mathring{I}} (\tau_1+2)(\tau_1-i)(\tau_2-i)m_i - \sum_{\substack{i\in\{1,\dots,N-3\}\\ i\notin\mathring{I},\, i\neq\tau_1,\tau_2,\tau_3}} (\tau_1+2)(\tau_1-i)(\tau_2-i)\varepsilon_i. \end{aligned}$$

Since K cannot exactly divide $(\tau_1+2)(\tau_1-\tau_3)(\tau_2-\tau_3)$, then from Lemma 14 we get $\Psi^{IS}_{(m_i)_{i\in\tilde{I}}}(\theta) = F_0(\theta)$. If $|\mathring{I}| = N-5$ (and thus $|\tilde{I}| = N-2$), then we have

$$\begin{aligned} \tau_1\tau_2(\tau_1+2)\theta \equiv {} & (\tau_1+2)m_N - (\tau_1+2)(\tau_1+\tau_2+2)m_{N-1} \\ & + (\tau_1+1)(\tau_1+2)(\tau_2+1)m_{N-2} - \sum_{i\in\mathring{I}} (\tau_1+2)(\tau_1-i)(\tau_2-i)m_i \pmod K. \end{aligned} \tag{R3}$$

Since K cannot exactly divide $\tau_1\tau_2(\tau_1+2)$, θ is uniquely pinned down by $(m_i)_{i\in\tilde{I}}$.

Similarly, we can prove that when $|\tilde{I}| = N-2$, truthful reports of $(m_i)_{i\in\tilde{I}}$ uniquely pin down ε_i for all $i \in \{1,\dots,N-3\} \setminus \mathring{I}$. Then with $\boldsymbol{\varepsilon} = \big(\theta, (\varepsilon_i)_{i\in\{1,\dots,N-3\}}\big)$ at hand, the remaining two signals is given by $m_i = \boldsymbol{\zeta_i} \cdot \boldsymbol{\varepsilon} \mod K$ for $i \in I \setminus \tilde{I}$. Thus, we finish the proof of properties (i) and (ii).

Next, we prove property (iii). From (ii), we can define a mapping

$$\theta^+ : \quad \big\{\text{truthful report } (m_i)_{i\in\tilde{I}} \text{ s.t. } |\tilde{I}| = N-2\big\} \to \Theta,$$

which is given by conditions R1, R2-1, R2-2, R2-2, R2-3 and R3. Notice that in each condition, none of the coefficients of $(m_i)_{i\in\tilde{I}}$ can be exactly divided by K. Thus, we must have

$\theta^+((m_i)_{i\in\tilde{I}}) \neq \theta^+((m_i)_{i\in\tilde{I}\setminus\{j\}}, m'_j)$ for any $m_j \neq m'_j$.[2]

Given an arbitrary report profile m where there exists at most one misreport, pick any $N-2$ signals $(m_i)_{i\in\tilde{I}}$, and by property (ii) we can calculate the remaining two signals. If they both coincide with the reports, then there is no misreport; otherwise, some agent lies about his signal. Suppose that unilateral deviation from truth-telling is detected and the principal cannot tell who misreports, that is, there exist i, j and $m'_i \neq m_i, m'_j \neq m_j$ such that no unilateral deviation is detected for both (m'_i, m_{-i}) and (m'_j, m_{-j}). By property (ii), $N-2$ signals m_{-ij} pin down the state $\theta = \theta^+(m_{-ij})$, as well as the remaining two signals, denoted by $\hat{m}_i, \hat{m}_j$. It follows that $m'_i = \hat{m}_i$, $m_j = \hat{m}_j$ (because (m'_i, m_{-i}) is a truthful report), and $m_i = \hat{m}_i$, $m'_j = \hat{m}_j$ (because (m'_j, m_{-j}) is also a truthful report), contradicting $m'_i \neq m_i, m'_j \neq m_j$. We conclude that the agent who misreports can be identified by the report of N signals.

B.1.7 Proof of Proposition 4

By assumption, the solution to (P^*), denoted by $\{x^*(\theta)\}_{\theta\in\Theta}$, also solves the following unconstrained problem

$$\sup_{x(\theta)\in\Delta(\mathscr{A})} \int_\theta u_0\big(x(\theta), v, \theta\big) dF_0(\theta).$$

It follows that $x^*(\theta) \in \arg\sup_x u_0(x, v, \theta)$ for any $\theta \in \Theta$. After observing any truthful reports $m \in M$ from the agents, the principal knows the true state is $\theta^+(m)$, and the allocation determined by the optimal private disclosure mechanism (Ξ^W, x^W) is given by $x^W(m) = x^*\big(\theta^+(m)\big)$. Thus, the principal already optimizes her payoff by obeying the allocation rule.

[2] Take condition R2-1 for example. Suppose there exist $m_{N-2} \neq m'_{N-2}$ satisfying

$$\begin{cases} \tau_1\theta \equiv (\tau_1+1)m_{N-2} - m_{N-1} - \sum_{i\in\hat{I}}(\tau_1 - i)m_i \pmod K \\ \tau_1\theta \equiv (\tau_1+1)m'_{N-2} - m_{N-1} - \sum_{i\in\hat{I}}(\tau_1 - i)m_i \pmod K, \end{cases}$$

then we have $0 \equiv (\tau_1+1)(m_{N-2} - m'_{N-2}) \pmod K$, which is impossible because K cannot exactly divide (τ_1+1). Thus, a change of one signal must induce a different state.

B.1.8 Proof of Lemma 13

We start from the information disclosure policy Φ^S defined in Section 1.4.1 where $M_i = \{1,2,\dots,T\}$. The associated belief matrix Ψ^S is given by

$$\Psi^S = \begin{pmatrix} \alpha_T & \alpha_1 & \cdots & \alpha_{T-1} \\ \alpha_{T-1} & \alpha_T & \cdots & \alpha_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_T \end{pmatrix},$$

which is a circulant matrix[3] generated by the row vector $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_T)$. The associated polynomial of circulant matrix Ψ^S is given by

$$f(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_T x^{T-1}.$$

Let $\rho_m = e^{\frac{2\pi m}{T} i}$ be the complex T^{th} root of unit, for $m = 0,1,\dots,T-1$, where i is the imaginary unit. Then circulant matrix Ψ^S has eigenvalue

$$\eta_m = \sum_{k=0}^{T-1} \alpha_{k+1} e^{\frac{2\pi mk}{T} i} = f(\rho_m), \text{ for } m = 0,1,\dots,T-1.$$

Thus, Ψ^S is nonsingular if and only if $\eta_m \neq 0$ for any $m \in \{0,1,\dots,T-1\}$. In other words, the necessary and sufficient condition for Ψ^S to be invertible is that, none of the complex T^{th} roots of unit is the root of polynomial $f(x)$. If Ψ^S has full rank, then we have found such invertible Ψ. Otherwise, pick T different prime numbers $\tau_1, \tau_2, \dots, \tau_T$ such that $T < \tau_1 < \tau_2 < \cdots \tau_T$, and for each $r \in \{1,2,\dots,T\}$ define $M_i^r = \{1,2,\dots,\tau_r\}$ for each $i \in I$, and construct the information disclosure policy Φ^r as follows

$$\Phi^r(\theta_t, m) = \begin{cases} \frac{\alpha_t}{\tau_r}, & \text{if } m_{l_1} \equiv m_{l_2} + (l_1 - l_2)t \pmod{\tau_r}, \quad \forall l_1, l_2 \in I \\ 0, & \text{otherwise.} \end{cases}$$

Then for each r, the induced belief matrix Ψ^r is given by

$$\Psi^r = \begin{pmatrix} \alpha_{\tau_r} & \alpha_1 & \cdots & \alpha_{\tau_r-1} \\ \alpha_{\tau_r-1} & \alpha_{\tau_r} & \cdots & \alpha_{\tau_r-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{\tau_r} \end{pmatrix},$$

[3] See Gray et al. (2006) for a review of the properties possessed by circulant matrices.

which is a circulant matrix generated by row vector $\vec{\alpha}_r = (\alpha_1, \alpha_2, \ldots, \alpha_{\tau_r})$, where $\alpha_t = 0$ for any $t > T$. Next, we are going to prove that there exists some r such that Ψ^r has full rank. The associated polynomial of Ψ^r is given by

$$f_r(x) = \alpha_1 + \alpha_2 x + \cdots + \alpha_{\tau_r} x^{\tau_r - 1}.$$

Suppose that for all $r = 1, \ldots, T$ we have Ψ^r is singular, which means there exists a complex τ_r^{th} root of unit, denoted by $\rho_{m_r} = e^{\frac{2\pi m_r}{\tau_r} i}$ for some $m_r \in \{0, 1, \ldots, \tau_r - 1\}$, satisfying $f_r(\rho_{m_r}) = 0$. Notice that $\alpha_t = 0$ for any $t > T$, then we have $f(\rho_{m_r}) = 0$, which means $\{\rho_{m_r}\}_{r=1}^T$ are the roots of $f(x)$. Moreover, we can prove that they are T different roots. Because $0 \leq \frac{2\pi m_r}{\tau_r} < 2\pi$ for each r, the mapping from $\frac{m_r}{\tau_r}$ to ρ_{m_r} is a bijection. Suppose we can find $r < r'$ such that $\rho_{m_r} = \rho_{m_{r'}}$, which means we must have $\frac{m_r}{\tau_r} = \frac{m_{r'}}{\tau_{r'}}$, then we have $m_r \tau_{r'} = m_{r'} \tau_r$. It follows that $\tau_{r'}$ exactly divides $m_{r'} \tau_r$, which is impossible because $\tau_{r'}$ is a prime number and $\tau_{r'} > \tau_r$ and $\tau_{r'} > m_{t'}$. Notice that the degree of polynomial $f(x)$ is equal to $T - 1$, then from the fundamental theorem of algebra we know that $f(x)$ has, counted with multiplicity, exactly $T - 1$ roots in the complex plane, contradicting the fact that we already find T distinct roots. Thus, there must exist some $\hat{r} \in \{1, \ldots, T\}$ such that $\Psi^{\hat{r}}$ is nonsingular.

B.2 Additional Results

B.2.1 Alternative timing

Consider an alternative timing as follows:

1. The principal makes public (Ξ, x), where $\Xi : \Theta \times V \to \Delta(M)$ and $x : V \times M \to \Delta(\mathscr{A})$, to which we assume that the principal can commit during the whole game.

2. Each agent i observes his own private type v_i, and then simultaneously reports $\hat{v}_i$ to the principal.

3. Each agent i privately observes his own signal m_i generated by $\Xi_{\theta,\hat{v}}$, and then simultaneously reports $\hat{m}_i$ to the principal.

4. The principal implements the social alternatives according to $x(\hat{v}, \hat{m})$.

First we consider the following relaxed problem where the signal profile m is directly observed by the principal so that each agent i cannot report $\hat{m}_i \neq m_i$, and participation is mandatory after stage 2.

$$\begin{aligned}
\sup_{\Xi,\, x} \quad & \int_\theta \int_v \int_m u_0\big(x(v,m),v,\theta\big) d\Xi_{\theta,v}(m) dF_V(v) dF_0(\theta) \\
s.t. \quad & \forall i,\; v_i \neq v_i' : \\
& \int_{v_{-i}} \int_\theta \int_m u_i\big(x(v_i,v_{-i},m),v,\theta\big) d\Xi_{\theta,v_i,v_{-i}}(m) dF_0(\theta) dF_{-i}(v_{-i}) \\
& \geq \max\Big\{0, \int_{v_{-i}} \int_\theta \int_m u_i\big(x(v_i',v_{-i},m),v,\theta\big) d\Xi_{\theta,v_i',v_{-i}}(m) dF_0(\theta) dF_{-i}(v_{-i})\Big\}.
\end{aligned}$$

Define $\tilde{x}(\hat{v}_i, v_{-i}, \theta) = \int_m x(\hat{v}_i, v_{-i}, m) d\Xi_{\theta,\hat{v}_i,v_{-i}}(m)$, and the above problem becomes:

$$\begin{aligned}
\sup_{\Xi,\, x} \quad & \int_\theta \int_v u_0\big(\tilde{x}(v,\theta),v,\theta\big) dF_V(v) dF_0(\theta) \\
s.t. \quad & \forall i,\; v_i \neq v_i' : \\
& \int_{v_{-i}} \int_\theta u_i\big(\tilde{x}(v_i,v_{-i},m),v,\theta\big) dF_0(\theta) dF_{-i}(v_{-i}) \\
& \geq \max\Big\{0, \int_{v_{-i}} \int_\theta u_i\big(\tilde{x}(v_i',v_{-i},m),v,\theta\big) dF_0(\theta) dF_{-i}(v_{-i})\Big\},
\end{aligned}$$

whose value is no larger than the value of (P^*). Since the optimal private disclosure mechanism in Theorem 1, 2, 3 already achieves the value of (P^*), we conclude that our result is robust to this alternative timing.

B.2.2 Relationship with Liu (2015)

Following the definition given by Liu (2015), we construct a non-redundant partition model $\langle \Omega, (\Pi_i, P_i)_{i\in I}, g\rangle$, where (i) $\Omega = \Theta$ is the state space, (ii) each agent i's information partition on Ω contains only one element, that is, $\Pi_i = \{\Theta\}$, (iii) agents share a common prior on Ω, that is, $P_i = F_0$ for any $i \in I$, (iv) $g : \Omega \to \{x^*(\cdot,\theta)\}_{\theta\in\Theta}$ specifies the associated direct mechanism the principal will implement at each state, that is, $g(\theta) = x^*(\cdot,\theta)$ for any $\theta \in \Theta$. Moreover, the set of belief hierarchies, denoted by $\delta(\Omega)$, satisfies that each belief hierarchy contains a unique element, because it is common knowledge that all agents share the same prior belief about the distribution of the associated direct mechanisms to be implemented.

As in Liu (2015), a correlating device $\langle (C^\theta)_{\theta\in\Theta}, (q_i^\theta)_{\theta\in\Theta, i\in I}\rangle$ on $\langle \Omega, (\Pi_i, P_i)_{i\in I}, g\rangle$, where each $C^\theta \subseteq \times_{i\in I} C_i$ and each q_i^θ is a probability measure with a support C^θ, satisfies for all i, $\theta, \theta' \in \Theta$: (1) $C_i^\theta = C_i^{\theta'}$, and (2) the marginal distributions of q_i^θ and $q_i^{\theta'}$ over C_i^θ coincide with each other. Then we can prove the following results.

Proposition 5. The optimal information disclosure policy Φ^S is an correlating device. Any aggregately-revealing correlating device on partition model $\langle \Omega, (\Pi_i, P_i)_{i\in I}, g\rangle$ is the solution to (P_1). The reverse is also true if $x^*(\cdot,\theta) \neq x^*(\cdot,\theta')$ for any $\theta \neq \theta'$.

Proof. By definition of Φ^S, for any i and θ_t, each $m_i \in M_i$ will occur with strictly positive probability, since we can always define $m_j := \big(m_i + (j-i)t\big) \mod T$ for all $j \neq i$, and then $\Phi^S(\theta_t, m_i, m_{-i}) > 0$. Thus, condition (1) is satisfied. Notice that

$$q_i^{\theta_t}[m_i] = \frac{\sum_{m_{-i}} \Phi(\theta_t, m_i, m_{-i})}{\sum_m \Phi(\theta_t, m_i, m_{-i})} = \frac{\Phi\big(\theta_t, m_i, m_{-i}^+(\theta_t, m_i)\big)}{\sum_{m_i} \Phi\big(\theta_t, m_i, m_{-i}^+(\theta_t, m_i)\big)} = \frac{\frac{\alpha_t}{T}}{\frac{\alpha_t}{T}\cdot T} = \frac{1}{T},$$

for any θ and m_i, then condition (2) is also satisfied. Thus, Φ^S is a correlating device.

Because correlating device does not change each agent i's belief about the distribution of the state of the world, the interim expected utility for each agent is based on the common prior F_0. Since the correlating device is aggregately revealing, the principal can infer the true state from m, and implement the correct mechanism $x^*\big(\cdot, \theta^+(m)\big)$. Notice that x^* is the solution to (P^*), then agents will participate in the mechanism and truthfully report v. Thus, any aggregately-revealing correlating device is the optimal information disclosure policy that solves (P_1).

Pick any solution (Ξ, x) to (P_1), and we denote the resulting partition model as $\langle \tilde{\Omega}, (\tilde{\Pi}_i, \tilde{P}_i)_{i\in I}, \tilde{g}\rangle$. Since the value of (P^*) is achieved by (Ξ, x), then for any $m \in M^\theta = \{m \in M \mid \Phi(\theta, m) > 0\}$, we have $x(\cdot, m) = x^*(\cdot,\theta)$. Suppose there exists $m' \in M^\theta \cap M^{\theta'}$ for some $\theta \neq \theta'$, then $x^*(\cdot,\theta) = x(\cdot, m') = x^*(\cdot, \theta')$, contradicting the assumption. Thus, $M^\theta \cap M^{\theta'} = \emptyset$ for any $\theta \neq \theta'$, which means the disclosure policy is aggregately revealing. On the other hand, we have proved that each agent's posterior belief about the state has to be coincide with F_0 in the optimal private disclosure mechanism,[4] then the set of belief hierarchies, denoted by $\delta(\tilde{\Omega})$, is equal to $\delta(\Omega)$. Notice that $\langle \Omega, (\Pi_i, P_i)_{i\in I}, g\rangle$ is non-redundant, then by Theorem 1 of Liu (2015), $\langle \tilde{\Omega}, (\tilde{\Pi}_i, \tilde{P}_i)_{i\in I}, \tilde{g}\rangle$ can be uniquely decomposed into a conjunction of a non-redundant model, that is, $\langle \Omega, (\Pi_i, P_i)_{i\in I}, g\rangle$, and a correlating device. We conclude that the information disclosure policy is the unique aggregately-revealing correlating device. □

B.2.3 General payoff environment

We are going to show that the optimality of IUAR disclosure policy does not hinge on the mutual independence of $(F_0, F_1, \ldots, F_N)$. We provide the proof for Theorem 1 where the least

[4] For simplicity, we assume that any individually informative disclosure policy is strictly outperformed by pooling the realizations of private signals for each agent.

restrictions are imposed on the environment. Since principal's information disclosure policy Ξ is chosen before she observes agents' reports about their private types v, the signal profile m and v are mutually independent conditional on the state θ. Let F_V be the joint distribution of v. Then the principal's problem (P) is given by

$$\begin{aligned}&\sup_{\Xi,\,x} \quad \int_\theta \int_m \int_v u_0\big(x(v,m),v,\theta\big)dF_V(v\mid\theta)d\Xi_\theta(m)dF_0(\theta)\\ &s.t. \quad \forall i,(m_i,v_i)\neq(m_i',v_i'): \quad U_i(m_i,v_i;m_i,v_i)\geq \max\{U_i(m_i,v_i;m_i',v_i'),0\},\end{aligned}$$

where

$$\begin{aligned}U_i(m_i,v_i;\hat{m}_i,\hat{v}_i) &= \int_{\theta,m_{-i},v_{-i}} u_i\big(x(\hat{v}_i,v_{-i},\hat{m}_i,m_{-i}),v,\theta\big)d\Psi_{m_i,v_i}(\theta,m_{-i},v_{-i})\\ &= \int_{\theta,m_{-i}}\int_{v_{-i}} u_i\big(x(\hat{v}_i,v_{-i},\hat{m}_i,m_{-i}),v,\theta\big)\underbrace{d\Psi_{\theta,v_i,m}(v_{-i})}_{=dF_{-i}(v_{-i}\mid\theta,v_i)}d\Psi_{m_i,v_i}(\theta,m_{-i}).\end{aligned}$$

And the relaxed problem (P^*) is given by

$$\begin{aligned}&\sup_{x} \quad \int_\theta \int_v u_0\big(x(v,\theta),v,\theta\big)dF_V(v\mid\theta)dF_0(\theta)\\ &s.t. \quad \forall i,v_i\neq v_i': \quad U_i(v_i,v_i)\geq \max\{U_i(v_i,v_i'),0\},\end{aligned}$$

where $U_i(v_i,\hat{v}_i)=\int_\theta\int_{v_{-i}} u_i\big(x(\hat{v}_i,v_{-i},\theta),v,\theta\big)dF_{-i}(v_{-i}\mid\theta,v_i)dF_0(\theta\mid v_i)$.

Let $x^*(v,\theta)$ be the solution to (P^*). Construct private disclosure mechanism (Ξ^W,x^W) in the same way as in Theorem 1, and we will show that it satisfies all the constraints in (P) and achieves the value of (P^*). Notice that

$$\begin{aligned}\Psi^W_{m_i,v_i}(\theta) &= \frac{\Pr(m_i\mid\theta,v_i)F_i(v_i\mid\theta)F_0(\theta)}{\sum_\theta \Pr(m_i\mid\theta,v_i)F_i(v_i\mid\theta)F_0(\theta)} \overset{(1)}{=} \frac{\Pr(m_i\mid\theta)F_i(v_i\mid\theta)F_0(\theta)}{\sum_\theta \Pr(m_i\mid\theta)F_i(v_i\mid\theta)F_0(\theta)}\\ &\overset{(2)}{=} \frac{F_i(v_i\mid\theta)F_0(\theta)}{\sum_\theta F_i(v_i\mid\theta)F_0(\theta)} = \frac{\Pr(v_i,\theta)}{\sum_\theta \Pr(v_i,\theta)} = F_0(\theta\mid v_i),\end{aligned}$$

where (1) comes from the independence between m and v conditional on θ, and (2) is due to the uniform distribution of m_i conditional on θ.[5] This property says when θ is correlated with v, each agent i learns something about the state from his own type v_i; while the signal m_i

[5] Fixed any $\theta_t\in\Theta$, for all $m_i\in M_i$, we have

$$\Pr(m_i\mid\theta_t)=\frac{\Pr(m_i,\theta_t)}{F_0(\theta_t)}\overset{(1)}{=}\frac{\Phi^W(\theta_t,m_i,m^+_{-i}(\theta_t,m_i))}{F_0(\theta_t)}=\frac{\alpha_t/K}{\alpha_t}=\frac{1}{K},$$

where (1) is due to property (ii) of Lemma 2. Thus, m_i is uniformly distributed conditional on θ.

provides no additional information about the state for him. Then we have $U_i(m_i, v_i; \hat{m}_i, \hat{v}_i) =$

$$\begin{aligned}
&\int_{\theta,m_{-i}}\int_{v_{-i}} u_i\big(x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v, \theta\big) dF_{-i}(v_{-i} \mid \theta, v_i) d\Psi^W_{m_i,v_i}(\theta, m_{-i}) \\
=&\int_{\theta}\int_{m_{-i}}\int_{v_{-i}} u_i\big(x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v, \theta\big) dF_{-i}(v_{-i} \mid \theta, v_i) d\Psi^W_{m_i,v_i,\theta}(m_{-i})\Psi^W_{m_i,v_i}(\theta) \\
=&\int_{\theta}\int_{m_{-i}}\int_{v_{-i}} u_i\big(x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}), v, \theta\big) dF_{-i}(v_{-i} \mid \theta, v_i) d\Psi^W_{m_i,\theta}(m_{-i}) dF_0(\theta \mid v_i) \\
=&\int_{\theta}\int_{v_{-i}} u_i\Big(\int_{m_{-i}} x^W(\hat{v}_i, v_{-i}, \hat{m}_i, m_{-i}) d\Psi^W_{m_i,\theta}(m_{-i}), v, \theta\Big) dF_{-i}(v_{-i} \mid \theta, v_i) dF_0(\theta \mid v_i) \\
=&\int_{\theta}\int_{v_{-i}} u_i\big(x^*(\hat{v}_i, v_{-i}, \theta), v, \theta\big) dF_{-i}(v_{-i} \mid \theta, v_i) dF_0(\theta \mid v_i),
\end{aligned}$$

where the last equality is due to Lemma 11 which guarantees that the principal can infer the true state even when there is unilateral deviation from truthfully reporting one's signal m_i. Thus, (Ξ^W, x^W) satisfies all the constraints of (P) because $x^*(v, \theta)$ is incentive compatible and individual rational. Notice that

$$\begin{aligned}
&\int_{\theta}\int_{m}\int_{v} u_0\big(x^W(v,m), v, \theta\big) dF_V(v \mid \theta) d\Xi^W_\theta(m) dF_0(\theta) \\
=&\int_{\theta}\int_{v} u_0\Big(\int_{m} x^W(v,m) d\Xi^W_\theta(m), v, \theta\Big) dF_V(v \mid \theta) dF_0(\theta) \\
=&\int_{\theta}\int_{v} u_0\big(x^*(v,\theta), v, \theta\big) dF_V(v \mid \theta) dF_0(\theta),
\end{aligned}$$

and then we conclude that (Ξ^W, x^W) constitutes a solution to (P).

B.2.4 Identify the truth teller

Assume $T = 2$ and $N = 3$, choose a prime number $K \geq 3$, and construct information disclosure policy Φ^W as in Theorem 1. Then we have for $t = 1, 2$, $\Xi^W_{\theta_t}(m) > 0$ if and only if $m_j \equiv m_i + (j - i)t \pmod{K}$, $\forall i, j \in \{1, 2, 3\}$. Suppose that there exists an off-path signal profile induced by unilateral deviation, but the principal cannot identify the truth teller. More precisely, $\exists m \in M$ such that $\Xi^W_{\theta_t}(m) > 0$ for some t, and $\exists m'_i \neq m_i$ for some i, satisfying: (i) $\exists m'_j \neq m_j$ for some $j \neq i$ such that $\Xi^W_{\theta_{t_1}}(m'_i, m'_j, m_k) > 0$ for some t_1, and (ii) $\exists m'_k \neq m_k$ for some $k \neq i, j$ such that $\Xi^W_{\theta_{t_2}}(m'_i, m_j, m'_k) > 0$ for some t_2, where $t, t_1, t_2 \in \{1, 2\}$. Without loss of generality, let $i = 1$, $j = 2$ and $k = 3$, then we have:

$$\begin{cases}
m_2 \equiv m_1 + t \pmod{K}, & m_3 \equiv m_1 + 2t \pmod{K} \\
m'_2 \equiv m'_1 + t_1 \pmod{K}, & m_3 \equiv m'_1 + 2t_1 \pmod{K} \\
m_2 \equiv m'_1 + t_2 \pmod{K}, & m'_3 \equiv m'_1 + 2t_2 \pmod{K}.
\end{cases}$$

It follows that $m_1 - m_1' \equiv t_2 - t \pmod K$ and $m_1 - m_1' \equiv 2t_1 - 2t \pmod K$, which means $t_1 - t \equiv t_2 - t_1 \pmod K$. Because two of $\{t, t_1, t_2\}$ must be equal to each other, then we have $t = t_1 = t_2$, contradicting $m_i' \neq m_i$. Thus, either condition (i) does not hold which means agent j is the truth teller, or condition (ii) does not hold which means agent k is the truth teller.

B.2.5 An alternative proof of Theorem 3

We provide an alternative way to prove Theorem 3. Instead of introducing additional signals to get the full-rank belief matrix, we can approximate any singular belief matrix by a sequence of invertible belief matrices. Specifically, consider the circulant matrix Ψ^S generated by the row vector $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_T)$. Without loss of generality we can assume that $\alpha_1 > 0$; otherwise we consider the circulant matrix $\Psi^{S'}$ generated by row vector $(\alpha_j, \alpha_{j+1}, \ldots, \alpha_{j-1})$ with $\alpha_j > 0$, which is a cyclic permutation of $\vec{\alpha}$. One can easily check that Ψ^S has full rank if and only if $\Psi^{S'}$ has full rank. If Ψ^S is singular, then define $R := \{m \in \{1, \ldots, T\} \mid f(\rho_m) = 0\}$ and $R^C := \{1, \ldots, T\} \setminus R$. Let $\bar{\varepsilon} = \min_{m \in R^C} |f(\rho_m)|$, and set $\bar{\varepsilon} = +\infty$ if $R^C = \emptyset$. We consider a sequence of belief matrices $\{\Psi^n\}_{n=n_0}^{\infty}$, where the positive integer n_0 satisfies $\frac{T}{n_0} < \min\{\bar{\varepsilon}, \alpha_1\}$, and each Ψ^n is a circulant matrix generated by row vector $\vec{\alpha}^n = (\alpha_1 - \frac{T-1}{n}, \alpha_2 + \frac{1}{n}, \ldots, \alpha_T + \frac{1}{n})$. Then for any $m = 1, \ldots, T$, we have

$$
\begin{aligned}
f^n(\rho_m) &= \big(\alpha_1 - \frac{T-1}{n}\big) + \big(\alpha_2 + \frac{1}{n}\big)\rho_m + \cdots + \big(\alpha_T + \frac{1}{n}\big)\rho_m^{T-1} \\
&= f(\rho_m) + \frac{1}{n}\big(1 + \rho_m + \cdots + \rho_m^{T-1}\big) - \frac{T}{n} \\
&= f(\rho_m) + \frac{1 - \rho_m^T}{1 - \rho} - \frac{T}{n} \\
&= f(\rho_m) - \frac{T}{n}, \quad \text{(because } \rho_m^T = e^{\frac{2\pi m}{T} i \cdot T} = 1\text{)}
\end{aligned}
$$

where $f^n(x)$ is the associated polynomial of circulant matrix Ψ^n. If $m \in R$, then $f^n(\rho_m) = 0 - \frac{T}{n} \neq 0$. If $m \in R^C$, then we have either $f^n(\rho_m) > \bar{\varepsilon} - \frac{T}{n} > 0$ or $f^n(\rho_m) < -\bar{\varepsilon} - \frac{T}{n} < 0$. Thus, we prove that $\lim_{n\to\infty} \Psi^n = \Psi^S$ and Ψ^n is nonsingular for each n. [6]

We start with (Φ^S, x^S), the optimal private disclosure policy with strong control over the disclosure process, and then for each common prior over Θ, denoted by $\vec{\alpha}^n$, we construct the private disclosure policy (Φ^n, x^S) with the corresponding full-rank belief matrix Ψ^n about the

[6] Basically, any singular matrix is the limit of a sequence of nonsingular matrices, which means in the topological space of circulant matrices generated by all $\vec{\alpha} \in \mathbb{R}^T$, every non-empty open subset contains at least one nonsingular circulant matrix. Thus, the subset of nonsingular circulant matrices is a dense set. On the other hand, any singular circulant matrix must satisfy $f(\rho_m) = 0$ for some $m = 0, \ldots, T-1$, which will reduce the degrees of freedom, and thus it is a *generic* property for any $\vec{\alpha} \in \mathbb{R}^T$ to generate a nonsingular circulant matrix.

opponents' signals. Let $\beta_i^{m_i \to m_i'|n}$ be the maximum gain from misreporting $m_i' \neq m_i$ when agent i observes m_i, which is given by

$$\beta_i^{m_i \to m_i'|n} = \max_{(v_i, v_i') \in V_i^2} \Big\{ -\int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^S(v_i, v_{-i}, m_i, m_{-i}), v, \theta\big) d\Psi^n_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}) \\ + \int_{v_{-i}} \int_{\theta, m_{-i}} \tilde{u}_i\big(x^S(v_i', v_{-i}, m_i', m_{-i}), v, \theta\big) d\Psi^n_{m_i}(\theta, m_{-i}) dF_{-i}(v_{-i}), 0 \Big\},$$

Set $\beta_i^{m_i \to m_i|n}$ equal to 0. Define a $T \times T$ matrix $\mathbf{B}_i^n = \big(\beta_i^{r \to s|n}\big)_{(r,s) \in \{1,\ldots,T\}^2}$. Let $\vec{t}_i^n(m_i) = \big(t_i^n(m_i, m_{-i})\big)_{m_{-i} \in M_{-i}}$ be the column vector representing agent i's transfer by reporting m_i. Since Ψ^n is invertible, we can find $\mathbf{t}_i^n := \big(\vec{t}_i^n(m_i)\big)_{m_i \in M_i}$ solving $\Psi^n \mathbf{t}_i^n = \mathbf{B}_i^n$. Then $(\Phi^n, x^S, \mathbf{t}^n)$ constitutes a candidate private disclosure policy with transfer. Let $(\Phi^n_{opt}, \tilde{x}^n_{opt}, t^n_{opt})$ be the optimal private disclosure policy with transfer when the prior over Θ is $\vec{\alpha}^n$, which achieves the ex ante expected utility $\tilde{U}_0^{opt}(\vec{\alpha}^n)$. Denote $\tilde{U}_0^S$ as the principal's ex ante expected utility if she has strong control. Hence we have

$$\begin{aligned}
&\tilde{U}_0^{opt}(\vec{\alpha}) - \tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha}) = \lim_{n\to\infty} \big(\tilde{U}_0(\Phi^n_{opt}, \tilde{x}^n_{opt}, t^n_{opt} \mid \vec{\alpha}^n) - \tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha})\big) \\
&\geq \lim_{n\to\infty} \big(\tilde{U}_0(\Phi^n, x^S, \mathbf{t}^n \mid \vec{\alpha}^n) - \tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha})\big) \\
&= \lim_{n\to\infty} \Big(\underbrace{\tilde{U}_0(\Phi^n, x^S, \mathbf{t}^n \mid \vec{\alpha}^n) - \tilde{U}_0^S(\Phi^n, x^S \mid \vec{\alpha}^n)}_{=0, \text{ by definition of } \mathbf{t}^n} + \tilde{U}_0^S(\Phi^n, x^S \mid \vec{\alpha}^n) - \tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha})\Big) \\
&= \lim_{n\to\infty} \int_v \sum_{\theta \in \Theta} \Big(\underbrace{(\alpha^n_\theta - \alpha_\theta)}_{\to 0} \cdot \underbrace{\tilde{u}_0\big(\tilde{x}^*(v,\theta), v, \theta\big)}_{\text{bounded}}\Big) dF_V(v) = 0.
\end{aligned}$$

Notice that $\tilde{U}_0^{opt}(\vec{\alpha})$ is bounded from above by the maximum expected utility in (P_1), which is $\tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha})$, then we conclude that $\tilde{U}_0^{opt}(\vec{\alpha}) = \tilde{U}_0^S(\Phi^S, x^S \mid \vec{\alpha})$.

B.2.6 Alternative implementation of the IUAR disclosure policy

In Section 1.5 we build the IUAR disclosure policy through the sample-product approach. Next, we provide an alternative way of implementation, that is, restrictions on agents' access to certain information. We illustrate this idea in a real-world example introduced by Dow and Gorton (1993).

Cray Research is a leading manufacturer of supercomputers. Its chief scientist engaged in an important research project is considering either staying in the company, or leaving, together with the project, to found an independent competitor. The value of Cray Research, which is either high (in a good state θ_G) or low (in a bad state θ_B), is determined as follows. If the scientist stays and the project has a high chance of success, the company would gain a strong competitive advantage, and thus have a high value. If the scientist leaves and the project is likely to fail, that is also good news for the company's value because it's likely to maintain a dominant position. However, it would be bad news if either the leaving scientist is likely to succeed in the project, resulting in a tough competitor; or the unpromising project continues to waste money.

Imagine that the Board of Cray Research decides whether to keep (a_1) the company or to sell (a_2) it. Table B.1 defines the Board's payoff $u_0(a,\theta)$ for each action-state pair. Basically we are assuming that the Board wants to match the state, that is, the optimal decision a^* is given by $a^*(\theta_G) = a_1$ and $a^*(\theta_B) = a_2$. Assume that the common prior over $\{\theta_G, \theta_B\}$ is given by $F_0(\theta_G) = F_0(\theta_B) = \frac{1}{2}$. The Board cannot observe the state; instead, she relies on Managers to collect information about the state, so as to implement the optimal outcome. This is possibly because Managers are better informed of their own departments' conditions than the Board, as is usually assumed in the organizational communication literature (e.g., Alonso, Dessein, and Matouschek, 2008); or because the Board's cost to acquire concrete information is too high. On the other hand, the Board can control what information and how precise each manager can learn, by authorizing access to particular database for each manager (e.g., through the Identity and Access Management Systems). Assume that there are two managers, each with payoff function $u_i(a,\theta)$ given by Table B.2.

Table B.1: Board's payoff

$u_0(a,\theta)$	θ_G	θ_B
a_1	1	0
a_2	0	1

Table B.2: Manager i's payoff

$u_i(a,\theta)$	θ_G	θ_B
a_1	2	0
a_2	0	-1

We can see that both managers prefer a_1 to a_2 regardless of the state. Intuitively, changing ownership of the company usually changes the management style and undermines the interests of current managers. Moreover, when the Board sells the company in a bad state, the current managers would most likely be fired by the new owner, which would damage their reputation and end up with a negative payoff (-1). We assume that managers can resign voluntarily and get their reservation utility 0 *before* the Board takes actions. The resignation of any manager would cause turmoil within the current management, and give 0 payoff to the Board regardless of her action.

Assume that the Board can implement any lottery over $\{a_1, a_2\}$, and let $x \in [0,1]$ be the probability of playing a_1. The timing is: (i) the Board designs (and commits to) an information disclosure policy and associated allocation rule; (ii) each agent i observes a private signal m_i

(generated by the disclosure policy), and then decides whether to report a message to Principal, or to resign at this stage; (iii) the Board keeps the company with probability x, and sells it with probability $(1-x)$ according to the allocation rule and agents' reports. Thus, the Board's problem is to maximize her expected payoff, such that each manager finds it optimal to truthfully report what he observes, and gets nonnegative interim expected utility.

Immediately, the first-best outcome for the Board is $x^*(\theta_G) = 1$ and $x^*(\theta_B) = 0$. By Proposition 1, it is optimal to employ the IUAR disclosure policy: $M_i = \{0,1\}$, and the joint distribution over $\{\theta_G, \theta_B\} \times M_1 \times M_2$, denoted by Φ, is given by Table B.3. Meanwhile, the allocation rule is given by $x(m) = x^*\left(\theta^+(m)\right)$, where $\theta^+(m)$ is the state inferred from the signal profile m. We can check that the allocation rule is Bayesian incentive compatible and guarantees nonnegative expected payoff for both managers.

Table B.3: Information disclosure policy

$\Phi(\theta_G, m)$	$m_2 = 0$	$m_2 = 1$
$m_1 = 0$	$\frac{1}{4}$	0
$m_1 = 1$	0	$\frac{1}{4}$

$\Phi(\theta_B, m)$	$m_2 = 0$	$m_2 = 1$
$m_1 = 0$	0	$\frac{1}{4}$
$m_1 = 1$	$\frac{1}{4}$	0

Recall that the two main factors (called the sub-states, denoted by s_1 and s_2) that determine the company's value are (i) whether the chief scientist stays ($s_1 = 0$) or leaves ($s_1 = 1$), and (ii) whether the project has a high chance of success ($s_2 = 0$) or not ($s_2 = 1$). Let $\alpha(s_1, s_2) \in \Delta(\{0,1\}^2)$ be the (full-support) common prior of sub-states, and $h : \{0,1\}^2 \to \{\theta_G, \theta_B\}$ be the mapping from sub-states profiles to states. From the previous illustration, we have that $h(0,0) = h(1,1) = \theta_G$ and $h(0,1) = h(1,0) = \theta_B$.[7] We can see that s_1 and s_2 are complements in Börgers, Hernando-Veciana, and Krähmer (2013), since neither of the sub-state can determine the state on itself.

The IUAR disclosure policy in Table B.3 can be constructed as follows. Manager 1 (or 2) is the manager of Human Resource department (or R&D department), who is authorized to collect information only about s_1 (or s_2). Let ε be a random variable which is independent of (s_1, s_2) and is drawn from $\{0,1\}$ equally likely, then we define manager i's signal by $m_i = s_i + \varepsilon \mod 2$, for $i = 1,2$. We can prove that (m_1, m_2) follows exactly the same distribution as in

[7]Obviously, $\alpha(s_1, s_2)$ satisfies $\alpha(0,0) + \alpha(1,1) = F_0(\theta_G) = \frac{1}{2}$ and $\alpha(0,1) + \alpha(1,0) = F_0(\theta_B) = \frac{1}{2}$.

Table B.3. To see this, for example,

$$\begin{aligned}
\Pr(\theta_G \mid m_1 = 0) &= \frac{\Pr\big(s_1 + \varepsilon \equiv 0 \ (\mathrm{mod}\ 2); (s_1, s_2) = (0,0) \text{ or } (1,1)\big)}{\Pr\big(s_1 + \varepsilon \equiv 0 \ (\mathrm{mod}\ 2)\big)} \\
&= \frac{\Pr\big((\varepsilon, s_1, s_2) = (0,0,0) \text{ or } (1,1,1)\big)}{\Pr\big((\varepsilon, s_1) = (0,0) \text{ or } (1,1)\big)} = \frac{\frac{1}{2}\alpha(0,0) + \frac{1}{2}\alpha(1,1)}{\frac{1}{2}(\alpha(0,0)+\alpha(0,1)) + \frac{1}{2}(\alpha(1,0)+\alpha(1,1))} \\
&= \alpha(0,0) + \alpha(1,1) = \Pr(\theta_G) = \frac{1}{2},
\end{aligned}$$

which implies the individually uninformative property. On the other hand, the mapping h can also be written as

$$h(s_1, s_2) = \begin{cases} \theta_G, & \text{if } s_1 - s_2 \equiv 0 \ (\mathrm{mod}\ 2) \\ \theta_B, & \text{if } s_1 - s_2 \equiv 1 \ (\mathrm{mod}\ 2). \end{cases}$$

From $m_1 - m_2 \equiv s_1 - s_2$ (mod 2), we get the aggregately revealing property.

In this example, we exploit the complementary nature of the two sub-states and the randomizing device to construct the IUAR disclosure policy. The above result suggests that in order to maximize the Board's interests, each department manager should only be authorized access to information concerning his own department, and certain noises should be deliberately added to what managers can observe. This result provides a novel rationale for sending noisy signals and preventing side-information among departments in organizational communications.

B.2.7 Extensions to the case with continuously distributed states

In Section 1.6.2 we have extended Proposition 1 to the case with continuously distributed states, and have derived the optimal private disclosure mechanism (Φ^S, x^S) when the principal has strong control over the disclosure process. Next, we prove the continuous version of Theorem 1, 2, 3. If the principal only has weak control, due to the same correlated structure of Φ^S as in the discrete case, the principal can apply the techniques in Section 1.4.2 to elicit truthful reports about the signal profiles from agents for free when there are three or more agents.

In the case where there are only two agents, the principal can still achieve the value of (P^*) through the approximation argument. Roughly speaking, in the vector space $\Delta(\Theta)$ endowed with the total variation norm, we choose a sequence of discrete distribution functions $\{F_0^n\}_{n=1}^{\infty}$ satisfying that (1) the sequence converges to the continuous distribution F_0, and (2) each discrete distribution induces a full-rank circulant belief matrix Ψ^n.[8] For each F_0^n, let $\Theta_n = \{\theta_1, \theta_2, \ldots, \theta_{K_n}\}$ be the states that occurs with strictly positive probabilities. Define $(\Phi_n^S, x_n^S, \mathbf{t}^n)$ as the private disclosure policy with transfer which implements the restriction of $x^*(v,\theta)$ to Θ_n. Then $(\Phi_n^S, x_n^S, \mathbf{t}^n)$ constitutes a candidate solution to the principal's problem with weak control when the common prior over the state of the world is F_0^n, and let $(\Phi_{opt}^n, x_{opt}^n, \mathbf{t}_{opt}^n)$ be the corresponding optimal private disclosure policy. Then the principal's maximum expected payoff by private disclosure with weak control under the prior F_0, denoted as $U_0^{opt}(F_0)$, satisfies:

$$\begin{aligned}
&U_0^{opt}(F_0) - U_0^S(\Phi^S, x^S \mid F_0) = \lim_{n\to\infty} \big(U_0(\Phi_{opt}^n, x_{opt}^n, t_{opt}^n \mid F_0^n) - U_0^S(\Phi^S, x^S \mid F_0)\big) \\
&\geq \lim_{n\to\infty} \big(U_0(\Phi_n^S, x_n^S, \mathbf{t}^n \mid F_0^n) - U_0^S(\Phi^S, x^S \mid F_0)\big) \\
&= \lim_{n\to\infty} \Big(\underbrace{U_0(\Phi_n^S, x_n^S, \mathbf{t}^n \mid F_0^n) - U_0^S(\Phi_n^S, x_n^S \mid F_0^n)}_{=0,\text{ by definition of } \mathbf{t}^n} + U_0^S(\Phi_n^S, x_n^S \mid F_0^n) - U_0^S(\Phi^S, x^S \mid F_0)\Big) \\
&= \lim_{n\to\infty} \int_v \int_{\theta\in\Theta} \underbrace{u_0\big(x^*(v,\theta), v, \theta\big)}_{\text{bounded}} d\big(F_0^n(\theta) - F_0(\theta)\big) dF_V(v).
\end{aligned}$$

Notice that $|u_0\big(x^*(v,\theta), v, \theta\big)| < H$, then we have

$$\left| \int_v \int_{\theta\in\Theta} u_0\big(x^*(v,\theta), v, \theta\big) d\big(F_0^n(\theta) - F_0(\theta)\big) dF_V(v) \right| < H \cdot \int_{\theta\in\Theta} \left| d\big(F_0^n(\theta) - F_0(\theta)\big) \right|,$$

where the right hand side goes to 0 as $n \to \infty$, since $\{F_0^n\}_{n=1}^\infty$ converges to F_0 under the total variation norm. Thus, we have $U_0^{opt}(F_0) \geq U_0^S(\Phi^S, x^S \mid F_0)$. Notice that $U_0^S(\Phi^S, x^S \mid F_0)$ is the upper bound of $U_0^{opt}(F_0)$, then we conclude that $U_0^{opt}(F_0) = U_0^S(\Phi^S, x^S \mid F_0)$.

B.2.8 Robustness to faulty agents

Recall that in the optimal private disclosure mechanism given by Theorem 1, any two truthful reports of signals uniquely pin down the state. Then there could be enough redundancy in cross checking the existence of unilateral deviations, so that the principal is still able to infer the true state even if there are some "faulty" agents who do not act according to incentives. Eliaz (2002) defines the k-fault tolerant Nash Equilibrium for an n-player game, where each non-faulty player has no incentive to deviate from the equilibrium action, regardless of the identity and actions of the faulty players, as long as there are $(n-k-1)$ other non-faulty

[8] We can always find such sequences because the subset of nonsingular circulant matrices generated by discrete distribution functions is a dense set. See Footnote 6 in this Supplemental Material (Appendix B).

players who stick to the equilibrium action. In the same spirit, we check whether the IUAR disclosure mechanism is k-fault tolerant.

We keep the setup in Section 1.3. For simplicity, we only consider the case with four or more agents, whose private information only consists of the signals generated by the disclosure policy. We construct the IUAR disclosure policy in the same way as in Section 1.6.3, denoted by Ξ^D, except that the state is determined by $D \in \{2,\ldots,N-2\}$ signals rather than $(N-2)$ signals. Define $M_i = \{1,\ldots,K\}$ for all $i \in I$, where $K \geq \max\{T,N\}$ is a prime number. Let $\varepsilon_1,\ldots,\varepsilon_{D-1} \in \{1,\ldots,K\}$ be mutually independent random variables, satisfying $\Pr(\varepsilon_\tau = k) = \frac{1}{K}$ for $\tau = 1,\ldots,D-1$ and $k = 1,\ldots,K$. Moreover, $(\varepsilon_1,\ldots,\varepsilon_{D-1})$ is independent of θ. The signal profile $m = (m_1,\ldots,m_N)$ is defined as follows:

$$\begin{pmatrix} m_1 \\ \vdots \\ m_{D-1} \\ m_D \\ \vdots \\ m_N \end{pmatrix} = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 1^0 & 2^0 & \cdots & D^0 \\ \vdots & \vdots & \ddots & \vdots \\ 1^{N-D} & 2^{N-D} & \cdots & D^{N-D} \end{pmatrix} \cdot \begin{pmatrix} \theta \\ \varepsilon_1 \\ \vdots \\ \varepsilon_{D-1} \end{pmatrix} \mod K,$$

where each signal m_i is equal to the *residue* of the dot product of i-th row vector (denoted as $\boldsymbol{\zeta}_i$) and column vector $(\theta,\varepsilon_1,\ldots,\varepsilon_{D-1})$ (denoted as $\boldsymbol{\varepsilon}$), modulo K.[9] Let $m_{\tilde{I}} := (m_i)_{i\in\tilde{I}}$ stand for the restriction of report profile m to agents included in the subset $\tilde{I}$. Using a similar argument with Lemma 6, we have that Ξ^D satisfies:

(i) For any $\tilde{I} \subseteq I$ such that $|\tilde{I}| \leq D-1$, the posterior belief about θ conditional on observing truthful reports of $m_{\tilde{I}}$, denoted by $\Psi^{IS}_{m_{\tilde{I}}}(\theta)$, coincides with F_0.

(ii) Any D signals uniquely pin down the realization of $\theta = \theta^+(m_{\tilde{I}})$ such that $|\tilde{I}| = D$, as well as $(\varepsilon_1,\ldots,\varepsilon_{D-1})$ and the remaining $(N-D)$ signals.

(iii) Any report of $(D+2)$ or more signals reveals whether there is unilateral deviation from truth-telling, and, if so, the identity of the agent who misreports.

We say a disclosure policy is k-fault tolerant if unilateral deviation from truthfully reporting one's signal by any non-faulty agent does not affect the principal's ability to infer the true state, regardless of the identity and reports of the faulty agents, as long as there are $(N-k-1)$

[9] The residue is calculated in the way of standard modular arithmetic, except that when the dot product can be exactly divided by K, we write $m_i = K$ instead of $m_i = 0$.

other non-faulty agents who stick to truth telling. More formally, for any $m \in M$ and $\theta \in \Theta$, define the largest subset of agents whose reports are consistent with the state θ as

$$I^{+}(m,\theta) := \bigcup_{I' \in I} \left\{ I' \mid \theta^{+}(m_{I'}) = \theta, \text{ and } |I'| = D \right\}.$$

Then the k-fault tolerant property requires that: fixed any $\theta \in \Theta$ and $m \in M$ such that $\Xi_\theta(m) > 0$, for any i, $\hat{m}_i \in M_i$, $(\hat{m}_j)_{j\in\tilde{I}} \in \prod_{j\in\tilde{I}} M_j$ and any $\tilde{I} \in I$ such that $|\tilde{I}| \leq k$, there exists no $\theta' \neq \theta$ satisfying

$$\left| I^{+}\left(\left(\hat{m}_i, (\hat{m}_j)_{j\in\tilde{I}}, (m_s)_{i\neq s\in I\setminus\tilde{I}}\right), \theta \right) \right| \leq \left| I^{+}\left(\left(\hat{m}_i, (\hat{m}_j)_{j\in\tilde{I}}, (m_s)_{i\neq s\in I\setminus\tilde{I}}\right), \theta' \right) \right|.$$

We derive the condition under which Ξ^D is k-fault tolerant. In the worst scenario, a non-faulty agent and k faulty agents jointly choose their reports to maximize the number of reports which are consistent with a false state; moreover, this subset of reports can include up to $(D-1)$ non-faulty agents' truthful reports, due to the property (i) of Ξ^D. Thus, for any $\theta' \neq \theta$, we have

$$\max_{\hat{m}_i, (\hat{m}_j)_{j\in\tilde{I}}} \left| I^{+}\left(\left(\hat{m}_i, (\hat{m}_j)_{j\in\tilde{I}}, (m_s)_{i\neq s\in I\setminus\tilde{I}}\right), \theta' \right) \right| = 1 + k + (D-1) = k + D.$$

Meanwhile, the number of reports that are consistent with the true state reaches its minimum value, which equals $(N-k-1)$. If the principal can infer the true state even in this worst scenario, Ξ^D is k-fault tolerant. Then, we get the condition:

$$N - k - 1 > k + D \quad \Longleftrightarrow \quad 2k \leq N - D - 2.$$

Thus, we conclude that as long as there are at most k faulty agents such that $2k \leq N - D - 2$, the principal can always infer the true state through Ξ^D, regardless of who are the faulty agents and which non-faulty agent unilaterally deviates.

Next, we define the allocation rule x^D:

$$x^D(m) = x^*(\theta), \text{ where } \theta = \arg\max_{\theta'} \left| I^{+}(m,\theta') \right|.$$

Since unilateral deviation from telling the truth does not affect the inferred state (and thus does not change the outcome), truth telling strategy profile constitutes a k-fault tolerant Bayesian Nash Equilibrium. In case $D = 1$ where Ξ^D fully reveals the state to all agents, we go back to the complete information setup, and the condition for k-fault tolerance becomes $k < \frac{N}{2} - 1$, as in Proposition 1 of Eliaz (2002).

Remark. Clearly, the more signals we need to determine the state, the less redundancy there is in cross checking, thus the less robust (Ξ^D, x^D) is to faulty agents. Let N_F denote the number of faulty agents that are allowed in implementing (Ξ^D, x^D). On the other hand, from Section 6.3, if the state is uniquely pinned down by D signals, then the optimal private disclosure mechanism is immune to information sharing among at most $(D-1)$ agents. Let N_{IS} denote the number of agents that are allowed to communicate about their private signals. Then we have

$$\left.\begin{array}{l} 2N_F \leq N-D-2 \\ N_{IS} \leq D-1 \end{array}\right\} \implies 2N_F + N_{IS} \leq N-3.$$

Thus, given a fixed number of agents, the principal faces a tradeoff between these two concepts of robustness.

B.3 Numerical Examples

B.3.1 A counterexample example - 3 agents

Assume $\Theta = \{\theta_1, \theta_2\}$ with $F_0(\theta_1) = F_0(\theta_2) = \frac{1}{2}$. Assume $\mathscr{A} = \{a_1, a_2\}$, and a feasible allocation is given by $xa_1 + (1-x)a_2$, where $x \in [0,1]$ is the probability according to which a_1 is implemented. Agents don't have private types. Table B.4 defines the principal's payoff function $u_0(a,\theta)$ and agent i's payoff function $u_i(a,\theta)$, for $i = 1,2,3$. We can easily check that Assumption 1 is not satisfied, since for any agent, a_1 is the worst social alternative at θ_1; while a_2 is the worst one at θ_2.

Table B.4: Payoff environment

$u_0(a,\theta)$	θ_1	θ_2
a_1	1	0
a_2	0	1

$u_i(a,\theta)$	θ_1	θ_2
a_1	$\frac{1}{2}$	1
a_2	1	$-\frac{1}{2}$

We first write the relaxed problem (P^*):

$$\begin{aligned} \max_x \quad & \frac{1}{2}\big(x(\theta_1)\cdot 1 + (1-x(\theta_1))\cdot 0\big) + \frac{1}{2}\big(x(\theta_2)\cdot 0 + (1-x(\theta_2))\cdot 1\big) \\ s.t. \quad & \frac{1}{2}\big(x(\theta_1)\cdot \frac{1}{2} + (1-x(\theta_1))\cdot 1\big) + \frac{1}{2}\big(x(\theta_2)\cdot 1 + (1-x(\theta_2))\cdot(-\frac{1}{2})\big) \geq 0. \end{aligned}$$

Immediately, the solution to (P^*) is given by $x^*(\theta_1) = 1$ and $x^*(\theta_2) = 0$.

The optimal private disclosure mechanism solves

$$(P) \quad \max_{\Xi,x} \quad \frac{1}{2}\cdot\int_m x(m)d\Xi_{\theta_1}(m)\cdot 1+\frac{1}{2}\cdot\int_m \big(1-x(m)\big)d\Xi_{\theta_2}(m)\cdot 1$$
$$s.t. \quad \forall i,\ \forall m_i \neq m_i': \quad U_i(m_i,m_i)\geq \max\{U_i(m_i,m_i'),0\}$$

where

$$U_i(m_i,\hat{m}_i) =\Psi_{m_i}(\theta_1)\int_{m_{-i}}\Big(x(\hat{m}_i,m_{-i})\big)\cdot\frac{1}{2}+\big(1-x(\hat{m}_i,m_{-i})\big)\cdot 1\Big)d\Psi_{m_i,\theta_1}(m_{-i})$$
$$+\Psi_{m_i}(\theta_2)\int_{m_{-i}}\Big(x(\hat{m}_i,m_{-i})\big)\cdot 1+\big(1-x(\hat{m}_i,m_{-i})\big)(-\frac{1}{2})\Big)d\Psi_{m_i,\theta_2}(m_{-i}).$$

Suppose the solution to (P), denoted by (Ξ,x), implements $x^*(\theta)$, then for any θ and any $m\in M$ such that $\Xi_\theta(m)>0$, we have $x(m)=x^*(\theta)$. Suppose there exists $m\in M$ such that $\Xi_{\theta_1}(m)>0$ and $\Xi_{\theta_2}(m)>0$, then $x(m)=x^*(\theta_1)=1\neq 0=x^*(\theta_2)=x(m)$. Thus, for all $m\in M$, $\Xi_\theta(m)>0$ for at most one θ; or in other words, Ξ is aggregately revealing. Then we have $U_i(m_i,\hat{m}_i=m_i)=\Psi_{m_i}(\theta_1)\cdot\frac{1}{2}+\Psi_{m_i}(\theta_2)\cdot(-\frac{1}{2})\geq 0$, which means $\Psi_{m_i}(\theta_1)\geq\frac{1}{2}$. One the other hand, we have $\int_{m_i}\Psi_{m_i}(\theta_1)d\Lambda_i(m_i)=F_0(\theta_1)=\frac{1}{2}$, since F_0 is the common prior shared by all. Thus, we must have $\Psi_{m_i}(\theta_1)=\frac{1}{2}$, $\forall m_i\in M_{-i}$. Notice that $U_i(m_i,\hat{m}_i)$ is minimized at $\hat{m}_i=m_i$ regardless of how we define the allocation rule off the equilibrium path, then fixed any θ, and any m such that $\Xi_\theta(m)>0$, we must have $x(\hat{m}_i,m_{-i})=x(m_i,m_{-i})$, for any $\hat{m}_i\neq m_i$.

Now pick any $m_1\in M_1$, since $\Psi_{m_1}(\theta_1)=\frac{1}{2}$, then there exist m_{-1} and m_{-1}' such that $\Xi_{\theta_1}(m_1,m_{-1})>0$ and $\Xi_{\theta_2}(m_1,m_{-1}')>0$. Then we have $x(m_1,m_2',m_3)=x(m_1,m_2,m_3)=x^*(\theta_1)=1$, and $x(m_1,m_2',m_3)=x(m_1,m_2',m_3')=x^*(\theta_2)=0$, which is infeasible. Thus, the solution to (P) cannot achieve the value of its relaxed problem (P^*).

B.3.2 A counterexample example - 2 agents

Assume $\Theta=\{\theta_1,\theta_2\}$ with $F_0(\theta_1)=F_0(\theta_2)=\frac{1}{2}$. Assume $\mathscr{A}=\{a_1,a_2,\underline{a}\}$, and a feasible allocation is given by $x=(x_1,x_2,x_3)\in\Delta(\mathscr{A})$. Agents don't have private types. Table B.5 gives the principal's payoff function $u_0(a,\theta)$ and agent i's payoff function $u_i(a,\theta)$, for $i=1,2$. Clearly, $\underline{a}$ is the uniformly worst social alternative.

We can show that the solution to relaxed problem (P^*) is given by $x^*(\theta_1)=(1,0,0)$ and $x^*(\theta_2)=(0,1,0)$. Suppose the optimal private disclosure mechanism, denoted by (Ξ,x), implements $x^*(\theta)$, then through similar arguments as in Appendix B.3.1, we have $\Psi_{m_i}(\theta_1)=\frac{1}{2}$,

Table B.5: Payoff environment

$u_0(a,\theta)$	θ_1	θ_2
a_1	1	0
a_2	0	1
$\underline{a}$	0	0

$u_i(a,\theta)$	θ_1	θ_2
a_1	$\frac{1}{2}$	1
a_2	1	$-\frac{1}{2}$
$\underline{a}$	$\frac{1}{2}$	$-\frac{1}{2}$

$\forall m_i \in M_{-i}$. Notice that $U_i(m_i, \hat{m}_i) =$

$$\Psi_{m_i}(\theta_1)\int_{m_{-i}}\Big(x_1(\hat{m}_i,m_{-i}))\cdot\frac{1}{2}+x_2(\hat{m}_i,m_{-i})\cdot 1+x_3(\hat{m}_i,m_{-i})\cdot\frac{1}{2}\Big)d\Psi_{m_i,\theta_1}(m_{-i})$$
$$+\Psi_{m_i}(\theta_2)\int_{m_{-i}}\Big(x_1(\hat{m}_i,m_{-i}))\cdot 1+x_2(\hat{m}_i,m_{-i})(-\frac{1}{2})+x_3(\hat{m}_i,m_{-i})(-\frac{1}{2})\Big)d\Psi_{m_i,\theta_2}(m_{-i})$$

is minimized at $\hat{m}_i = m_i$ regardless of how we define the allocation rule off the equilibrium path, then incentive compatibility constraints require that: for any m such that $\Xi_{\theta_1}(m) > 0$, we must have $x_2(\hat{m}_i, m_{-i}) = 0$, $\forall \hat{m}_i \neq m_i$; for any m such that $\Xi_{\theta_2}(m) > 0$, we must have $x_1(\hat{m}_i, m_{-i}) = 0$, $\forall \hat{m}_i \neq m_i$.

Fixed any $m_1 \in M_1$, since $\Psi_{m_1}(\theta_1) = \frac{1}{2}$, then there exist $m_2, m_2' \in M_2$ such that $\Xi_{\theta_1}(m_1, m_2) > 0$ and $\Xi_{\theta_2}(m_1, m_2') > 0$. Then we have $x(m_1,m_2) = x^*(\theta_1) = (1,0,0)$, and $x(m_1,m_2') = x^*(\theta_2) = (0,1,0)$, contradicting $x_2(m_1, m_2') = 0$ if m_2' is viewed as agent 2's misreport and the true signal profile is (m_1, m_2). Thus, (Ξ, x) cannot achieve the value of relaxed problem (P^*).

B.3.3 A numerical example where private disclosure is strictly better than public disclosure

We illustrate how private disclosure outperforms public disclosure by revealing no information about the state of the world in the following example. We consider the previous single-object auction environment with two bidders, that is $I = \{1,2\}$. Let $\Theta = \{1,2\}$, where $\theta \sim (1,\frac{1}{2};2,\frac{1}{2})$. Let $V_1 = \{1,2,3\}$, $V_2 = \{1,2\}$, where $v_1 \sim (1,\frac{2}{3};2,\frac{1}{9};3,\frac{2}{9})$ and $v_2 \sim (1,\frac{4}{7};2,\frac{3}{7})$. The valuation functions and the corresponding virtual value functions for both agents are given in Table B.6, B.7.

From the previous results, we only need to consider the case where the principal has strong control over the disclosure process. Since the optimal public disclosure mechanism fully reveals θ to both agents, we find the allocation rules for each realization of θ to maximize the total expected virtual value subject to the monotonicity constraints in (P'_{pub}). When $\theta = 1$, define $q^1(v)$ as the solution to the relaxed public disclosure problem, where all monotonicity

Table B.6: Agent 1's (virtual) value

$y_1(v_1,\theta),\gamma_1(v_1,\theta)$	1	2
1	1, $\frac{1}{2}$	1, 0
2	2, 0	3, 1
3	3, 3	4, 4

Table B.7: Agent 2's (virtual) value

$y_2(v_2,\theta),\gamma_2(v_2,\theta)$	1	2
1	1, $\frac{1}{4}$	1, $\frac{1}{4}$
2	2, 2	2, 2

constraints are ignored (Table B.8); and define $q^2(v)$ as the allocation rule in the optimal public disclosure mechanism (Table B.9).

Table B.8: Solution to relaxed (P'_{pub}), $\theta = 1$

$q^1_1(v_1,v_2),q^1_2(v_1,v_2)$	1	2
1	(1,0)	(0,1)
2	(0,1)	(0,1)
3	(1,0)	(1,0)

Table B.9: Solution to (P'_{pub}), $\theta = 1$

$q^2_1(v_1,v_2),q^2_2(v_1,v_2)$	1	2
1	(1,0)	(0,1)
2	(1,0)	(0,1)
3	(1,0)	(1,0)

We can see that $q^1(v)$ violates only one monotonicity constraint, that is,

$$\mathbb{E}_{v_2}[q^1_1(1,v_2)] = \frac{4}{7} > 0 = \mathbb{E}_{v_2}[q^1_1(2,v_2)].$$

When $\theta = 2$, the solution to the relaxed public disclosure problem, denoted by $q^3(v)$, is indeed the allocation rule in the optimal public disclosure mechanism (Table B.10).

Table B.10: Solution to (P'_{pub}), $\theta = 2$

$q^3_1(v_1,v_2),q^3_2(v_1,v_2)$	1	2
1	(0,1)	(0,1)
2	(1,0)	(0,1)
3	(1,0)	(1,0)

The allocation rule of the optimal private disclosure mechanism is given by (P^*). Particularly, in the linear auction environment it is equivalent to the following problem:

$$(P'_2) \quad \sup_{q(v,\theta)\in Q} \int_\theta \int_v \sum_{i=1}^N q_i(v,\theta)\gamma_i(v_i,\theta)dF_V(v)dF_0(\theta)$$

$$s.t. \quad \forall i, v_i < v'_i:$$

$$\int_\theta \Big(\mathbb{E}_{v_{-i}}\big[q_i(v'_i,v_{-i},\theta)\big] - \mathbb{E}_{v_{-i}}\big[q_i(v_i,v_{-i},\theta)\big]\Big)\big(y_i(v'_i,\theta) - y_i(v_i,\theta)\big)dF_0(\theta) \geq 0.$$

Immediately, the solution to the relaxed problem of (P_2'), where monotonicity constraints ($Mon_{v_i<v_i'}$) are ignored, is given by $q^1(v)$ for $\theta=1$ and $q^3(v)$ for $\theta=2$. To check whether such solution is also feasible for (P_2'), the only monotonicity constraint in (P_2') we need to check is the one involving bidder 1's types $\{1,2\}$, which is satisfied because

$$\begin{aligned}
&\int_\theta \Big(\mathbb{E}_{v_2}\big[q_1(2,v_2,\theta)\big]-\mathbb{E}_{v_2}\big[q_1(1,v_2,\theta)\big]\Big)\big(y_1(2,\theta)-y_1(1,\theta)\big)dF_0(\theta)\\
&=\frac{1}{2}\Big(\mathbb{E}_{v_2}[q_1^1(2,v_2)]-\mathbb{E}_{v_2}[q_1^1(1,v_2)]\Big)\big(y_1(2,1)-y_1(1,1)\big)\\
&\qquad\qquad+\frac{1}{2}\Big(\mathbb{E}_{v_2}[q_1^3(2,v_2)]-\mathbb{E}_{v_2}[q_1^3(1,v_2)]\Big)\big(y_1(2,2)-y_1(1,2)\big)\\
&=\frac{1}{2}\big(0-\frac{4}{7}\big)(2-1)+\frac{1}{2}\big(\frac{4}{7}-0\big)(3-1)=\frac{2}{7}>0.
\end{aligned}$$

Thus, the optimal private disclosure policy Φ^S, with signal sets $M_1=\{a_1,a_2\}$ and $M_2=\{b_1,b_2\}$, is given by Table B.11.

Table B.11: Optimal private disclosure policy Φ^S

$\Phi(1,m_1,m_2)$	b_1	b_2
a_1	0	$\frac{1}{4}$
a_2	$\frac{1}{4}$	0

$\Phi(2,m_1,m_2)$	b_1	b_2
a_1	$\frac{1}{4}$	0
a_2	0	$\frac{1}{4}$

Moreover, for any $v\in V$, the associated allocation rule is defined as

$$q^S(v,m)=\begin{cases} q^1(v), & \text{if } m\in\{(a_1,b_2),(a_2,b_1)\}\\ q^3(v), & \text{if } m\in\{(a_1,b_1),(a_2,b_2)\}\end{cases}$$

The difference in the principal's expected payoff between private disclosure (with strong control) and public disclosure is given by

$$\begin{aligned}
\Pi_{pri}-\Pi_{pub}&=\Pr(\theta=1)\cdot\Pr(v_1=2,v_2=1)\cdot\Big(\gamma_2(1,1)-\gamma_1(2,1)\Big)\\
&=\frac{1}{2}\cdot\frac{1}{9}\cdot\frac{4}{7}\cdot\big(\frac{1}{4}-0\big)=\frac{1}{126}>0.
\end{aligned}$$

Appendix C

Appendix for Chapter 2

C.1 Proof of Theorem 6

Because $\succ_i^{\theta_{-i}} = \succ_i^{\theta'_{-i}}$ for all i, θ_{-i}, and θ'_{-i}, we denote this ordering by $\succ_i$ with no superscript. Also, let $\Theta_i = \{\theta_i^1, \ldots, \theta_i^N\}$ (where $N = |\Theta_i|$) so that $\theta_i^n \prec_i \theta_i^{n+1}$ for all $n = 1, \ldots, N-1$.

Consider the simple type space $\widehat{\mathscr{T}}^f = (T_i, \widehat{\theta}_i, \widehat{\pi}_i)_{i=1}^I$ with $T_i = \Theta_i$ and the agents' beliefs defined by $\widehat{\pi}_i(\theta_i^n)[\theta_{-i}] = \big(\sum_{\theta'_{-i}\in\Theta_{-i}} G_i(\theta_i^n, \theta'_{-i})\big)^{-1} G_i(\theta_i^n, \theta_{-i})$ for all $\theta_{-i} \in \Theta_{-i}$, where $G_i(\theta_i^n, \theta_{-i}) = \sum_{k=n}^N f(\theta_i^k, \theta_{-i})$. By convention, $G_i(\theta_i^{N+1}, \theta_{-i}) = 0$.

The optimal Bayesian mechanism given this simple type space achieves:

$$
\begin{aligned}
V(f) = & \max_{(q,p):\Theta\to Q\times\mathbb{R}^I} \quad \sum_{\theta\in\Theta} f(\theta) \sum_{i\in I} p_i(\theta) \\
s.t. \quad & \forall i \in I, \forall n, l \in \{1, \ldots, N\}, \forall \theta \in \Theta: \\
& \sum_{\theta_{-i}\in\Theta_{-i}} \widehat{\pi}_i(\theta_i^n)[\theta_{-i}] \big(v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i(\theta_i^n, \theta_{-i})\big) \geq 0, \qquad (BIR_i^n) \\
& \sum_{\theta_{-i}\in\Theta_{-i}} \widehat{\pi}_i(\theta_i^n)[\theta_{-i}] \big(v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i(\theta_i^n, \theta_{-i})\big) \\
\geq & \sum_{\theta_{-i}\in\Theta_{-i}} \widehat{\pi}_i(\theta_i^n)[\theta_{-i}] \big(v_i(q_i(\theta_i^l, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i(\theta_i^l, \theta_{-i})\big). \qquad (BIC_i^{n\to l})
\end{aligned}
$$

Because the identity function $\widehat{\theta}_i$ is one-to-one, by Lemma 7, $\widehat{\mathscr{T}}^f$ can be embedded in the universal type space $\mathscr{T}^*$ through a bijection h such that $t_i^n = h_i(\theta_i^n)$. Thus, $V(f)$ provides an upper bound for the best expected revenue given the universal type space $\mathscr{T}^*$ (and the principal's belief $\mu^* \in \mathscr{M}$ such that $\mu^*(h(\widehat{\theta}^{-1}(\theta))) = f(\theta)$). Therefore, in order to show the Bayesian foundation for EPIC mechanisms given f, it suffices to show that $V(f) \leq R_f^{EP}$.

We first prove the claim by imposing the non-singularity condition on f, which assumes that $\Omega_i = (f(\theta_i^1, \cdot), \ldots, f(\theta_i^N, \cdot))^\intercal$ has rank N for each i, where $f(\theta_i^n, \cdot) = (f(\theta_i^1, \theta_{-i}))_{\theta_{-i}\in\Theta_{-i}}$

is a $(I-1)N$-dimensional vector.

Lemma 15. In the solution of $V(f)$, $(BIC_i^{n\to n-1})$ holds with equality for all i and $n \neq 1$, and (BIR_i^n) holds with equality for all i and n.

The lemma implies that, for all i and n:

$$\begin{aligned}&\sum_{\theta_{-i}\in\Theta_{-i}} \widehat{\pi}_i(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^n,\theta_{-i}),\theta_i^n,\theta_{-i}) - p_i(\theta_i^n,\theta_{-i})\big)\\ =&\sum_{\theta_{-i}\in\Theta_{-i}} \widehat{\pi}_i(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^{n-1},\theta_{-i}),\theta_i^n,\theta_{-i}) - p_i(\theta_i^{n-1},\theta_{-i})\big) = 0,\end{aligned}$$

or equivalently:

$$\begin{aligned}&\sum_{\theta_{-i}\in\Theta_{-i}} \Big(\sum_{\theta'_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta'_{-i})\Big)^{-1} G_i(\theta_i^n,\theta_{-i})\big(v_i(q_i(\theta_i^n,\theta_{-i}),\theta_i^n,\theta_{-i}) - p_i(\theta_i^n,\theta_{-i})\big) = 0,\\ &\sum_{\theta_{-i}\in\Theta_{-i}} \Big(\sum_{\theta'_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta'_{-i})\Big)^{-1} G_i(\theta_i^n,\theta_{-i})\big(v_i(q_i(\theta_i^{n-1},\theta_{-i}),\theta_i^n,\theta_{-i}) - p_i(\theta_i^{n-1},\theta_{-i})\big) = 0.\end{aligned}$$

This implies:

$$\begin{aligned}&\sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta_{-i}) v_i(q_i(\theta_i^n,\theta_{-i}),\theta_i^n,\theta_{-i}) = \sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta_{-i}) p_i(\theta_i^n,\theta_{-i}),\\ &\sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta_{-i}) v_i(q_i(\theta_i^{n-1},\theta_{-i}),\theta_i^n,\theta_{-i}) = \sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta_{-i}) p_i(\theta_i^{n-1},\theta_{-i})\big),\end{aligned}$$

and therefore, the objective becomes:

$$\begin{aligned}&\sum_{i\in I}\sum_{n=1}^{N_i}\sum_{\theta_{-i}\in\Theta_{-i}} f(\theta_i^n,\theta_{-i}) p_i(\theta_i^n,\theta_{-i})\\ =&\sum_{i\in I}\sum_{n=1}^{N_i}\sum_{\theta_{-i}\in\Theta_{-i}} \big(G_i(\theta_i^n,\theta_{-i}) - G_i(\theta_i^{n+1},\theta_{-i})\big) p_i(\theta_i^n,\theta_{-i})\\ =&\sum_{i\in I}\sum_{n=1}^{N_i}\left(\sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n,\theta_{-i}) p_i(\theta_i^n,\theta_{-i}) - \sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^{n+1},\theta_{-i}) p_i(\theta_i^n,\theta_{-i})\right)\\ =&\sum_{i\in I}\sum_{n=1}^{N_i}\sum_{\theta_{-i}\in\Theta_{-i}} \big(G_i(\theta_i^n,\theta_{-i}) v_i(q_i(\theta_i^n,\theta_{-i}),\theta_i^n,\theta_{-i}) - G_i(\theta_i^{n+1},\theta_{-i}) v_i(q_i(\theta_i^n,\theta_{-i}),\theta_i^{n+1},\theta_{-i})\big)\\ =&\sum_{i\in I}\sum_{\theta\in\Theta} f(\theta)\gamma_i(q_i,\theta).\end{aligned}$$

Therefore, under Assumption 3, we have $V(f) = R_f^{EP}$.

Proof of Lemma 15. We first show that each upward incentive constraint, $(BIC_i^{n\to l})$ with $n < l$, can be ignored without loss. Let $\Pi_i = \big(\widehat{\pi}_i(\theta_i^1),\ldots,\widehat{\pi}_i(\theta_i^N)\big)^\intercal$ denote the matrix of agent i's beliefs, where each $\widehat{\pi}_i(\theta_i^n)$ is a $(I-1)N$-dimensional vector. Then:

$$\Pi_i = \begin{pmatrix} \kappa_i^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \kappa_i^N \end{pmatrix}_{N\times N} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{N\times N} \Omega,$$

where $\kappa_i^n = \left(\sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^n, \theta_{-i})\right)^{-1}$, and hence Π_i has a rank N. Thus, there exists $\lambda \in \mathbb{R}^{(I-1)N}$ such that:

$$\Pi_i\lambda = (1,\ldots,1, \underbrace{0}_{l\text{-th element}}, \ldots, 0)^\intercal.$$

If we add λ to $p_i(\theta_i^l, \cdot)$, each $BIC_i^{n\to l}$ with $n < l$ is relaxed, while no other (*BIC*) and (*BIR*) constraints are affected. Moreover, from $\widehat{\pi}_i(\theta_i^l)\cdot\lambda = 0$ and $\widehat{\pi}_i(\theta_i^{l+1})\cdot\lambda = 0$, we obtain:

$$\sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^l, \theta_{-i})\lambda(\theta_{-i}) = 0, \qquad \sum_{\theta_{-i}\in\Theta_{-i}} G_i(\theta_i^{l+1}, \theta_{-i})\lambda(\theta_{-i}) = 0,$$

which implies that $\sum_{\theta_{-i}\in\Theta_{-i}} f(\theta_i^l, \theta_{-i})\lambda(\theta_{-i}) = 0$, that is, the principal's expected revenue is also unaffected.

Next, we show that for any mechanism (q,p) satisfying the remaining constraints, there exists a mechanism (q',p') which satisfies not only the remaining constraints, but also (BIR_i^n) for $n = 1,\ldots,N$ and ($BIC_i^{n\to n-1}$) for $n = 2,\ldots,N$ with equality, and raises at least as high expected revenue as (q,p).

Given any such mechanism (q,p), if ($BIC_i^{n\to n-1}$) is satisfied with strict inequality for some i and n, then let $\beta_i^{n\to n-1}$ be the amount of the slackness of this constraint ($BIC_i^{n\to n-1}$). Let Π_i' be the matrix generated by substituting the n-th row of Π_i with the vector $f(\theta^{n-1}, \cdot)$. That is:

$$\Pi_i' = \begin{pmatrix} \kappa_i^1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \kappa_i^{n-1} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \kappa_i^{n+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & \kappa_i^N \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \Omega,$$

and hence, Π_i' has a rank N. Thus, there exists $\lambda \in \mathbb{R}^{(I-1)N}$ such that:

$$\Pi_i'\lambda = (0,\ldots,0, \underbrace{1}_{n\text{-th element}}, 0,\ldots,0)^\intercal.$$

Because $\widehat{\pi}_i(\theta_i^{n-1}) \cdot \lambda = 0$ and $f(\theta^{n-1}, \cdot) \cdot \lambda = 1$, we have:

$$\widehat{\pi}_i(\theta_i^n) \cdot \lambda = \frac{\kappa_i^n}{\kappa_i^{n-1}} \widehat{\pi}_i(\theta_i^{n-1}) \cdot \lambda - \kappa_i^n f(\theta^{n-1}, \cdot) \cdot \lambda < 0,$$

and thus, $\varepsilon = -\beta_i^{n \to n-1} / (\widehat{\pi}_i(\theta_i^n) \cdot \lambda) > 0$. If we add $\varepsilon\lambda$ to $p_i(\theta_i^{n-1}, \cdot)$, then all the constraints for types θ_i^l with $l \neq n$ are unaffected because $\widehat{\pi}_i(\theta_i^l) \cdot \lambda = 0$ for all $l \neq n$, and for type θ_i^n only constraint ($BIC_i^{n \to n-1}$) is changed, which holds with equality under the new payment rule. Because $f(\theta^{n-1}, \cdot) \cdot (\varepsilon\lambda) = \varepsilon > 0$, the expected revenue increases under the new payment rule.

Similarly, if (BIR_i^n) is satisfied with strict inequality for some i and n, then let β_i^n be the amount of the slackness of this constraint (BIR_i^n). Because Π_i has a rank N, there exists $\lambda \in \mathbb{R}^{(I-1)N}$ such that:

$$\Pi_i \lambda = (\beta_i^1, \ldots, \beta_i^N)^\intercal \geq 0.$$

Adding λ to each $p_i(\theta_i^n, \cdot)$ does not affect any (BIC) constraint, while all the participation constraints are satisfied with equality in the new mechanism. The change in the total expected revenue is:

$$\begin{aligned}
\sum_{n=1}^{N} \sum_{\theta_{-i} \in \Theta_{-i}} f(\theta_i^n, \theta_{-i}) \lambda(\theta_{-i}) &= \sum_{\theta_{-i} \in \Theta_{-i}} \lambda(\theta_{-i}) \sum_{n=1}^{N} f(\theta_i^n, \theta_{-i}) \\
&= \sum_{\theta_{-i} \in \Theta_{-i}} \lambda(\theta_{-i}) G_i(\theta_i^1, \theta_{-i}) \\
&= \frac{1}{\kappa_i^1} \sum_{\theta_{-i} \in \Theta_{-i}} \lambda(\theta_{-i}) \widehat{\pi}_i(\theta_i^1)[\theta_{-i}] \\
&= \beta_i^1,
\end{aligned}$$

which is non-negative. □

Next, we consider the case where f is singular, that is, for some i, Ω_i has a rank strictly less than N. Consider a sequence of distributions over Θ, $\{f^r\}_{r=1}^{\infty}$, such that each f^r is full-support and $f_r \to f$ (in the standard Euclidean distance).[1] By Assumption 3, without loss of generality, we assume that the monotonicity constraints (M) are not binding in the problem of $R_{f_r}^{EP}$.

We prove the following continuity lemma.

Lemma 16. For each $\varepsilon > 0$, there exists $r_\varepsilon \in \mathbb{N}$ such that, for any $r \geq r_\varepsilon$, $R_{f_r}^{EP} \leq R_f^{EP} + \varepsilon$ and $V(f_r) \geq V(f) - \varepsilon$.

[1] We can always find such a sequence because the set of all non-singular distributions is a dense subset of the set of all distributions over Θ.

Proof of Lemma 16. For the first inequality, recall that $R_f^{EP} = \sum_i \sum_\theta \max\{\gamma_i(\theta), 0\} f(\theta)$, which is obviously continuous in f.

For the second inequality, let (q, p) be a solution to the problem of $V(f)$.

In the following, for each r, we construct another mechanism (q, p^r) (note that we keep the same q), so that it satisfies all the constraints of the problem of $V(f_r)$, namely:

$$\sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i^r(\theta_i^n, \theta_{-i})\big) \geq 0, \tag{$BIR_i^n(r)$}$$

$$\sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i^r(\theta_i^n, \theta_{-i})\big)$$
$$\geq \sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^l, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i^r(\theta_i^l, \theta_{-i})\big). \tag{$BIC_i^{n\to l}(r)$}$$

Let:

$$S_i^n(r) = \max\left\{0, \sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(p_i(\theta_i^n, \theta_{-i}) - v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i})\big)\right\},$$

denote the size of violation of ($BIR_i^n(r)$) by p. If we consider a modified payment rule p' so that $p_i'(\theta_i^n, \cdot) = p_i(\theta_i^n, \cdot) - S_i^n(r)\mathbf{1}$, then this new payment rule satisfies the participation constraints, but may not satisfy the incentive compatibility constraints. Thus, let:

$$L_i^{n\to l}(r) = \max\Big\{ 0, \sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^l, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i'(\theta_i^l, \theta_{-i})\big)$$
$$- \sum_{\theta_{-i} \in \Theta_{-i}} \widehat{\pi}_i^r(\theta_i^n)[\theta_{-i}]\big(v_i(q_i(\theta_i^n, \theta_{-i}), \theta_i^n, \theta_{-i}) - p_i'(\theta_i^n, \theta_{-i})\big)\Big\},$$

denote the size of violation of ($BIC_i^{n\to l}(r)$) by p'. As in the first part of the proof, the matrix of agent i's belief in the simple type space $\widehat{\mathscr{T}^{f_r}}$, $\Pi_i^r = \big(\widehat{\pi}_i^r(\theta_i^1), \ldots, \widehat{\pi}_i^r(\theta_i^N)\big)^\mathsf{T}$, has a rank N, and hence, there exists $\lambda_i^1(r), \ldots, \lambda_i^N(r) \in \mathbb{R}^{(I-1)N}$ such that:

$$\Pi_i^r\left(\lambda_i^1(r), \ldots, \lambda_i^N(r)\right) = \left(L_i^{n\to l}(r)\right)_{N\times N},$$

which we denote by $\mathbf{L}_r$. Or equivalently:

$$\mathbf{L}_r = \underbrace{\begin{pmatrix} \kappa_i^1(r) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \kappa_i^N(r) \end{pmatrix}_{N\times N}}_{\triangleq \mathbf{K}_r} \underbrace{\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{N\times N}}_{\triangleq \mathbf{A}} \Omega_r\left(\lambda_i^1(r), \ldots, \lambda_i^N(r)\right).$$

Define $p_i^r(\theta_i^n,\cdot) = p_i(\theta_i^n,\cdot) - S_i^n(r)\mathbf{1} + \lambda_i^n(r)$. Then, together with q, it satisfies all the constraints of the problem of $V(f_r)$.

We complete the proof by showing that $\sum_\theta \sum_i (p_i^r(\theta) - p_i(\theta)) f_r(\theta) \to 0$ as $r \to \infty$. Because it is obvious that $S_i^n(r) \to 0$, it suffices to show that:

$$\sum_{n=1}^{N} f_r(\theta_i^n,\cdot)\cdot\lambda_i^n(r) \to 0.$$

Indeed:

$$\sum_{n=1}^{N} f_r(\theta_i^n,\cdot)\cdot\lambda_i^n(r) = tr\big(\mathbf{A}^{-1}\mathbf{K}_r^{-1}\mathbf{L}_r\big) \to 0,$$

as $r \to \infty$, because $\mathbf{L}_r \to 0$. □

Finally, contrarily to the original claim, suppose that $V(f) > R_f^{EP}$, and let $\varepsilon \in (0, \frac{V(f)-R_f^{EP}}{2})$. Then, there exists r_ε such that:

$$V(f_r) - R_{f_r}^{EP} \geq V(f) - R_f^{EP} - 2\varepsilon > 0,$$

which contradicts the first part of this proof.

C.2 Proof of Theorem 7

The previous theorem already states that EPIC mechanisms have the foundation if we do not have ordinal interdependence. Therefore, we only prove its converse in this proof.

We first observe an implication of ordinal interdependence under Assumptions 5.

Lemma 17. Under Assumptions 5, ordinal interdependence implies at least one of the following: (i) there exists i, q_i, q_i', θ_{-i} and θ_{-i}' such that $\theta_i^*(q_i,\theta_{-i}) \notin \Theta_i^*(q_i',\theta_{-i}')$ and $\theta_i^*(q_i',\theta_{-i}') \notin \Theta_i^*(q_i,\theta_{-i})$; or (ii) there exists i, θ_i, q_i, q_i', θ_{-i} and θ_{-i}' such that $\theta_i \in \Theta_i^*(q_i,\theta_{-i}) \setminus \Theta_i^*(q_i',\theta_{-i}')$ and $\theta_i^*(q_i,\theta_{-i}) \in \Theta_i^*(q_i',\theta_{-i}')$.

Proof. By definition of ordinal interdependence, there exists i, $\tilde{\theta}_{-i}$ and $\tilde{\theta}_{-i}'$ such that $\prec_i^{\tilde{\theta}_{-i}} \neq \prec_i^{\tilde{\theta}_{-i}'}$. Single-crossing condition implies that, for any $q_i > 0$, any θ_{-i}, and any distinct pair $\theta_i \neq \theta_i'$, we have $v_i(q_i,\theta_i',\theta_{-i}) < v_i(q_i,\theta_i,\theta_{-i})$ if and only if $\theta_i' \prec_i^{\theta_{-i}} \theta_i$. Thus, there exists θ_i and θ_i' such that $v_i(q_i,\theta_i',\tilde{\theta}_{-i}) < v_i(q_i,\theta_i,\tilde{\theta}_{-i})$ and $v_i(q_i,\theta_i',\tilde{\theta}_{-i}') > v_i(q_i,\theta_i,\tilde{\theta}_{-i}')$ hold for any $q_i > 0$. Fixed any $q_i > 0$, by Assumption 5, there exists θ_{-i} and θ_{-i}' such that

$$\begin{cases} \theta_i \in \Theta_i^*(q_i,\theta_{-i}), & \theta_i' \notin \Theta_i^*(q_i,\theta_{-i}); \\ \theta_i \notin \Theta_i^*(q_i,\theta_{-i}'), & \theta_i' \in \Theta_i^*(q_i,\theta_{-i}'). \end{cases}$$

Next, we show that if (i) is violated, then we must have (ii). Without loss of generality, we can assume that $\theta_i^*(q_i,\theta_{-i}) \in \Theta_i^*(q_i,\theta'_{-i})$. Since $\theta_i \in \Theta_i^*(q_i,\theta_{-i})$ and $\theta_i \notin \Theta_i^*(q_i,\theta'_{-i})$, we have $\theta_i \in \Theta_i^*(q_i,\theta_{-i}) \setminus \Theta_i^*(q_i,\theta'_{-i})$, which means (ii) holds. Therefore, we must have either (i) or (ii) is satisfied.

□

We show that, for each of these cases, there exists a mechanism that yields a strictly higher expected revenue than the optimal EPIC mechanism.

Case (i): $\theta_1^*(q_1,\theta_{-1}) \notin \Theta_1^*(q'_1,\theta'_{-1})$ and $\theta_1^*(q'_1,\theta'_{-1}) \notin \Theta_1^*(q_1,\theta_{-1})$.

Consider a new mechanism (M,q^*,p^*) such that $M_1 = \Theta_1 \times [0,1]$, $M_j = \Theta_j$ for $j \neq 1$, and for each $((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) \in M$,

$$\begin{aligned} q^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) &= q^{EP}(\tilde{\theta}), \\ p_j^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) &= p_j^{EP}(\tilde{\theta}),\ \forall j \neq 1, \end{aligned}$$

and for p_1^*, we set $p_1^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) = p_1^{EP}(\tilde{\theta})$ unless $\tilde{\theta}_1 \in \Theta_1^*(q_1,\theta_{-1}) \cap \Theta_1^*(q'_1,\theta'_{-1})$ and $\tilde{\theta}_{-1} \in \{\theta_{-1},\theta'_{-1}\}$; and for each $\tilde{\theta}_1 \in \Theta_1^*(q_1,\theta_{-1}) \cap \Theta_1^*(q'_1,\theta'_{-1})$, we set

$$\begin{aligned} p_1^*((\tilde{\theta}_1,x),\theta_{-1}) &= p_1^{EP}(\tilde{\theta}_1,\theta_{-1}) + \eta(1-x), \\ p_1^*((\tilde{\theta}_1,x),\theta'_{-1}) &= p_1^{EP}(\tilde{\theta}_1,\theta'_{-1}) + \eta\psi(x), \end{aligned}$$

where $\psi(x) = 1 - \sqrt{1-x^2}$.

Intuitively, $x \in [0,1]$ is related to agent 1's first-order belief over θ_{-i} and θ'_{-i} (more precisely, their likelihood ratio). Indeed, if agent 1 reports his payoff type θ_1 truthfully, his optimal choice of x is given by $x^*(\beta,\beta') = \sqrt{\frac{(\beta/\beta')^2}{1+(\beta/\beta')^2}}$, where β is 1's first-order belief for θ_{-1} and β' is 1's first-order belief for θ'_{-1}. Note that, given any $\mu \in \mathring{\mathscr{M}}$, agent 1 chooses $x \in (0,1)$ with probability one.

It is then obvious that, if the agents report their payoff types truthfully (and agent 1 chooses x optimally), then this new mechanism yields a strictly higher expected revenue than the optimal EPIC mechanism.

For any agent $j \neq 1$, the new mechanism is outcome-equivalent to the optimal EPIC mechanism, and hence satisfies EPIC and EPIR.

We show the incentive compatibility of agent 1 with $\tilde{\theta}_1 \in \Theta_1^*(q_1,\theta_{-1}) \cap \Theta_1^*(q'_1,\theta'_{-1})$ (for the other payoff types, the new mechanism is outcome-equivalent to the optimal EPIC mechanism, and hence satisfies EPIC and EPIR). First, obviously, any deviation to $\hat{\theta}_1 \in \Theta_1^*(q_1,\theta_{-1}) \cap$

$\Theta_1^*(q_1', \theta_{-1}')$ is not profitable. Second, any deviation to $\hat{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}')$ is not profitable either, because, letting β and β' be his first-order beliefs for θ_{-1} and θ_{-1}' respectively, the expected gain by deviation is at most

$$\beta[\eta(1-x^*(\beta,\beta'))] + \beta'[-\eta + \eta\psi(x^*(\beta,\beta'))] \leq 0.$$

Similarly, we can show that any deviation to $\hat{\theta}_1 \in \Theta_1^*(q_1', \theta_{-1}') \setminus \Theta_1^*(q_1, \theta_{-1})$ and $\hat{\theta}_1 \notin \Theta_1^*(q_1', \theta_{-1}') \cup \Theta_1^*(q_1, \theta_{-1})$ is not profitable either.

Case (ii): $\Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}') \neq \emptyset$ and $\theta_1^*(q_1, \theta_{-1}) \in \Theta_1^*(q_1', \theta_{-1}')$.

Consider a new mechanism (M, q^*, p^*) such that $M_1 = \Theta_1 \times [0,1]$, $M_j = \Theta_j$ for $j \neq 1$, and for each $((\tilde{\theta}_1, x), \tilde{\theta}_{-1}) \in M$,

$$\begin{aligned} q^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) &= q^{EP}(\tilde{\theta}), \\ p_j^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) &= p_j^{EP}(\tilde{\theta}),\ \forall j \neq 1, \end{aligned}$$

and for p_1^*, we set $p_1^*((\tilde{\theta}_1,x),\tilde{\theta}_{-1}) = p_1^{EP}(\tilde{\theta})$ unless $\tilde{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}')$ and $\tilde{\theta}_{-1} \in \{\theta_{-i}, \theta_{-i}'\}$; and for each $\tilde{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}')$, we set

$$\begin{aligned} p_1^*((\tilde{\theta}_1,x),\theta_{-1}) &= p_1^{EP}(\tilde{\theta}_1,\theta_{-1}) + \eta(1-x), \\ p_1^*((\tilde{\theta}_1,x),\theta_{-1}') &= p_1^{EP}(\tilde{\theta}_1,\theta_{-1}') + \eta\psi(x), \end{aligned}$$

where $\psi(x) = 1 - \sqrt{1-x^2}$.

Again, $x \in [0,1]$ is related to agent 1's first-order belief over θ_{-i} and θ_{-i}'. Indeed, if agent 1 reports his payoff type θ_1 truthfully, his optimal choice of x is given by $x^*(\beta,\beta') = \sqrt{\frac{(\beta/\beta')^2}{1+(\beta/\beta')^2}}$, where β is 1's first-order belief for θ_{-1} and β' is 1's first-order belief for θ_{-1}'. Note that, given any $\mu \in \mathring{\mathscr{M}}$, agent 1 chooses $x \in (0,1)$ with probability one.

It is obvious that, if the agents report their payoff types truthfully (and agent 1 chooses x optimally), then this new mechanism yields a strictly higher expected revenue than the optimal EPIC mechanism.

For any agent $j \neq 1$, the new mechanism is outcome-equivalent to the optimal EPIC mechanism, and hence satisfies EPIC and EPIR.

We show the incentive compatibility of agent 1 with $\tilde{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}')$ (for the other payoff types, the new mechanism is outcome-equivalent to the optimal EPIC mechanism, and hence satisfies EPIC and EPIR). First, obviously, any deviation to $\hat{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \setminus \Theta_1^*(q_1', \theta_{-1}')$ is not profitable. Second, any deviation to $\hat{\theta}_1 \in \Theta_1^*(q_1, \theta_{-1}) \cap \Theta_1^*(q_1', \theta_{-1}')$ is not profitable either, because, letting β and β' be his first-order beliefs for θ_{-1} and θ'_{-1} respectively, the expected gain by deviation is at most

$$\beta[\eta(1-x^*(\beta,\beta'))]+\beta'[-\eta+\eta\psi(x^*(\beta,\beta'))]\leq 0.$$

Similarly, we can show that any deviation to $\hat{\theta}_1 \in \Theta_1^*(q'_1,\theta'_{-1})\setminus\Theta_1^*(q_1,\theta_{-1})$ and $\hat{\theta}_1 \notin \Theta_1^*(q'_1,\theta'_{-1})\cup\Theta_1^*(q_1,\theta_{-1})$ is not profitable either.

In conclusion, EPIC mechanisms do not have the strong foundation.

C.3 Proof of Proposition 5

Assume that $(i,\theta_i,\theta_j,q_j,\tilde{\theta}_{-ij})$ satisfies the definition of strong improvability. We use the same mechanism as above, except that the allocation for agent i changes in case he reports θ_i and $x=1$. Recall that, given his truthfully reporting θ_i, agent 1's optimal choice of x is $\sqrt{\frac{(\beta/\beta')^2}{1+(\beta/\beta')^2}}$ where β, β' are his first-order beliefs for θ_{-i}, θ'_{-i}, respectively, with $\theta_{-i}=(\theta_j,\tilde{\theta}_{-ij})$ and $\theta'_{-i}=(\theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij}),\tilde{\theta}_{-ij})$; $x=1$ means that he predicts that j does not have a threshold type for q_j given $\tilde{\theta}_{-ij}$. The allocations from agents i and j are then modified as follows (and all the other parts of the mechanism are the same as before):

$$\begin{aligned}
q_j^{**}((\theta_i,1),\theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij}),\tilde{\theta}_{-ij}) &= q_j^*((\theta_i,1),\hat{\theta}_j,\tilde{\theta}_{-ij}),\\
p_j^{**}((\theta_i,1),\theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij}),\tilde{\theta}_{-ij}) &= p_j^*((\theta_i,1),\hat{\theta}_j,\tilde{\theta}_{-ij}),\\
p_j^{**}((\theta_i,1),\theta_j,\tilde{\theta}_{-ij}) &= p_j^*(\theta_i,\theta_j,\tilde{\theta}_{-ij})+\eta, \qquad \forall\theta_j \succ_j^{\theta_i,\tilde{\theta}_{-ij}} \theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij})\\
p_i^{**}((\theta_i,1),\theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij}),\tilde{\theta}_{-ij}) &= M,
\end{aligned}$$

where $\hat{\theta}_j$ is j's payoff type that is just below $\theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij})$ with respect to $\prec_j^{\theta_i,\tilde{\theta}_{-ij}}$, and $M>0$ is sufficiently large.

Observe that the modified mechanism satisfies all the constraints. First, except for agents i and j, the allocations are the same as in the previous mechanism. For agent i, large fine M is irrelevant unless he assigns zero probability for θ_{-i} (because $x=1$ is not optimal for him); on the other hand, if he assigns zero probability for θ_{-i}, then this large fine is payoff-irrelevant for him. Finally, for agent j, we only need to check his incentive if i reports $(\theta_i,1)$ and $-ij$ report $\tilde{\theta}_{ij}$: in such a case, j with payoff type $\tilde{\theta}_j \precsim_i^{\theta_i,\tilde{\theta}_{-ij}} \theta_j^*(q_j,\theta_i,\tilde{\theta}_{-ij})$ has no incentive of misreporting, because their on-path payoffs would be the same as in the original mechanism, while the other types' payments are higher than in the original mechanism. For $\tilde{\theta}_j \succ_i^{\theta_i,\tilde{\theta}_{-ij}}$

$\theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij})$, his payoff by deviation is at most

$$\begin{aligned} & v_j(q_j^*((\theta_i,1),\hat{\theta}_j,\tilde{\theta}_{-ij}),\theta_i,\tilde{\theta}_j,\tilde{\theta}_{-ij}) - p_j^*((\theta_i,1),\hat{\theta}_j,\tilde{\theta}_{-ij}) \\ \leq \; & v_j(q_j^*((\theta_i,1),\tilde{\theta}_j,\tilde{\theta}_{-ij}),\theta_i,\tilde{\theta}_j,\tilde{\theta}_{-ij}) - p_j^*((\theta_i,1),\tilde{\theta}_j,\tilde{\theta}_{-ij}) - \eta, \end{aligned}$$

but the right-hand side is precisely his on-path payoff. The individual rationality constraints can be checked similarly.

Finally, we show that this modified mechanism achieves a strictly higher expected revenue than the original mechanism. First, observe that it does not yield a lower payoff given any payoff-type profile. It is obvious except when a payoff type $(\theta_i, \theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij}), \tilde{\theta}_{-ij})$ and agent i chooses $x = 1$; if this is the realized payoff-type profile, and agent i reports $x = 1$, agent i pays a large fine M. Therefore, the principal would be better off by setting M large enough.

Moreover, consider a payoff-type profile $(\theta_i, \tilde{\theta}_j, \tilde{\theta}_{-ij})$ such that $\tilde{\theta}_j \succ_i^{\theta_i, \tilde{\theta}_{-ij}} \theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij})$. If agent i chooses $x < 1$ (at least with a positive probability), then i pays $\eta(1-x)(>0)$ more than in the original mechanism, and hence, strict improvement is achieved. If agent i chooses $x = 1$ (with probability one), then the principal increases j's payment by η as explained above, and thus, again strict improvement is achieved.

C.4 Proof of Theorem 8

It suffices to show that ordinal interdependence implies strong improvability.

By ordinal interdependence, there exist $i, \theta_{-i}, \theta'_{-i}$ such that $\prec_i^{\theta_{-i}} \neq \prec_i^{\theta'_{-i}}$. We first observe the following lemma.

Lemma 18. Ordinal interdependence implies that there exist $j \neq i$, θ_j, θ'_j, and $\tilde{\theta}_{-ij}$ such that $\prec_i^{\theta_j, \tilde{\theta}_{-ij}} \neq \prec_i^{\theta'_j, \tilde{\theta}_{-ij}}$.

Proof. Let $i = 1$ without loss of generality, and for each $n = 1, \ldots, I$, let $\theta_{-1}^n = ((\theta'_j)_{j=2}^n, (\theta_j)_{j=n+1}^I)$. Note that $\theta_{-1}^1 = \theta_{-1}$ and $\theta_{-1}^I = \theta'_{-1}$.

If $\prec_1^{\theta_{-1}^{n-1}} = \prec_1^{\theta_{-1}^n}$ for all $n = 2, \ldots, I$, then we have $\theta_{-1}^1 = \theta_{-1}^I$, contradicting that $\prec_1^{\theta_{-1}} \neq \prec_1^{\theta'_{-1}}$. Therefore, there exists $n \in \{2, \ldots, I\}$ such that $\prec_1^{\theta_{-1}^{n-1}} \neq \prec_1^{\theta_{-1}^n}$. We complete the proof of the lemma by setting $j = n$ and $\tilde{\theta}_{-1j} = ((\theta'_k)_{k=2}^{n-1}, (\theta_k)_{k=n+1}^I)$. □

By the lemma, there exists θ_i, θ'_i such that $\theta_i \succ_i^{(\theta_j, \tilde{\theta}_{-ij})} \theta'_i$ and $\theta'_i \succ_i^{(\theta'_j, \tilde{\theta}_{-ij})} \theta_i$. Letting $q_i = q_i^{EP}(\theta'_i, \theta_j, \tilde{\theta}_{-ij})$ and $q'_i = q_i^{EP}(\theta'_i, \theta'_j, \tilde{\theta}_{-ij})$, by Assumption 6, we have $\theta'_i = \theta_i^*(q'_i, \theta'_j, \tilde{\theta}_{-ij}) =$

$\theta_i^*(q_i, \theta_j, \tilde{\theta}_{-ij})$. It follows that $\theta_i \in \Theta_i^*(q_i, \theta_i, \tilde{\theta}_{-ij}) \setminus \Theta_i^*(q_i', \theta_j', \tilde{\theta}_{-ij})$ and $\theta_i^*(q_i, \theta_j, \tilde{\theta}_{-ij}) \in \Theta_i^*(q_i', \theta_j', \tilde{\theta}_{-ij})$. Then, revenue from agent i is improvable with respect to $\big(\theta_i, (\theta_j, \tilde{\theta}_{-ij}), (\theta_j', \tilde{\theta}_{-ij})\big)$.

Without loss of generality, we assume that $\theta_j' \prec_j^{\theta_i, \tilde{\theta}_{-ij}} \theta_j$. Letting $q_j = q_j^{EP}(\theta_j', \theta_i, \tilde{\theta}_{-ij})$, by Assumption 6, we have $\theta_j' = \theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij})$ and $\theta_j \in \Theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij}) \setminus \{\theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij})\}$. Thus, revenue from i is improvable with respect to $\big(\theta_i, (\theta_j, \tilde{\theta}_{-ij}), (\theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij}), \tilde{\theta}_{-ij})\big)$, where $\theta_j \in \Theta_j^*(q_j, \theta_i, \tilde{\theta}_{-ij})$, which establishes the strong improvability.

Appendix D

Appendix for Chapter 3

D.1 Omitted Proofs

D.1.1 Proof of Theorem 9

Proof. Let $\mathring{w} = \frac{y-x}{t_2-t_1}$, then we have $x-(t_1-t_0+1)\mathring{w} = y-(t_2-t_0+1)\mathring{w}$ for any t_0. It follows that $U_{i|t_0}(C^x_{t_1}, w) \geq U_{i|t_0}(C^y_{t_2}, w)$ for all $w \geq \mathring{w}$, and $U_{i|t_0}(C^x_{t_1}, w) \leq U_{i|t_0}(C^y_{t_2}, w)$ for all $w \leq \mathring{w}$. Moreover, $\mathring{w}$ and $\bar{w}(C^x_{t_1}, t_0) = \frac{x}{t_1-t_0+1}$, $\bar{w}(C^y_{t_2}, t_0) = \frac{y}{t_2-t_0+1}$ satisfy

$$\bar{w}(C^y_{t_2}, t_0) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \bar{w}(C^x_{t_1}, t_0) \Longrightarrow \mathring{w} \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \bar{w}(C^y_{t_2}, t_0) \begin{Bmatrix} \leq \\ \geq \end{Bmatrix} \bar{w}(C^x_{t_1}, t_0). \tag{D.1}$$

(Part 1) If $C^{\text{eff}} = C^x_{t_1}$, then we have $U_{P|0}(C^x_{t_1}) \geq U_{P|0}(C^y_{t_2})$, which means

$$\int_0^{\frac{x}{t_1+1}} \big(x-(t_1+1)w\big) dF(w) - \int_0^{\frac{y}{t_2+1}} \big(y-(t_2+1)w\big) dF(w) \geq 0.$$

We must have $\bar{w}(C^y_{t_2}, 0) < \bar{w}(C^x_{t_1}, 0)$; otherwise from (D.1) we would have $\mathring{w} \geq \bar{w}(C^y_{t_2}, 0) \geq \bar{w}(C^x_{t_1}, 0)$, which means agents who remain active all prefer $C^y_{t_2}$ regardless of the principal's time-0 choice, and thus the efficient outcome should also be $C^y_{t_2}$.

If $x \geq y$, $C^{\text{equ}} = C^x_{t_1}$ holds trivially, because all agents prefer an earlier higher consumption to a later lower one. If $\frac{t_1+1}{t_2+1} y < x < y$, we show that $C^x_{t_1}$ is the unique SPE-NCF outcome. In any SPE, agents either quit immediately at time 0 or stay until the public good is consumed. Let $I(w) := \{i \in I \mid w(i) \leq w\}$ denote the subset of agents whose outside options are at most w. Thus, SPE outcome is restricted to two cases: $I^*_{t_d} = I\big(\bar{w}(C, 0)\big)$, where $C \in \{C^x_{t_1}, C^y_{t_2}\}$. Next, we check whether C is the principal's optimal time-t_d choice when the subset of active agents is $I\big(\bar{w}(C,0)\big)$. Consider

$$
\begin{aligned}
& U_{P|t_d}\big(C^x_{t_1} \mid \frac{x}{t_1+1}\big) - U_{P|t_d}\big(C^y_{t_2} \mid \frac{x}{t_1+1}\big) \\
&= \int_0^{\frac{x}{t_1+1}} \big(x-(t_1-t_d+1)w\big)dF(w) - \int_0^{\frac{x}{t_1+1}} \max\big\{0, y-(t_2-t_d+1)w\big\}dF(w) \\
&= \int_0^{\frac{x}{t_1+1}} \big(x-(t_1-t_d+1)w\big)dF(w) - \int_0^{\frac{y}{t_2+1}} \max\big\{0, \underbrace{y-(t_2-t_d+1)w}_{\geq 0}\big\}dF(w) \\
&\quad - \int_{\frac{y}{t_2+1}}^{\frac{x}{t_1+1}} \max\big\{0, y-(t_2-t_d+1)w\big\}dF(w) \\
&= \Big(U_{P|0}(C^x_{t_1}) - U_{P|0}(C^y_{t_2})\Big) + \int_{\frac{y}{t_2+1}}^{\frac{x}{t_1+1}} \Big(t_d w - \max\big\{0, y-(t_2-t_d+1)w\big\}\Big)dF(w).
\end{aligned}
$$

Notice that for all $w \in [\frac{y}{t_2+1}, \frac{x}{t_1+1}]$ we have

$$
t_d w - \max\big\{0, y-(t_2-t_d+1)w\big\} = \min\{t_d w, (t_2+1)w - y\} \geq 0,
$$

then we have $U_{P|t_d}\big(C^x_{t_1} \mid \frac{x}{t_1+1}\big) - U_{P|t_d}\big(C^y_{t_2} \mid \frac{x}{t_1+1}\big) \geq 0$, which means $C = C^x_{t_1}$ is a SPE outcome. If $C = C^y_{t_2}$ fails to be a SPE outcome, then the unique SPE is indeed the unique SPE-NCF. If $C = C^y_{t_2}$ is a SPE outcome, it is still not a SPE-NCF outcome, because agents in the subset $I\big(\bar{w}(C^x_{t_1},0)\big) \setminus I\big(\bar{w}(C^y_{t_2},0)\big)$ may form a coalition and jointly deviate from quitting at time 0 to quitting at time t_1+1, which alters the principal's time-t_d choice from $C^y_{t_2}$ to $C^x_{t_1}$, and (weakly) benefits all agents in the coalition. Thus, $C^x_{t_1}$ is the unique SPE-NCF outcome, i.e. $C^{\text{equ}} = C^x_{t_1}$.

(Part 2) If $\alpha = \beta$, all agents (except a measure-zero subset) have the same outside option. Then the principal is essentially faced with a single agent. By Lemma 9, there does not exist such x and y satisfying $C^{\text{eff}} = C^y_{t_2}$ and $C^{\text{equ}} = C^x_{t_1}$.

If $\alpha < \beta$, choose ε such that $0 < \varepsilon < \min\big\{\beta-\alpha, (t_2-t_1)\frac{\beta-\alpha}{2}F\big(\frac{\beta+\alpha}{2}\big)\big\}$. Let $y := (t_2+1)\beta - \varepsilon$. Define $\varphi(x)$ as the difference between $U_{P|0}(C^x_{t_1})$ and $U_{P|0}(C^y_{t_2})$:

$$
\varphi(x) = \int_0^{\frac{x}{t_1+1}} \big(x-(t_1+1)w\big)dF(w) - \int_0^{\frac{y}{t_2+1}} \big(y-(t_2+1)w\big)dF(w).
$$

First, we have

$$
\begin{aligned}
\varphi\big(\beta(t_1+1)\big) &= \int_0^{\beta} (\beta-w)(t_1+1)dF(w) - \int_0^{\beta-\frac{\varepsilon}{t_2+1}} \big((\beta-w)(t_2+1)-\varepsilon\big)dF(w) \\
&= -\int_0^{\beta-\frac{\varepsilon}{t_2+1}} (\beta-w)(t_2-t_1)dF(w) + \int_{\beta-\frac{\varepsilon}{t_2+1}}^{\beta} (\beta-w)(t_1+1)dF(w) + \varepsilon F\big(\beta-\frac{\varepsilon}{t_2+1}\big) \\
&\leq -\int_0^{\beta-\frac{\beta-\alpha}{2}} (\beta-w)(t_2-t_1)dF(w) + \frac{t_1+1}{t_2+1}\varepsilon\big[1-F\big(\beta-\frac{\varepsilon}{t_2+1}\big)\big] + \varepsilon F\big(\beta-\frac{\varepsilon}{t_2+1}\big) \\
&\leq -(t_2-t_1)\frac{\beta-\alpha}{2}F\big(\frac{\beta+\alpha}{2}\big) + \varepsilon < 0.
\end{aligned}
$$

By choosing $x = y - \alpha(t_2 - t_1)$, we have $\mathring{w} = \alpha$. Notice that $U_{i|0}(C^x_{t_1}, w) \geq U_{i|0}(C^y_{t_2}, w)$ for all $w \geq \mathring{w}$, and strict inequality holds for $\frac{x}{t_1+1} > w > \mathring{w}$, then we have

$$\varphi\big(y - \alpha(t_2 - t_1)\big) = \int_\alpha^\beta U_{i|0}(C^x_{t_1}, w)dF(w) - \int_\alpha^\beta U_{i|0}(C^y_{t_2}, w)dF(w) > 0.$$

Since $\frac{d}{dx}\varphi(x) = F\big(\frac{x}{t_1+1}\big) \geq 0$, then there exists unique $\tilde{x} \in \Big(\beta(t_1+1), y - \alpha(t_2 - t_1)\Big)$ such that $\varphi(\tilde{x}) = 0$. Immediately, we have

$$\begin{aligned} &\alpha < \frac{y}{t_2+1} < \beta < \frac{\tilde{x}}{t_1+1} < \frac{\tilde{x}}{t_1 - t_d + 1}, \quad \mathring{w} = \frac{y - \tilde{x}}{t_2 - t_1} > \alpha, \\ &\frac{y}{t_2 - t_d + 1} = \frac{(t_2+1)\beta - \varepsilon}{t_2 - t_d + 1} > \frac{(t_2+1)\beta - (\beta - \alpha)}{t_2 - t_d + 1} = \frac{t_2\beta + \alpha}{t_2 - t_d + 1} > \beta. \end{aligned}$$

Thus, in SPE where $C^{\tilde{x}}_{t_1}$ is implemented, all agent quit at time $t_1 + 1$; while in SPE where $C^y_{t_2}$ is implemented, agents with $w \leq \frac{y}{t_2+1}$ quit at time $t_2 + 1$ and the rest quit at time 0. Given $I^*_{t_d} = I$, we have $U_{P|t_d}\big(C^{\tilde{x}}_{t_1} \mid I\big) - U_{P|t_d}\big(C^y_{t_2} \mid I\big) =$

$$\begin{aligned} &\int_\alpha^\beta \big(\tilde{x} - (t_1 - t_d + 1)w\big)dF(w) - \int_\alpha^\beta \big(y - (t_2 - t_d + 1)w\big)dF(w) \\ =&\int_\alpha^\beta (\tilde{x} - (t_1+1)w)dF(w) - \int_\alpha^{\frac{y}{t_2+1}} (y - (t_2+1)w)dF(w) - \int_{\frac{y}{t_2+1}}^\beta (y - (t_2+1)w)dF(w) \\ =&\varphi(\tilde{x}) - \int_{\frac{y}{t_2+1}}^\beta (y - (t_2+1)w)dF(w) = \int_{\frac{y}{t_2+1}}^\beta \big((t_2+1)w - y\big)dF(w), \end{aligned}$$

which is bounded from below by some real number $\varepsilon' > 0$, and is independent of $\tilde{x}$. Since $\frac{d\varphi(x)}{dx}\,|_{x=\tilde{x}} = F\big(\frac{\tilde{x}}{t_1+1}\big) = 1$, choose x' such that $\max\{\beta(t_1+1), \tilde{x} - \varepsilon'\} < x' < \tilde{x}$, and we have $\varphi(x') < 0$ and $U_{P|t_d}\big(C^{x'}_{t_1} \mid I\big) > U_{P|t_d}\big(C^y_{t_2} \mid I\big)$. Thus, $C^{x'}_{t_1}$ is indeed a SPE outcome. Through the same argument as in Part 1, $C^{x'}_{t_1}$ is the unique SPE-NCF outcome. We conclude that x' and y satisfy $C^{\text{eff}} = C^y_{t_2}$ and $C^{\text{equ}} = C^{x'}_{t_1}$. □

D.1.2 Proof of Lemma 10

Proof. At time 1, the present value of common consumption stream (c_1, c_2) for agent with outside option w is given by[1]

$$\text{If } c_1 \geq c_2: \quad U_{i|1}(c_1,c_2,w) = \begin{cases} c_1+c_2-2w & \text{if } w \leq c_2 & \text{quit at } t=3 \\ c_1-w & \text{if } c_2 < w \leq c_1 & \text{quit at } t=2 \\ 0 & \text{if } w > c_1 & \text{quit at } t=1; \end{cases}$$

$$\text{If } c_1 < c_2: \quad U_{i|1}(c_1,c_2,w) = \begin{cases} c_1+c_2-2w & \text{if } w \leq \frac{c_1+c_2}{2} & \text{quit at } t=3 \\ 0 & \text{if } w > \frac{c_1+c_2}{2} & \text{quit at } t=1. \end{cases}$$

Thus, at time 1 the principal's problem is to maximize $U_{P|1}\big(\gamma\delta y, (1-\gamma)y \mid I_1^*\big) =$

$$\begin{cases} \int_0^{(1-\gamma)y} \big(\gamma\delta y + (1-\gamma)y - 2w\big)\rho(w,I_1^*)dF(w) \\ \qquad\qquad + \int_{(1-\gamma)y}^{\gamma\delta y} \big(\gamma\delta y - w\big)\rho(w,I_1^*)dF(w) & \text{if } \gamma\delta \geq 1-\gamma \\ \int_0^{\frac{\gamma\delta y+(1-\gamma)y}{2}} \big(\gamma\delta y + (1-\gamma)y - 2w\big)\rho(w,I_1^*)dF(w) & \text{if } \gamma\delta < 1-\gamma, \end{cases}$$

where

$$\rho(w,I_1^*) = \frac{\int_{w(i)=w} \mathbf{1}_{\{i\in I_1^*\}}di}{\int_{w(i)=w} di}.$$

Since we have $\frac{d}{d\gamma}U_{P|1}\big(\gamma\delta y, (1-\gamma)y \mid I_1^*\big) =$

$$\begin{cases} \delta y \int_0^{\gamma\delta y} \rho(w,I_1^*)dF(w) - y\int_0^{(1-\gamma)y} \rho(w,I_1^*)dF(w) & \text{if } \gamma\delta \geq 1-\gamma \\ (\delta-1)y\int_0^{\frac{\gamma\delta y+(1-\gamma)y}{2}} \rho(w,I_1^*)dF(w) & \text{if } \gamma\delta < 1-\gamma, \end{cases}$$

and $\frac{d^2}{d\gamma^2}U_{P|1}\big(\gamma\delta y, (1-\gamma)y \mid I_1^*\big) =$

$$\begin{cases} (\delta y)^2\rho(\gamma\delta y, I_1^*)f(\gamma\delta y) + y^2\rho((1-\gamma)y, I_1^*)f((1-\gamma)y) & \text{if } \gamma \geq \frac{1}{1+\delta} \\ \frac{1}{2}(\delta-1)^2y^2\rho(\frac{\gamma\delta y+(1-\gamma)y}{2}, I_1^*)f(\frac{\gamma\delta y+(1-\gamma)y}{2}) & \text{if } \gamma < \frac{1}{1+\delta}, \end{cases}$$

immediately we have $\frac{d^2}{d\gamma^2}U_{P|1} \geq 0$ for any $\gamma \in (0,1)$, which means $U_{P|1}$ is convex over $[0,1]$ with respect to γ. Thus, principal's program has corner solutions, that is,

$$\max_{\gamma\in[0,1]} U_{P|1}(\gamma) = \max\{U_{P|1}(0), U_{P|1}(1)\}.$$

□

[1] Quitting at time 3 means quitting at the end of time 2. Alternatively, we can assume that agents are infinitely lived in discrete periods $\{0,1,2,\dots\}$. Since there is no positive consumption after time 2, all agents will quit at the beginning of time 3.

D.1.3 Proof of Proposition 6

Proof. Take derivatives of the implicit function which gives the definition of $\tilde{x}(y)$, and we get $\frac{d\tilde{x}(y)}{dy} = \frac{F(\frac{y}{3})}{F(\frac{\tilde{x}(y)}{2})}$. It follows that

$$\frac{d\tilde{\delta}(y)}{dy} = \frac{1}{y}\Big(\frac{d\tilde{x}(y)}{dy} - \frac{\tilde{x}(y)}{y}\Big) = \frac{1}{y^2 F(\frac{\tilde{x}(y)}{2})}\underbrace{\Big(yF(\frac{y}{3}) - F(\frac{\tilde{x}(y)}{2})\tilde{x}(y)\Big)}_{:=\tau(y)}.$$

Since $\tilde{x}(0) = 0$, we have $\tau(0) = 0$. Moreover, we have

$$\frac{d\tau(y)}{dy} = F(\frac{y}{3})\Big[\frac{f(\frac{y}{3})}{F(\frac{y}{3})}\frac{y}{3} - \frac{f(\frac{\tilde{x}(y)}{2})}{\frac{\tilde{x}(y)}{2}}\frac{\tilde{x}(y)}{2}\Big] = F(\frac{y}{3})\Big[e_F(\frac{y}{3}) - e_F(\frac{\tilde{x}(y)}{2})\Big].$$

If $e_F(\cdot)$ is monotonically increasing, since $\frac{y}{3} < \frac{\tilde{x}(y)}{2}$ for any y, we have $\frac{d\tau(y)}{dy} \leq 0$, which means $\tau(y) \leq 0$ for any y. It follows that $\frac{d\tilde{\delta}(y)}{dy} \leq 0$. Then $\tilde{\delta}(y)$ is monotonically decreasing. Through a symmetric argument we can show that if $e_F(\cdot)$ is monotonically decreasing, then $\tilde{\delta}(y)$ is monotonically increasing.

To prove the remaining statement of the proposition, we define $\hat{e}_F(w) := \frac{F(w)w}{\int_0^w F(w')dw'}$. We will show that if $e_F(w)$ is monotonically increasing (or decreasing), then $\hat{e}_F(w)$ is also monotonically increasing (or decreasing). Obviously, $e_F(w)$ is continuous and $\hat{e}_F(w)$ is differentiable. First, we have

$$\begin{aligned}\frac{d\hat{e}_F(w)}{dw} &= \frac{wf(w)\int_0^w F(w')dw' + F(w)\int_0^w F(w')dw' - w\big(F(w)\big)^2}{\big(\int_0^w F(w')dw'\big)^2} \\ &= \frac{F(w)}{\big(\int_0^w F(w')dw'\big)^2}\Big[\big(1 + \frac{wf(w)}{F(w)}\big)\int_0^w F(w')dw' - wF(w)\Big] \\ &= \frac{F(w)}{\big(\int_0^w F(w')dw'\big)^2}\Big[\underbrace{\big(1 + e_F(w)\big)\int_0^w F(w')dw' - wF(w)}_{:=\tau(w)}\Big].\end{aligned}$$

Pick any small $\varepsilon > 0$, and omit higher-order infinitesimal terms, then we have

$$\begin{aligned}\tau(w+\varepsilon) - \tau(w) &= \big(e_F(w+\varepsilon) - e_F(w)\big)\int_0^w F(w')dw' + \int_w^{w+\varepsilon} F(w')dw' \\ &\quad + e_F(w+\varepsilon)\int_w^{w+\varepsilon} F(w')dw' - w\big(F(w+\varepsilon) - F(w)\big) - \varepsilon F(w+\varepsilon) \\ &\simeq \big(e_F(w+\varepsilon) - e_F(w)\big)\int_0^w F(w')dw' + e_F(w+\varepsilon)F(w)\varepsilon - wf(w)\varepsilon \\ &= \big(e_F(w+\varepsilon) - e_F(w)\big)\int_0^w F(w')dw' + \Big(e_F(w+\varepsilon) - \frac{wf(w)}{F(w)}\Big)F(w)\varepsilon \\ &= \big(e_F(w+\varepsilon) - e_F(w)\big)\Big(\int_0^w F(w')dw' + F(w)\varepsilon\Big).\end{aligned}$$

If $e_F(\cdot)$ is monotonically increasing, then $\tau(w)$ is also increasing. Notice that $\tau(0) = 0$, then $\tau(w) \geq 0$ for all w, which means $\hat{e}_F(w)$ is also increasing. Similarly, If $e_F(\cdot)$ is monotonically decreasing, then $\hat{e}_F(w)$ is also monotonically decreasing.

Take derivatives of the implicit function which defines $\hat{x}(y)$, and we get

$$\frac{d\hat{x}(y)}{dy} = \frac{F(\frac{\hat{x}(y)}{2})}{\frac{1}{2}\big(\frac{3}{2}\hat{x}(y) - y\big)f(\frac{\hat{x}(y)}{2}) + F(\frac{\hat{x}(y)}{2})} = \frac{\big(F(\frac{\hat{x}(y)}{2})\big)^2}{\frac{1}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw + \big(F(\frac{\hat{x}(y)}{2})\big)^2}.$$

Then we have

$$\begin{aligned}
\frac{d\hat{\delta}(y)}{dy} &= \frac{1}{y}\Big[\frac{d\hat{x}(y)}{dy} - \frac{\tilde{x}(y)}{y}\Big] = \frac{1}{y}\Big[\frac{\big(F(\frac{\hat{x}(y)}{2})\big)^2}{\frac{1}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw + \big(F(\frac{\hat{x}(y)}{2})\big)^2} - \frac{\tilde{x}(y)}{y}\Big] \\
&= \frac{y\big(F(\frac{\hat{x}(y)}{2})\big)^2 - \tilde{x}(y)\big[\frac{1}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw + \big(F(\frac{\hat{x}(y)}{2})\big)^2\big]}{y^2\big[\frac{1}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw + \big(F(\frac{\hat{x}(y)}{2})\big)^2\big]} \\
&= \frac{\frac{\tilde{x}(y)}{2}\big(F(\frac{\hat{x}(y)}{2})\big)^2 - F(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw - \frac{\tilde{x}(y)}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw}{y^2\big[\frac{1}{2}f(\frac{\hat{x}(y)}{2})\int_0^{\frac{\hat{x}(y)}{2}} F(w)dw + \big(F(\frac{\hat{x}(y)}{2})\big)^2\big]} \\
&= \frac{x\big(F(x)\big)^2 - F(x)\int_0^x F(w)dw - xf(x)\int_0^x F(w)dw}{y^2\big[\frac{1}{2}f(x)\int_0^x F(w)dw + \big(F(x)\big)^2\big]}\Bigg|_{x=\frac{\hat{x}(y)}{2}} \\
&= \frac{-\big(\int_0^x F(w)dw\big)^2}{y^2\big[\frac{1}{2}f(x)\int_0^x F(w)dw + \big(F(x)\big)^2\big]}\frac{d}{dx}\Big(\frac{xF(x)}{\int_0^x F(w)dw}\Big)\Bigg|_{x=\frac{\hat{x}(y)}{2}},
\end{aligned}$$

which means the sign of $\frac{d\hat{\delta}(y)}{dy}$ is always the opposite of the sign of $\frac{d\hat{e}_F(x)}{dx}\Big|_{x=\frac{\hat{x}(y)}{2}}$. Thus we conclude that $\hat{\delta}(y)$ is monotonically decreasing (or increasing) as long as $\hat{e}_F(w)$ is monotonically increasing (or decreasing). □

D.1.4 Proof of Proposition 7

Proof. By assumption, $F(w)$ has full-support continuous density, which means $e_F(w)$ is continuous. First we show that $F(w) = e^{-\int_w^1 \frac{e_F(x)}{x}dx}$ for all $w \in (0,1]$ and meanwhile $\lim_{w\to 0} e^{-\int_w^1 \frac{e_F(x)}{x}dx} = 0 = F(0)$. By definition, $F(w)$ can be viewed as the solution of the following ordinary differential equation:

$$\begin{cases} \dot{v} = \frac{1}{w}e_F(w)v := \phi(w,v) \\ v(1) = 1. \end{cases}$$

Since $f(w)$ is continuous over $[0,1]$, we have $f(w)$ is bounded over $[0,1]$, and $e_F(w)$ is continuous and bounded over $[\varepsilon,1]$ for any $\varepsilon > 0$. Notice that

$$\lim_{w\to 0} e_F(w) = \lim_{w\to 0} \frac{wf(w)}{F(w)} = \lim_{w\to 0} \frac{wF^{(k+1)}(w)+kF^{(k)}(w)}{F^{(k)}(w)} = k,$$

where $F^{(n)}(w)$ stands for the n^{th}-order derivative of $F(w)$ and $k := \min_{n\in\mathbb{N}}\{F^{(n)}(0) > 0\}$, which is a finite integer. Thus, $e_F(w)$ is bounded by some positive number M over $(0,1]$. Let $\mathscr{D} = (0,1+\varepsilon)\times[0,1]$. For any $(w_0,v_0)\in\mathscr{D}$, choose $0<r<\min\{\varepsilon,w_0\}$, and define $\mathbf{I}\times\mathbf{U} = (w_0-r,w_0+r)\times(0,1)$, $L = \frac{M}{w_0-r}$. Then $\mathbf{I}\times\mathbf{U}\subseteq\mathscr{D}$ and contains (w_0,v_0) in its interior. Since for any $v_1,v_2\in\mathbf{U}$ and any $w\in\mathbf{I}$ we have

$$|\phi(w,v_1)-\phi(w,v_2)| = \frac{1}{w}|e_F(w)|\cdot|v_1-v_2| \le \frac{M}{w_0-r}|v_1-v_2| = L|v_1-v_2|,$$

then $\phi(w,v)$ is locally Lipschitz continuous with respect to v. Thus the Picard-Lindelöf Theorem indicates that the above ODE has a solution existing on some interval of the form (w_-,w_+) and that the solution is *unique* on that interval; moreover, since $(w,v)=(w,F(w))\in[0,1]^2$ which is bounded, then either $w_+=1$ or $v(w)\to 1$ as $w\uparrow w_+$, and either $w_-=0$ or $v(w)\to 0$ as $w\downarrow w_-$. In other words, we establish a bijection between a distribution function and its elasticity. We solve the above ODE as follows. From

$$\big(\ln v(w)\big)' = \frac{e_F(w)}{w},$$

we get the general solution:

$$\ln v(w) = \int_0^w \frac{e_F(x)}{x}dx + C.$$

Because $v(1)=1$, we get $C = -\int_0^1 \frac{e_F(x)}{x}dx$. Therefore, $v(w) = e^{-\int_w^1 \frac{e_F(x)}{x}dx}$. It is straightforward to check that as $w\downarrow 0$, we have $v(w)\to e^{-\infty}=0$.

Consider a function $e(w,a)$ satisfying $e(w,a)=e_{F_1}(w)$ when $w\le a$, and $e(w,a)=e_{F_2}(w)$ when $w>a$. Define a virtual distribution function $F(w,a) := e^{-\int_w^1 \frac{e(x,a)}{x}dx}$ and compute the corresponding $\tilde{\delta}(y,a)$ given by

$$\int_0^{\frac{y}{3}} 3e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw = \int_0^{\frac{\tilde{\delta}(y,a)y}{2}} 2e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw.$$

Clearly, $F(w,a)$ and $\tilde{\delta}(y,a)$ are both continuous. Fixed any y, *Case 1*. If $a < \frac{y}{3}$, then

$$\begin{aligned}&\int_0^a 3e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw+\int_a^{\frac{y}{3}} 3e^{-\int_w^1 \frac{e_{F_2}(x)}{x}dx}dw\\ =&\int_0^a 2e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw+\int_a^{\frac{\tilde{\delta}(y,a)y}{2}} 2e^{-\int_w^1 \frac{e_{F_2}(x)}{x}dx}dw.\end{aligned}$$

Take derivatives with respect to a on both side, and we get

$$\int_0^a e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw\big(\underbrace{-\frac{e_{F_1}(a)}{a}+\frac{e_{F_2}(a)}{a}}_{\geq 0}\big)=ye^{-\int_{\frac{\tilde{\delta}(y,a)y}{2}}^1 \frac{e_{F_2}(x)}{x}dx}\frac{d\tilde{\delta}(y,a)}{da},$$

which means $\frac{d\tilde{\delta}(y,a)}{da}\geq 0$;

Case 2. If $\frac{y}{3}\leq a<\frac{\tilde{\delta}(y,a)y}{2}$, then

$$\begin{aligned}&\int_0^{\frac{y}{3}} 3e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw\\ =&\int_0^a 2e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw+\int_a^{\frac{\tilde{\delta}(y,a)y}{2}} 2e^{-\int_w^1 \frac{e_{F_2}(x)}{x}dx}dw.\end{aligned}$$

Take derivatives with respect to a on both side, and we get

$$\begin{aligned}&\Big(\underbrace{\int_0^{\frac{y}{3}} 3e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw-\int_0^a 2e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw}_{\geq 0,\ \because a\leq\frac{\tilde{\delta}(y,a)y}{2}}\Big)\big(\underbrace{-\frac{e_{F_1}(a)}{a}+\frac{e_{F_2}(a)}{a}}_{\geq 0}\big)\\ =&ye^{-\int_{\frac{\tilde{\delta}(y,a)y}{2}}^1 \frac{e_{F_2}(x)}{x}dx}\frac{d\tilde{\delta}(y,a)}{da},\quad\Longrightarrow\quad\frac{d\tilde{\delta}(y,a)}{da}\geq 0;\end{aligned}$$

Case 3. If $a\geq\frac{\tilde{\delta}(y,a)y}{2}$, then

$$\int_0^{\frac{y}{3}} 3e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw=\int_0^{\frac{\tilde{\delta}(y,a)y}{2}} 2e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw.$$

Take derivatives with respect to a on both side, and we get

$$\begin{aligned}&\Big(\underbrace{\int_0^{\frac{y}{3}} 3e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw-\int_0^{\frac{\tilde{\delta}(y,a)y}{2}} 2e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw}_{=0}\Big)\big(\underbrace{-\frac{e_{F_1}(a)}{a}+\frac{e_{F_2}(a)}{a}}_{\geq 0}\big)\\ =&ye^{-\int_{\frac{\tilde{\delta}(y,a)y}{2}}^1 \frac{e(x,a)}{x}dx}\frac{d\tilde{\delta}(y,a)}{da},\quad\Longrightarrow\quad\frac{d\tilde{\delta}(y,a)}{da}=0.\end{aligned}$$

Thus we have $\frac{d\tilde{\delta}(y,a)}{da} \geq 0$ for all $a \in (0,1)$. Together with the continuity of $\tilde{\delta}(y,a)$, we have $\tilde{\delta}(y,1) \geq \tilde{\delta}(y,0)$, which means $\tilde{\delta}_{F_1}(y) \geq \tilde{\delta}_{F_2}(y)$.

Similarly, the corresponding $\hat{\delta}(y,a)$ is given by

$$y\big(\frac{3}{2}\hat{\delta}(y,a)-1\big)e^{-\int_{\frac{\hat{\delta}(y,a)y}{2}}^{1}\frac{e(x,a)}{x}dx} = \int_0^{\frac{\hat{\delta}(y,a)y}{2}} e^{-\int_w^1 \frac{e(x,a)}{x}dx}dw.$$

Fixed any y,

Case 1. If $a < \frac{\hat{\delta}(y,a)y}{2}$, then

$$y\big(\frac{3}{2}\hat{\delta}(y,a)-1\big)e^{-\int_{\frac{\hat{\delta}(y,a)y}{2}}^{1}\frac{e_{F_2}(x)}{x}dx}$$
$$=\int_0^a e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw+\int_a^{\frac{\hat{\delta}(y,a)y}{2}} 2e^{-\int_w^1 \frac{e_{F_2}(x)}{x}dx}dw.$$

Take derivatives with respect to a on both side, and we get

$$y\Big[1+\big(\underbrace{\frac{3\hat{\delta}(y,a)}{2}-1}_{\geq 0}\big)\frac{e_{F_2}(\frac{\hat{\delta}(y,a)y}{2})}{\hat{\delta}(y,a)}\Big]e^{-\int_{\frac{\hat{\delta}(y,a)y}{2}}^{1}\frac{e_{F_2}(x)}{x}dx}\frac{d\hat{\delta}(y,a)}{da}$$
$$=\int_0^a e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1 \frac{e_{F_2}(x)}{x}dx}dw\big(\underbrace{-\frac{e_{F_1}(a)}{a}+\frac{e_{F_2}(a)}{a}}_{\geq 0}\big),$$

which means $\frac{d\hat{\delta}(y,a)}{da} \geq 0$;

Case 2. If $a \geq \frac{\hat{\delta}(y,a)y}{2}$, then

$$y\big(\frac{3}{2}\hat{\delta}(y,a)-1\big)e^{-\int_{\frac{\hat{\delta}(y,a)y}{2}}^{a}\frac{e_{F_1}(x)}{x}dx-\int_a^1\frac{e_{F_2}(x)}{x}dx} = \int_0^{\frac{\hat{\delta}(y,a)y}{2}} e^{-\int_w^a \frac{e_{F_1}(x)}{x}dx-\int_a^1\frac{e_{F_2}(x)}{x}dx}dw.$$

Take derivatives with respect to a on both side, and we get $\frac{d\hat{\delta}(y,a)}{da} = 0$.

Thus we have $\frac{d\hat{\delta}(y,a)}{da} \geq 0$ on each part of the domain. Notice that $\hat{\delta}(y,a)$ is continuous, then we have $\hat{\delta}(y,a)$ is increasing over $[0,1]$. It follows that $\hat{\delta}(y,1) \geq \hat{\delta}(y,0)$, which means $\hat{\delta}_{F_1}(y) \geq \hat{\delta}_{F_2}(y)$. □

D.1.5 Proof of Proposition 8

Proof. We first show that $w_{\mu_0}(\delta)$ is quasiconcave over $[\underline{\delta},1]$. Pick any small $\varepsilon > 0$, we have $w_{\mu_0}(\delta+\varepsilon) - w_{\mu_0}(\delta) =$

$$y\frac{(\mu_0(\delta+\varepsilon)-\mu_0(\delta))(2-\int_{\delta'\in(\delta,1]}\delta' d\mu_0(\delta'))-(2+\mu_0(\delta))\int_{\delta'\in(\delta,\delta+\varepsilon]}\delta' d\mu_0(\delta')}{(2+\mu_0(\delta))(2+\mu_0(\delta+\varepsilon))}.$$

Since μ_0 is right-continuous, we have $\mu_0(\delta^+) := \lim_{\delta' \downarrow \delta} \mu_0(\delta') = \mu_0(\delta)$, that is, for infinitesimal ε, $w_{\mu_0}(\delta+\varepsilon) - w_{\mu_0}(\delta)$ is an infinitesimal of same order. Apply Lebesgue version of integration by parts, and we have

$$\int_{\delta' \in (\delta, \delta+\varepsilon]} \delta' d\mu_0(\delta') = (\delta+\varepsilon)\mu_0(\delta+\varepsilon) - \delta\mu_0(\delta) - \int_{\delta' \in (\delta, \delta+\varepsilon]} \mu_0(\delta'^-) d\delta',$$
$$\int_{\delta' \in (\delta, 1]} \delta' d\mu_0(\delta') = 1 - \delta\mu_0(\delta) - \int_{\delta' \in (\delta, 1]} \mu_0(\delta'^-) d\delta',$$

where $\mu_0(\delta^-) := \lim_{\delta' \uparrow \delta} \mu_0(\delta)$. Notice that we have $\delta'^- \in (\delta, \delta+\varepsilon)$ for any $\delta' \in (\delta, \delta+\varepsilon]$, then $w_{\mu_0}(\delta'^-)$ is only infinitesimal different from $w_{\mu_0}(\delta)$. Omit all higher-order infinitesimal terms, and we have

$$\int_{\delta' \in (\delta, \delta+\varepsilon]} \delta' d\mu_0(\delta') \simeq (\delta+\varepsilon)\mu_0(\delta+\varepsilon) - \delta\mu_0(\delta) - \mu_0(\delta)\varepsilon$$
$$\simeq (\delta+\varepsilon)\mu_0(\delta+\varepsilon) - \delta\mu_0(\delta) - \mu_0(\delta+\varepsilon)\varepsilon = \delta(\mu_0(\delta+\varepsilon) - \mu_0(\delta)),$$
$$(\mu_0(\delta+\varepsilon) - \mu_0(\delta)) \int_{\delta' \in (\delta, 1]} \mu_0(\delta'^-) d\delta' \simeq (\mu_0(\delta+\varepsilon) - \mu_0(\delta)) \int_{\delta' \in (\delta, 1]} \mu_0(\delta') d\delta'.$$

Thus, we have

$$w_{\mu_0}(\delta+\varepsilon) - w_{\mu_0}(\delta) \simeq \frac{y(\mu_0(\delta+\varepsilon) - \mu_0(\delta))}{(2+\mu_0(\delta))(2+\mu_0(\delta+\varepsilon))} \underbrace{\left[1 - 2\delta + \int_{\delta' \in (\delta, 1]} \mu_0(\delta') d\delta'\right]}_{:=\tau(\delta)}.$$

Since $\tau(\delta)$ is decreasing with respect to δ, $\tau(\frac{1}{2}) \geq 0$ and $\tau(1) < 0$, then there exists $[a,b] \subseteq [\frac{1}{2}, 1)$, where $a \leq b$, such that $[a,b] = \arg\max_{\delta'} w_{\mu_0}(\delta')$, $w_{\mu_0}(\delta)$ is increasing over $[\frac{1}{2}, a)$ and is decreasing over $(b, 1]$. It follows that $w_{\mu_0}(\delta)$ is quasiconcave.

Second, we prove that If $w_{\mu_0}(\delta)$ and $\bar{w}(\delta)$ intersect, then the intersection is always on the non-increasing side of $w_{\mu_0}(\delta)$.Pick any small $\varepsilon > 0$, we have

$$\int_{\delta' \in [0, \delta]} \big(y - 3w_{\mu_0}(\delta)\big) d\mu_0(\delta') + \int_{\delta' \in (\delta, 1]} \big(\delta' y - 2w_{\mu_0}(\delta)\big) d\mu_0(\delta') = 0$$
$$\int_{\delta' \in [0, \delta+\varepsilon]} \big(y - 3w_{\mu_0}(\delta+\varepsilon)\big) d\mu_0(\delta') + \int_{\delta' \in (\delta+\varepsilon, 1]} \big(\delta' y - 2w_{\mu_0}(\delta+\varepsilon)\big) d\mu_0(\delta') = 0.$$

It follows that

$$3w_{\mu_0}(\delta) = y + \int_{\delta' \in (\delta, 1]} \big(\delta' y + w_{\mu_0}(\delta) - y\big) d\mu_0(\delta')$$
$$3w_{\mu_0}(\delta+\varepsilon) = y + \int_{\delta' \in (\delta+\varepsilon, 1]} \big(\delta' y + w_{\mu_0}(\delta+\varepsilon) - y\big) d\mu_0(\delta').$$

Then we have $3w_{\mu_0}(\delta+\varepsilon)-3w_{\mu_0}(\delta)=$

$$\int_{\delta'\in(\delta+\varepsilon,1]}\big(\delta' y+w_{\mu_0}(\delta+\varepsilon)-y\big)d\mu_0(\delta')-\int_{\delta'\in(\delta,1]}\big(\delta' y+w_{\mu_0}(\delta)-y\big)d\mu_0(\delta')$$
$$=\int_{\delta'\in(\delta+\varepsilon,1]}\big(w_{\mu_0}(\delta+\varepsilon)-w_{\mu_0}(\delta)\big)d\mu_0(\delta')-\int_{\delta'\in(\delta,\delta+\varepsilon]}\big(\delta' y+w_{\mu_0}(\delta)-y\big)d\mu_0(\delta')$$
$$\simeq\big(w_{\mu_0}(\delta+\varepsilon)-w_{\mu_0}(\delta)\big)\big(1-\mu_0(\delta+\varepsilon)\big)-\big(\delta y+w_{\mu_0}(\delta)-y\big)\big(\mu_0(\delta+\varepsilon)-\mu_0(\delta)\big),$$

which means

$$w_{\mu_0}(\delta+\varepsilon)-w_{\mu_0}(\delta)=-\frac{\mu_0(\delta+\varepsilon)-\mu_0(\delta)}{2+\mu_0(\delta+\varepsilon)}\big(\delta y+w_{\mu_0}(\delta)-y\big).$$

When $w_{\mu_0}(\delta)$ and $\bar{w}(\delta)$ intersect at some δ'', we have $w_{\mu_0}(\delta'')=\bar{w}(\delta'')$, then $\delta'' y+w_{\mu_0}(\delta'')-y=\delta'' y+\bar{w}(\delta'')-y\geq 0$, then $w_{\mu_0}(\delta''+\varepsilon)-w_{\mu_0}(\delta'')\leq 0$. We conclude that $w_{\mu_0}(\delta)$ intersects with $\bar{w}(\delta)$ on the non-increasing side of $w_{\mu_0}(\delta)$.

Third, we establish the existence of $\delta^*\in[\underline{\delta},\bar{\delta}]$ such that $w_{\mu_0}(\delta^*)=\bar{w}(\delta^*)$, where $\bar{\delta}$ satisfies $\bar{w}(\bar{\delta})=\frac{y}{3}$, and $\underline{\delta}$ satisfies $w_{\mu_0}(\underline{\delta})=\frac{y}{2}$. Consider a correspondence $\Gamma(\delta):[a,1]\mapsto 2^{[a,1]}$, which is defined by $\Gamma(\delta)=\{\delta'\in[a,1]\mid w_{\mu_0}(\delta)\leq\bar{w}(\delta')\leq w_{\mu_0}(\delta^-)\}$, where $w_{\mu_0}(\delta^-)=\lim_{\delta'\uparrow\delta}w_{\mu_0}(\delta')$; in case $w_{\mu_0}(a)>w_{\mu_0}(a^-)$, $\Gamma(a):=\{\delta'\in[a,1]\mid w_{\mu_0}(a)=\bar{w}(\delta')\}$. Obviously, $[a,1]$ is a non-empty, convex and compact subset of Euclidean space $\mathbb{R}$. We have showed that $\bar{w}(\delta)$ is continuous and decreasing over $[a,1]$, $w_{\mu_0}(\delta)$ is right-continuous with left limits and non-increasing over $[a,1]$, $w_{\mu_0}(a)\leq\bar{w}(a)$ and $w_{\mu_0}(\bar{\delta})\geq\frac{y}{3}=\bar{w}(\bar{\delta})$, where $[a,b]=\arg\max_{\delta'}w_{\mu_0}(\delta')$. Thus we can easily check that for any $\delta\in[a,1]$, $\Gamma(\delta)$ is non-empty and convex. Moreover, Γ has a closed graph $Gr(\Gamma):=\{(s,t)\in[a,1]\times 2^{[a,1]}\mid t\in\Gamma(s)\}$.[2] Thus, by Kakutani's fixed-point theorem, Γ has a fixed point, that is, there exists δ such that $\delta\in\Gamma(\delta)$. Immediately, any fixed point satisfies $\delta\leq\bar{\delta}$, since $\frac{y}{3}\leq w_{\mu_0}(\delta)\leq\bar{w}(\delta)$.

[2] To prove this, consider the complementary set of $Gr(\Gamma)$ in $[a,1]\times 2^{[a,1]}$, which is given by $Gr^c(\Gamma):=\{(s,t)\in[a,1]\times 2^{[a,1]}\mid w_{\mu_0}(s)>\bar{w}(t),\text{ or }\bar{w}(t)>w_{\mu_0}(s^-)\}$. Pick any $(s,t)\in Gr^c(\Gamma)$, if $w_{\mu_0}(s)>\bar{w}(t)$, choose $0<\varepsilon<w_{\mu_0}(s)-\bar{w}(t)$, then $\exists r_1>0$ such that for any $t'\in(t-r_1,t+r_1)$ we have $\bar{w}(t')\in(\bar{w}(t)-\frac{\varepsilon}{3},\bar{w}(t)+\frac{\varepsilon}{3})$, and $\exists r_2>0$ such that for any $s'\in(s-r_2,s+r_2)$ we have $w_{\mu_0}(s')>w_{\mu_0}(s)-\frac{\varepsilon}{3}$. Since $w_{\mu_0}(s')-\bar{w}(t')>w_{\mu_0}(s)-\frac{\varepsilon}{3}-\bar{w}(t)-\frac{\varepsilon}{3}>\frac{\varepsilon}{3}>0$, then for all (s',t') such that $|s'-s|<r_2$ and $|t'-t|<r_1$, we have $(s',t')\in Gr^c(\Gamma)$. If $\bar{w}(t)>w_{\mu_0}(s^-)$, choose $0<\varepsilon<\bar{w}(t)-w_{\mu_0}(s^-)$, then $\exists r_1>0$ such that for all $t'\in(t-r_1,t+r_1)$ we have $\bar{w}(t')\in(\bar{w}(t)-\frac{\varepsilon}{3},\bar{w}(t)+\frac{\varepsilon}{3})$. The fact that left limit $w_{\mu_0}(s^-)$ exists and $w_{\mu_0}(\cdot)$ is non-increasing implies that $\exists r_2>0$ such that for all $s'\in(s-r_2,s)$ we have $0\leq w_{\mu_0}(s')-w_{\mu_0}(s^-)<\frac{\varepsilon}{3}$. It follows that for all $s''\in(s',s)$ we have $w_{\mu_0}(s''^-)\leq w_{\mu_0}(s')<w_{\mu_0}(s^-)+\frac{\varepsilon}{3}$. For all $s''\in(s,s+r_2)$, we have $w_{\mu_0}(s''^-)\leq w_{\mu_0}(s)\leq w_{\mu_0}(s^-)+\frac{\varepsilon}{3}$. Then $\bar{w}(t')-w_{\mu_0}(s''^-)>\bar{w}(t)-\frac{\varepsilon}{3}-w_{\mu_0}(s^-)-\frac{\varepsilon}{3}>\frac{\varepsilon}{3}>0$ for all (s'',t') such that $s''\in(s',s+r_2)$ and $|t'-t|<r_1$, which means $(s'',t')\in Gr^c(\Gamma)$. Thus, (s,t) is in the interior of $Gr^c(\Gamma)$, which means $Gr^c(\Gamma)$ is an open set. As a result, $Gr(\Gamma)$ is closed.

Let $\mathscr{F} := \{\delta \in [\underline{\delta}, \bar{\delta}] \mid \delta \in \Gamma(\delta)\}$ collect all the fixed points. We have $\mathscr{F}$ is a compact subset.[3] If there exists $\delta \in \mathscr{F}$ such that $w_{\mu_0}(\delta) = \bar{w}(\delta)$, then we find such δ^*. Suppose for all $\delta \in \mathscr{F}$ we have $w_{\mu_0}(\delta) < \bar{w}(\delta)$. Pick any $\delta_1 \in \mathscr{F}$, since $w_{\mu_0}(\delta_1) < \bar{w}(\delta_1)$, then we can apply the fixed-point theorem on domain $[\delta_1, \bar{\delta}]$ and find $\delta_1 < \delta_2 \in \mathscr{F}$ satisfying $w_{\mu_0}(\delta_2) < \bar{w}(\delta_2)$. Repeat this process and we can find a strictly increasing sequence $\{\delta_n\}_{n=1}^{\infty} \in \mathscr{F}$ such that $\delta_n \in [\underline{\delta}, \bar{\delta}]$ and $w_{\mu_0}(\delta_n) < \bar{w}(\delta_n)$ for all n. Define $\phi(\delta_n) := \bar{w}(\delta_n) - w_{\mu_0}(\delta_n) > 0$. Since $\phi(\delta_n)$ is bounded, it converges. If $\{\phi(\delta_n)\}$ has a unique limit which is 0, then $\lim_{n\to\infty}\phi(\delta_n) = 0$ and $\lim_{n\to\infty}\delta_n = \check{\delta}$. Since $\mathscr{F}$ is compact, we have $\check{\delta} \in \mathscr{F}$ and $w_{\mu_0}(\check{\delta}) = \bar{w}(\check{\delta})$, contradiction. If there exists a subsequence $\{\phi(\delta_{k_n})\}$ such that $\lim_{n\to\infty}\phi(\delta_{k_n}) = \varepsilon > 0$, then $\exists N_0$ such that for all $n \geq N_0$ we have $\phi(\delta_{k_n}) > \frac{\varepsilon}{2}$, which means $w_{\mu_0}(\delta_{k_n}^-) - w_{\mu_0}(\delta_{k_n}) \geq \bar{w}(\delta_{k_n}) - w_{\mu_0}(\delta_{k_n}) > \frac{\varepsilon}{2}$. It follows that

$$\frac{y}{6} = \frac{y}{2} - \frac{y}{3} \geq \max_{\delta} w_{\mu_0}(\delta) - \min_{\delta} w_{\mu_0}(\delta) > \sum_{n=1}^{\infty}\left(w_{\mu_0}(\delta_n^-) - w_{\mu_0}(\delta_n)\right)$$
$$> \sum_{n=N_0}^{\infty}\left(w_{\mu_0}(\delta_n^-) - w_{\mu_0}(\delta_n)\right) > \sum_{n=N_0}^{\infty}\frac{\varepsilon}{2} = +\infty\cdot\frac{\varepsilon}{2} > \frac{y}{6},$$

contradiction. Thus we conclude there exists $\delta^* \in [\underline{\delta}, \bar{\delta}]$ such that $w_{\mu_0}(\delta^*) = \bar{w}(\delta^*)$.

With a slight abuse of notation, let $\delta^* = \min\{\delta \mid w_{\mu_0}(\delta) = \bar{w}(\delta)\}$.[4] The last step is to show that, given any message inducing the posterior belief μ_0, there exists a unique pure-strategy SPE-NCF, where all agents with $w \leq w^*(\mu_0) := w_{\mu_0}(\delta^*)$ participate at time 0, and principal chooses $C_1^{\delta y}$ if $\delta > \delta^*$ and C_2^y if $\delta \leq \delta^*$.

Let $\{\delta \mid w_{\mu_0}(\delta) = \bar{w}(\delta)\} = \{\delta_1, \delta_2, \ldots, \delta_n, \ldots\}$, with $\delta_1 < \delta_2 < \ldots < \delta_n < \ldots$. Pick any $\delta \in (\delta_t, \delta_{t+1})$, if $w_{\mu_0}(\delta) < \bar{w}(\delta)$, then given $I_1^* = \{i \in I \mid w(i) \leq w_{\mu_0}(\delta)\}$, the group leader will choose C_2^y if δ is realized, then the agent with outside option $w_{\mu_0}(\delta)$ will get negative expected utility, thus he prefers to quit at time 0. This process will not stop until $I_1^* = \{i \in I \mid w(i) \leq w_{\mu_0}(\delta_{t+1})\}$. If $w_{\mu_0}(\delta) > \bar{w}(\delta)$, given $I_1^* = \{i \in I \mid w(i) \leq w_{\mu_0}(\delta)\}$, the group leader will choose $C_1^{\delta y}$ at $\delta - \varepsilon_1$, which will make participation at time 0 profitable for agents with outside option $w_{\mu_0}(\delta) + \varepsilon_2$. This process will keep going until $I_1^* = \{i \in I \mid w(i) \leq w_{\mu_0}(\delta_t)\}$. Thus only $\{\delta_1, \delta_2, \ldots, \delta_n, \ldots\}$ are SPE. Take any $\delta_t < \delta_{t+1}$, since there is no coordination failure among agents, then agents whose outside options belong to $[w_{\mu_0}(\delta_{t+1}), w_{\mu_0}(\delta_t)]$ will

[3] Define $\mathscr{I} := \{(s,t) \in [a,1] \times 2^{[a,1]} \mid s = t\}$, which is closed. Then $\mathscr{F}$ is also closed because $\mathscr{F} = Gr(\Gamma) \cap \mathscr{I}$. Notice that $\mathscr{F}$ is bounded, then it is compact.

[4] We must have that $w_{\mu_0}(\delta)$ is continuous at δ^*; otherwise $w_{\mu_0}(\delta^{*-}) > w_{\mu_0}(\delta^*) = \bar{w}(\delta^*)$, which means there exists $\delta' \in [a, \delta^*)$ such that $w_{\mu_0}(\delta') = \bar{w}(\delta')$, contradicting the definition of δ^*.

jointly participate at time 0 because they can get utility gains in SPE-δ_t compared with no participation in SPE-δ_{t+1}. Thus, the only subgame perfect equilibrium that survives from elimination of all possible coordination failure is SPE-δ^*, where the participating subset of outside options is the largest among all subgame perfect equilibriums. □

D.1.6 Proof of Theorem 11

Proof. We first show two properties of $w_{\mu_0}(\delta,\delta^T)$.

(i) Fixed $\forall\delta^T > \frac{2}{3}$, $w_{\mu_0}(\delta,\delta^T)$ is quasi-concave with respect to δ over $[\underline{\delta},1]$. To prove this, Pick any small $\varepsilon > 0$, we have $w_{\mu_0}(\delta+\varepsilon,\delta^T) - w_{\mu_0}(\delta,\delta^T) =$

$$\frac{y}{(2\mu_0(\delta^T)+\mu_0(\delta))(2\mu_0(\delta^T)+\mu_0(\delta+\varepsilon))}\Big[\big(\mu_0(\delta+\varepsilon)-\mu_0(\delta)\big)\Big(2\mu_0(\delta^T) - \int_{\delta'\in(\delta,\delta^T]}\delta' d\mu_0(\delta')\Big) - \big(2\mu_0(\delta^T)+\mu_0(\delta)\big)\int_{\delta'\in(\delta,\delta+\varepsilon]}\delta' d\mu_0(\delta')\Big].$$

Through exactly the same argument as in the previous observation where $\delta^T = 1$, we omit all higher-order infinitesimal terms, and get $w_{\mu_0}(\delta+\varepsilon,\delta^T) - w_{\mu_0}(\delta,\delta^T) \simeq$

$$\frac{y(\mu_0(\delta+\varepsilon)-\mu_0(\delta))\overbrace{\big[(2-\delta^T)\mu_0(\delta^T) - 2\mu_0(\delta^T)\delta + \int_{\delta'\in(\delta,\delta^T]}\mu_0(\delta')d\delta'\big]}^{:=\tau(\delta)}}{(2\mu_0(\delta^T)+\mu_0(\delta))(2\mu_0(\delta^T)+\mu_0(\delta+\varepsilon))}.$$

Since $\tau(\delta)$ is decreasing with respect to δ, $\tau(\frac{\delta^T}{2}) \geq 0$ and $\tau(\delta^T) = (2-3\delta^T)\mu_0(\delta^T) < 0$, then there exists $[a,b] \subseteq [\frac{1}{2},1)$, where $a \leq b$, such that $[a,b] = \arg\max_{\delta'} w_{\mu_0}(\delta',\delta^T)$, $w_{\mu_0}(\delta,\delta^T)$ is increasing over $[\frac{1}{2},a)$ and decreasing over $(b,1]$. It follows that $w_{\mu_0}(\delta,\delta^T)$ is quasiconcave.

(ii) Fixed $\forall\delta$, $w_{\mu_0}(\delta,\delta^T)$ is non-decreasing in δ^T over $(\frac{2}{3},1]$. To prove this, Pick any small $\varepsilon > 0$, we have $w_{\mu_0}(\delta,\delta^T+\varepsilon) - w_{\mu_0}(\delta,\delta^T) =$

$$\frac{y}{(2\mu_0(\delta^T)+\mu_0(\delta))(2\mu_0(\delta^T+\varepsilon)+\mu_0(\delta))}\Big[\big(2\mu_0(\delta^T)+\mu_0(\delta)\big)\int_{\delta'\in(\delta^T,\delta^T+\varepsilon]}\delta' d\mu_0(\delta') - \big(\mu_0(\delta^T+\varepsilon)-\mu_0(\delta^T)\big)\Big(2\mu_0(\delta)+2\int_{\delta'\in(\delta,\delta^T]}\delta' d\mu_0(\delta')\Big)\Big].$$

Through a similar argument we can omit all higher-order infinitesimal terms, and get

$$w_{\mu_0}(\delta,\delta^T+\varepsilon) - w_{\mu_0}(\delta,\delta^T) \simeq \frac{y\big(\mu_0(\delta^T+\varepsilon)-\mu_0(\delta^T)\big)}{(2\mu_0(\delta^T)+\mu_0(\delta))(2\mu_0(\delta^T+\varepsilon)+\mu_0(\delta))} \times \underbrace{\Big[\big(\delta^T-2+2\delta\big)\mu_0(\delta)+2\int_{\delta'\in(\delta,\delta^T]}\mu_0(\delta')d\delta'\Big]}_{:=\tau(\delta^T)}.$$

Since $\tau(\delta^T) \geq (\delta^T - 2 + 2\delta)\mu_0(\delta) + 2(\delta^T - \delta)\mu_0(\delta) = (3\delta^T - 2)\mu_0(\delta) \geq 0$, we conclude that $w_{\mu_0}(\delta, \delta^T)$ is non-decreasing in δ^T when $\delta^T > \frac{2}{3}$.

With these properties we can prove the following claim.

Claim. In the upper revealing truncation of μ_0 at δ^T with $\mu_0'(\delta) = \frac{\mu_0(\delta)}{\mu_0(\delta^T)}$,
(i) If $\delta^T = \hat{\delta}$, then $w^*(\mu_0') = \frac{y}{3}$;
(ii) If $\delta^* > \hat{\delta}$, then $w^*(\mu_0') = \frac{y}{3}$ for $\delta^T \leq \delta^*$.

Proof of the Claim. (i) Notice that when $\delta^T = \hat{\delta}$ we have $w_{\mu_0'}(\delta) < \frac{\hat{\delta}y}{2} = \bar{w}(\hat{\delta}) < \bar{w}(\delta)$ for all $\delta < \hat{\delta}$, and $w_{\mu_0'}(\delta) \equiv \frac{y}{3}$ for all $\delta \geq \hat{\delta}$, then $w_{\mu_0'}(\delta) < \bar{w}(\delta)$ for all $\delta \in [\underline{\delta}, \bar{\delta})$. Thus, we have $w^*(\mu_0') = \frac{y}{3}$.

(ii) By definition of δ^*, for all $\delta < \delta^*$ we have $w_{\mu_0}(\delta) < \bar{w}(\delta)$. If $\delta^T > \frac{2}{3}$, we have showed in the previous observation that $w_{\mu_0}(\delta, \delta^T)$ is non-decreasing in δ^T. Thus, fixed $\forall \delta^T \leq \delta^*$, we have $w_{\mu_0'}(\delta) = w_{\mu_0}(\delta, \delta^T) \leq w_{\mu_0}(\delta, 1) = w_{\mu_0}(\delta) < \bar{w}(\delta)$ for all $\delta < \delta^T$, and $w_{\mu_0'}(\delta) \equiv \frac{y}{3}$ for all $\delta \geq \delta^T$, which means $w_{\mu_0'}(\delta) < \bar{w}(\delta)$ for all $\delta \in [\underline{\delta}, \bar{\delta})$. If $\delta^T \leq \frac{2}{3}$, we have $w_{\mu_0'}(\delta) \leq \frac{y}{3} < \bar{w}(\delta)$ for all $\delta \in [\underline{\delta}, \bar{\delta})$. Thus, we have $w^*(\mu_0') = \frac{y}{3}$. □

Next we will study several properties of the optimal information structure.

Property 1. Optimal information structure always completely separates any $\delta \geq \tilde{\delta}$, that is, for all $\delta \geq \tilde{\delta}$, any $m_{0|\delta} \in M$ that includes δ as a possible outcome should induce a posterior belief $\mu_0[m_{0|\delta}] \in \Delta([0,1])$ such that $\mathbf{Pr}(\delta \mid m_{0|\delta}) = 1$.

Proof of the property. Pick $\forall \mu_0 \in \Delta([0,1])$ induced by some m_0, and in equilibrium $I_1^* = \{i \in I \mid w(i) \leq w^*(\mu_0)\}$.

Case 1. If $w^*(\mu_0) \geq \frac{\hat{\delta}y}{2}$, we have $\delta^* \in (\underline{\delta}, \hat{\delta})$. The group's total expected utility conditional on m_0 is given by $\mathbb{E}_{\mu_0}[U_{p|0}] =$

$$\int_0^{\delta^*} \Big(\int_0^{w^*(\mu_0)} (y - 3w) dF(w)\Big) d\mu_0(\delta) + \int_{\delta^*}^1 \Big(\int_0^{w^*(\mu_0)} (\delta y - 2w) dF(w)\Big) d\mu_0(\delta).$$

Consider the upper revealing truncation of μ_0 at $\hat{\delta}$, that is, $\delta^T = \hat{\delta}$. We have proved that $w^*(\mu_0') = \frac{y}{3}$ and group leader chooses default option at time 1. Thus, the group's total expected utility under the new message set conditional on μ_0 is given by

$$\mathbb{E}'_{\mu_0}[U_{p|0}] = \int_0^{\delta^*} \Big(\int_0^{\frac{y}{3}} (y - 3w) dF(w)\Big) d\mu_0(\delta) + \int_{\delta^*}^{\hat{\delta}} \Big(\int_0^{\frac{y}{3}} (y - 3w) dF(w)\Big) d\mu_0(\delta)$$
$$+ \int_{\hat{\delta}}^1 \Big(\int_0^{\frac{\delta y}{2}} (\delta y - 2w) dF(w)\Big) d\mu_0(\delta) \geq \mathbb{E}_{\mu_0}[U_{p|0}].$$

Case 2. If $\bar{w}(\tilde{\delta}) < w^*(\mu_0) < \frac{\hat{\delta}y}{2}$, we have $\delta^* \in (\hat{\delta}, \tilde{\delta})$. The group's total expected utility conditional on m_0 is given as before. Consider the upper revealing truncation of μ_0 at δ^*, that is, $\delta^T = \delta^*$. We have proved that $w^*(\mu_0') = \frac{y}{3}$ and group leader chooses default option at time 1. Thus, the group's total expected utility under the new message set conditional on μ_0 is given by $\mathbb{E}'_{\mu_0}[U_{p|0}] =$

$$\int_0^{\delta^*}\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big)d\mu_0(\delta)+\int_{\delta^*}^1\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)\Big)d\mu_0(\delta) \geq \mathbb{E}_{\mu_0}[U_{p|0}].$$

Case 3. If $w^*(\mu_0) \leq \bar{w}(\tilde{\delta})$, we have $\delta^* \geq \tilde{\delta}$. The group's total expected utility conditional on m_0 is given by

$$\begin{aligned}\mathbb{E}_{\mu_0}[U_{p|0}] &= \int_0^{\tilde{\delta}}\Big(\int_0^{w^*(\mu_0)}(y-3w)dF(w)\Big)d\mu_0(\delta)\\ &+\int_{\tilde{\delta}}^{\delta^*}\Big(\int_0^{w^*(\mu_0)}(y-3w)dF(w)\Big)d\mu_0(\delta)+\int_{\delta^*}^1\Big(\int_0^{w^*(\mu_0)}(\delta y-2w)dF(w)\Big)d\mu_0(\delta).\end{aligned}$$

Consider the upper revealing truncation of μ_0 at $\tilde{\delta}$, that is, $\delta^T = \tilde{\delta} \leq \delta^*$. We have proved that $w^*(\mu_0') = \frac{y}{3}$ and group leader chooses default option at time 1. Thus, the group's total expected utility under the new message set conditional on μ_0 is given by

$$\begin{aligned}\mathbb{E}'_{\mu_0}[U_{p|0}] &= \int_0^{\tilde{\delta}}\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big)d\mu_0(\delta)+\int_{\tilde{\delta}}^{\delta^*}\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)\Big)d\mu_0(\delta)\\ &+\int_{\delta^*}^1\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)\Big)d\mu_0(\delta) \geq \mathbb{E}_{\mu_0}[U_{p|0}],\end{aligned}$$

since for all $\delta \in (\tilde{\delta}, \delta^*)$, we have

$$\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w) > \int_0^{\frac{y}{3}}(y-3w)dF(w) > \int_0^{w^*(\mu_0)}(y-3w)dF(w).$$

We can see that in none of the three cases should we pool any $\delta > \tilde{\delta}$ with the other realizations. Thus in the optimal information structure, the group leader always completely separates any $\delta > \tilde{\delta}$. The resulting information structure will improve the total expected utility and possess the following three properties: (i) $\forall\delta > \tilde{\delta}$ is completely revealed; (ii) On receiving $\forall m_0$ that pools some realizations of δ, $w^*(\mu_0) = \frac{y}{3}$ and the first-best outcome is achieved; (iii) If $\delta \leq \tilde{\delta}$ and is revealed by some m_0, then the first-best outcome under that realization is the default option.

Property 2. For any μ_0, μ_0' induced by $m_0, m_0' \in M_0$, let μ_0'' be the posterior belief induced by m_0'' which pools m_0 and m_0', then we have $w^*(\mu_0'') \leq \max\{w^*(\mu_0), w^*(\mu_0')\}$.

Proof of the property. Without loss of generality, we assume $w^*(\mu_0) \geq w^*(\mu_0')$, which means the threshold under μ_0, denoted by δ^*, is no larger than the threshold under μ_0', denoted by δ'^*. By definition, we have $\mu_0'' = \frac{\pi(m_0)}{\pi(m_0)+\pi(m_0')}\mu_0 + \frac{\pi(m_0')}{\pi(m_0)+\pi(m_0')}\mu_0'$. Since

$$\begin{aligned}
&\int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0''}(\delta)\big)d\mu_0''(\delta') + \int_{\delta' \in (\delta,1]} \big(\delta' y - 2w_{\mu_0''}(\delta)\big)d\mu_0''(\delta') = 0 \\
\Leftrightarrow\ & \frac{\pi(m_0)}{\pi(m_0)+\pi(m_0')}\Big[\int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0''}(\delta)\big)d\mu_0(\delta') + \int_{\delta' \in (\delta,1]} \big(\delta' y - 2w_{\mu_0''}(\delta)\big)d\mu_0(\delta')\Big] \\
&+ \frac{\pi(m_0')}{\pi(m_0)+\pi(m_0')}\Big[\int_{\delta' \in [0,\delta]} \big(y - 3w_{\mu_0''}(\delta)\big)d\mu_0'(\delta') + \int_{\delta' \in (\delta,1]} \big(\delta' y - 2w_{\mu_0''}(\delta)\big)d\mu_0'(\delta')\Big] = 0,
\end{aligned}$$

we have

$$w_{\mu_0''}(\delta) = \frac{\pi(m_0)[\mu_0(\delta)+2]\cdot w_{\mu_0}(\delta)}{\pi(m_0)[\mu_0(\delta)+2]+\pi(m_0')[\mu_0'(\delta)+2]} + \frac{\pi(m_0')[\mu_0'(\delta)+2]\cdot w_{\mu_0'}(\delta)}{\pi(m_0)[\mu_0(\delta)+2]+\pi(m_0')[\mu_0'(\delta)+2]}.$$

Notice that for any $\delta < \delta^* \leq \delta'^*$, we have $w_{\mu_0}(\delta) < \bar{w}(\delta)$ and $w_{\mu_0'}(\delta) < \bar{w}(\delta)$, then $w_{\mu_0''}(\delta) \leq \max\{w_{\mu_0}(\delta), w_{\mu_0'}(\delta)\} < \bar{w}(\delta)$ for all $\delta < \delta^*$, which means the threshold of μ_0'', denoted by δ''^*, satisfies $\delta''^* \geq \delta^*$. It follows that

$$w^*(\mu_0'') \leq w^*(\mu_0) = \max\{w^*(\mu_0), w^*(\mu_0')\}.$$

We have proved that after upper revealing truncation, if a message $m_0 \in M_0$ pools some realizations of δ, then the possible outcomes are restricted in $[0, \tilde{\delta}]$, that is, the induced posterior belief μ_0 satisfying $\mu_0(\tilde{\delta}) = 1$. Pick any pair of such pooling messages m_0, m_0', we can prove that mixing m_0 and m_0' will not affect the total expected utility. Let μ_0, μ_0' be the induced posterior belief, which satisfy $w^*(\mu_0) = w^*(\mu_0') = \frac{y}{3}$. Replace m_0 and m_0' by a new message $m_0'' := \big(m_0, \frac{\pi(m_0)}{\pi(m_0)+\pi(m_0')}; m_0', \frac{\pi(m_0')}{\pi(m_0)+\pi(m_0')}\big)$, which induces the posterior belief μ_0''. From Property 2 we have

$$\frac{y}{3} \leq w^*(\mu_0'') \leq \max\{w^*(\mu_0), w^*(\mu_0')\} = \frac{y}{3},$$

then $w^*(\mu_0'') = \frac{y}{3}$, implying that the default option is still implemented in equilibrium. We can repeat this process until we pool all such messages while the total expected utility remains unaffected. On the other hand, for any revealing message $m_{0|\delta}$ such that $\delta \leq \hat{\delta}$, the induced posterior belief $\mu_{0|\delta}$ satisfies $\mathbf{Pr}(\delta \mid m_{0|\delta}) = 1$, and thus we have $w^*(\mu_{0|\delta}) = \frac{y}{3}$. As a result we can pool all these revealing messages with the pooling message, and still implement the default option.

Now the optimal information structure is a collection of messages consisting of a unique pooling message m_0^p and a subset of revealing messages $\{m_{0|\delta}\}$. The optimal revealing rule denoted by $\sigma(\delta) : [0,1] \mapsto \Delta(M_0)$, satisfies $\sigma(\delta)[m_0^p] = 1$ for all $\delta \leq \hat{\delta}$ and $\sigma(\delta)[m_{0|\delta}] = 1$

for all $\delta > \tilde{\delta}$. The remaining problem is to determine the revealing rule for $\delta \in (\hat{\delta}, \tilde{\delta}]$.

Property 3. Fixed $\forall \delta \in (\hat{\delta}, \tilde{\delta}]$, if $\sigma(\delta)[m_0^p] > 0$, then $\sigma(\delta')[m_0^p] = 1$ for all $\delta' \in (\hat{\delta}, \delta)$.
Proof of the property. Suppose, to the contrary, that there exist $\hat{\delta} < \delta' < \delta \leq \tilde{\delta}$ such that $\sigma(\delta)[m_0^p] > 0$ and $\sigma(\delta')[m_{0|\delta'}] > 0$. Since δ follows a continuous distribution, the probability measure of any countable subset is equal to 0. If we only have countable number of such δ', we can set $\sigma(\delta')[m_{0|\delta'}] = 1$, that is, mix $m_{0|\delta'}$ with m_0^p, and still have $w^*(\mu_0^p) = \frac{y}{3}$, where μ_0^p is the posterior belief induced by m_0^p. Thus, we only need to consider the case where $\exists r_1, r_2 > 0$ such that $\sigma(\delta_1)[m_0^p] > 0$ for all $\delta_1 \in \mathscr{B}_{r_1}(\delta)$, $\sigma(\delta_2)[m_{0|\delta_2}] > 0$ for all $\delta_2 \in \mathscr{B}_{r_2}(\delta')$, and $\mathscr{B}_{r_1}(\delta) \cap \mathscr{B}_{r_2}(\delta') = \emptyset$. Let $\pi \in \Delta(M_0)$ be the probability measure over M_0 induced by the revealing rule σ. Since we have

$$\varepsilon := \min\left\{\pi(m_0^p)\int_{\delta_1 \in \mathscr{B}_{r_1}(\delta)} d\mu_0^p(\delta_1), \int_{\delta_2 \in \mathscr{B}_{r_2}(\delta')} d\pi(m_{0|\delta_2})\right\} > 0,$$

there exist $C_\delta \subseteq \mathscr{B}_{r_1}(\delta)$ and $C_{\delta'} \subseteq \mathscr{B}_{r_2}(\delta')$ such that

$$\pi(m_0^p)\int_{\delta_1 \in C_\delta} d\mu_0^p(\delta_1) = \int_{\delta_2 \in C_{\delta'}} d\pi(m_{0|\delta_2}) = \varepsilon.$$

Modify the original information structure by setting $\sigma(\delta_1)[m_{0|\delta_1}] = 1$ for all $\delta_1 \in C_\delta$, and $\sigma(\delta_2)[m_0^p] = 1$ for all $\delta_2 \in C_{\delta'}$. Next, we prove that under the new $\tilde{\mu}_0^p$ we still have $w^*(\tilde{\mu}_0^p) = \frac{y}{3}$ while the total expected utility increases. Because $w^*(\mu_0^p) = \frac{y}{3}$, for any $\check{\delta} \leq \tilde{\delta}$ we have

$$\chi_{\mu_0^p}(\check{\delta}) := \int_{\delta' \in [0,\check{\delta}]} \left(y - 3\bar{w}(\check{\delta})\right) d\mu_0^p(\delta') + \int_{\delta' \in (\check{\delta},\tilde{\delta}]} \left(\delta' y - 2\bar{w}(\check{\delta})\right) d\mu_0^p(\delta') < 0.$$

Notice that

(i) if $\check{\delta} < \delta_2 < \delta_1$, then $\underbrace{\delta_2 y - 2\bar{w}(\check{\delta})}_{\delta_2} < \underbrace{\delta_1 y - 2\bar{w}(\check{\delta})}_{\delta_1}$;

(ii) if $\delta_2 \leq \check{\delta} < \delta_1$, then $\left(\delta_1 y - 2\bar{w}(\check{\delta})\right) - \left(y - 3\bar{w}(\check{\delta})\right) = \delta_1 y + \bar{w}(\check{\delta}) - y > \check{\delta} y + \bar{w}(\check{\delta}) - y > 0$, which means $\underbrace{y - 3\bar{w}(\check{\delta})}_{\delta_2} < \underbrace{\delta_1 y - 2\bar{w}(\check{\delta})}_{\delta_1}$;

(iii)if $\delta_2 < \delta_1 \leq \check{\delta}$, then $\underbrace{y - 3\bar{w}(\check{\delta})}_{\delta_2} < \underbrace{y - 3\bar{w}(\check{\delta})}_{\delta_1}$.

Thus, for any $\check{\delta} \leq \tilde{\delta}$ we have $\chi_{\tilde{\mu}_0^p}(\check{\delta}) < \chi_{\mu_0^p}(\check{\delta}) < 0$. It follows that $w^*(\tilde{\mu}_0^p) = \frac{y}{3}$. The difference in total expected utility between the new information structure and the original one is given by $\tilde{\mathbb{E}}[U_{p|0}] - \mathbb{E}[U_{p|0}] =$

$$
\begin{aligned}
&\Big\{\pi(m_0^p)\int_{\delta_1\in C_\delta}\int_0^{\frac{\delta_1 y}{2}}(\delta_1 y-2w)dF(w)d\mu_0^p(\delta_1)+\int_{\delta_2\in C_{\delta'}}\int_0^{\frac{y}{3}}(y-3w)dF(w)d\pi(m_{0|\delta_2})\Big\}\\
&-\Big\{\pi(m_0^p)\int_{\delta_1\in C_\delta}\int_0^{\frac{y}{3}}(y-3w)dF(w)d\mu_0^p(\delta_1)+\int_{\delta_2\in C_{\delta'}}\int_0^{\frac{\delta_2 y}{2}}(\delta_2 y-2w)dF(w)d\pi(m_{0|\delta_2})\Big\}\\
\simeq&\Big\{\pi(m_0^p)\int_{\delta_1\in C_\delta}d\mu_0^p(\delta_1)\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)+\int_{\delta_2\in C_{\delta'}}d\pi(m_{0|\delta_2})\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big\}\\
&-\Big\{\pi(m_0^p)\int_{\delta_1\in C_\delta}d\mu_0^p(\delta_1)\int_0^{\frac{y}{3}}(y-3w)dF(w)+\int_{\delta_2\in C_{\delta'}}d\pi(m_{0|\delta_2})\int_0^{\frac{\delta' y}{2}}(\delta' y-2w)dF(w)\Big\}\\
=&\varepsilon\cdot\Big\{\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)-\int_0^{\frac{\delta' y}{2}}(\delta' y-2w)dF(w)\Big\}>0.
\end{aligned}
$$

We can repeat this process for all such pairs of $\{\delta,\delta'\}$ until there exist no $\hat{\delta}<\delta'<\delta\leq\tilde{\delta}$ such that $\sigma(\delta)[m_0^p]>0$ and $\sigma(\delta')[m_{0|\delta'}]>0$.

Since $\delta\sim G[0,1]$ follows a continuous distribution, Property 3 implies that in the optimal information structure, there exists a threshold δ^{SB} such that $\sigma(\delta)[m_0^p]=1$ for all $\delta\in[0,\delta^{SB}]$, and $\sigma(\delta)[m_{0|\delta}]=1$ for all $\delta\in[\delta^{SB},1]$. The induced posterior belief of m_0^p is given by $\mu_0^p(\delta)=\frac{G(\delta)}{G(\delta^{SB})}$ for $\delta\leq\delta^{SB}$. The last step is to characterize the threshold.

By definition, $\mu_0^p(\delta)=\frac{G(\delta)}{G(\delta^{SB})}$, then $\chi_{\mu_0^p}(\check{\delta})\leq 0$ if and only if $\chi_G(\check{\delta},\delta^T)\leq 0$. Take $\forall\delta^T<\delta^{SB}$, since $\delta^T\in[\hat{\delta},\tilde{\delta}]$, then the difference in total expected utility between δ^T and δ^{SB} is given by $\mathbb{E}_G[U_{p|0}(\delta^{SB})]-\mathbb{E}_G[U_{p|0}(\delta^T)]=$

$$
\begin{aligned}
&\int_{\delta'\in(\delta^T,\delta^{SB}]}\int_0^{\frac{y}{3}}(y-3w)dF(w)dG(\delta')-\int_{\delta'\in(\delta^T,\delta^{SB}]}\int_0^{\frac{\delta' y}{2}}(\delta' y-2w)dF(w)dG(\delta')\\
=&\int_{\delta'\in(\delta^T,\delta^{SB}]}\underbrace{\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)-\int_0^{\frac{\delta' y}{2}}(\delta' y-2w)dF(w)\Big)}_{>0,\ \because\delta'\leq\tilde{\delta}}dG(\delta')>0.
\end{aligned}
$$

Suppose the optimal information structure takes a different threshold $\delta^T>\delta^{SB}$. Since we have proved that $\delta^T\leq\tilde{\delta}$, then we have $\delta^{SB}<\delta^T\leq\tilde{\delta}$, which means there exists $\check{\delta}\in[\underline{\delta},\delta^T]$ such that $\chi_G(\check{\delta},\delta^T)>0$. It is equivalent to $\chi_{\mu_0^p}(\check{\delta})>0$, and thus we have $w_{\mu_0^p}(\check{\delta})>\bar{w}(\check{\delta})$, which means in equilibrium $\delta^*\in[\underline{\delta},\check{\delta})$ and $w^*(\mu_0^p)>\bar{w}(\tilde{\delta})$. From Property 1, there exists an upper revealing truncation of m_0^p which strictly increases the total expected utility, contradicting that δ^T is assumed to be optimal.

□

D.1.7 Proof of Proposition 9

Proof. Given any SPE where t_0^* is implemented, only agents with $w \le \bar{w}(t_0^*)$ participate at time 0. Principal's objective function from the time-t perspective by choosing stopping time t_0 is given by

$$\begin{aligned} U_{P|t}(t_0 \mid \bar{w}(t_0^*)) &= \int_0^{\bar{w}(t_0^*)} \Big(\int_{t'=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t')dt' - (\bar{t}(w,t_0) - t) \cdot w \Big) dF(w) \\ &= \int_0^{\bar{w}(t_0^*)} \underbrace{\Big(\int_{t'=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t')dt' - \bar{t}(w,t_0) \cdot w \Big)}_{=U_{A|0}(w,t_0)} dF(w) + \int_0^{\bar{w}(t_0^*)} t \cdot w dF(w). \end{aligned}$$

Since $U_{A|t_0^*}(w,t_0^*) > 0$ for all $w < \bar{w}(t_0^*)$, by continuity for any $t, t_0 \in (t_0^* - \varepsilon, t_0^* + \varepsilon)$, we have $U_{A|t}(w,t_0) > 0$, which means agents who participate at time 0 will never quit in the neighborhood of equilibrium. By definition of SPE, t_0^* has to be a local maximum point for $\max_{t_0} U_{P|t_0^*}(t_0 \mid \bar{w}(t_0^*))$, then we have the first order condition

$$\left. \frac{dU_{P|t_0^*}(t_0 \mid \bar{w}(t_0^*))}{dt_0} \right|_{t_0=t_0^*} = \int_0^{\bar{w}(t_0^*)} \frac{dU_{A|0}(w,t_0)}{dt_0} dF(w) \Big|_{t_0=t_0^*} = 0.$$

On the other hand, the first order derivative of principal's objective when computing the first-best solution is given by

$$\frac{dU_{P|0}(t_0)}{dt_0} = \frac{d\bar{w}(t_0)}{dt_0} \underbrace{U_{A|0}(\bar{w}(t_0),t_0)}_{=0} f(\bar{w}(t_0)) + \int_0^{\bar{w}(t_0)} \frac{dU_{A|0}(w,t_0)}{dt_0} dF(w).$$

Thus, we conclude that t_0^* must be a local maximum point of $\max_{t_0 \in [0,T]} U_{P|0}(t_0)$. Let $\mathbb{O}$ collect all these local maximum points. Every subgame perfect equilibrium SPE-t_0^* is contained in $\mathbb{O}$, and satisfies

$$t_0^* \in \arg \max_{t_0 \in [0,T]} \int_0^{\bar{w}(t_0^*)} \Big(\int_{t=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t)dt - \bar{t}(w,t_0) \cdot w \Big) dF(w).$$

Let $\mathbb{O}_1$ collect all subgame perfect equilibriums. Clearly, $\mathbb{O}_1 \neq \emptyset$, since $t_0^{FB} \in \mathbb{O}_1$.[5] Then we have $t_0^{SB} = \arg\max_{t_0 \in \mathbb{O}_1} \bar{w}(t_0)$, because in any other SPE with $t_0^* \neq t_0^{SB}$, agents with $w \in$

[5]Pick $\forall t_0 \neq t_0^{FB}$, by definition of t_0^{FB} we have

$$\begin{aligned} &\int_0^{\bar{w}(t_0^{FB})} \Big(\int_{t=t_0^{FB}}^{\bar{t}(w,t_0^{FB})} \delta(t_0^{FB},t)dt - \bar{t}(w,t_0^{FB}) \cdot w \Big) dF(w) \ge \int_0^{\bar{w}(t_0)} \Big(\int_{t=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t)dt - \bar{t}(w,t_0) \cdot w \Big) dF(w) \\ &\ge \int_0^{\bar{w}(t_0^{FB})} \Big(\int_{t=t_0}^{\bar{t}(w,t_0)} \delta(t_0,t)dt - \bar{t}(w,t_0) \cdot w \Big) dF(w). \end{aligned}$$

$(\bar{w}(t_0^*), \bar{w}(t_0^{SB})]$ quit at time 0; however, the joint participation of them will make the principal change the stopping time from t_0^* to t_0^{SB}, and strictly benefit those agents. Thus, there always exists coordination failure among agents except for t_0^{SB}.

□

D.1.8 Proof of Proposition 10

Proof. We first compute $\tilde{\delta}^t$, which determines the efficient outcome. Principal's time-0 objective functions under C_2^y and $C_1^{\delta y}$ are given by

$$U_{P|0}(C_2^y) = \int_0^{\frac{y(\bar{w}(C_2^y))}{3}} \Big(y\big(\bar{w}(C_2^y)\big) - 3w\Big) dF(w), \quad \text{where } \bar{w}(C_2^y) = \frac{1}{3} y\big(\bar{w}(C_2^y)\big)$$
$$U_{P|0}(C_1^{\delta y}) = \int_0^{\frac{\delta y(\bar{w}(C_1^{\delta y}))}{2}} \Big(\delta y\big(\bar{w}(C_1^{\delta y})\big) - 2w\Big) dF(w), \quad \text{where } \bar{w}(C_1^{\delta y}) = \frac{\delta}{2} y\big(\bar{w}(C_1^{\delta y})\big).$$

If $\delta < \frac{2}{3}$, we can easily show that $\bar{w}(C_1^{\delta y}) < \bar{w}(C_2^y)$, which means C_2^y is both the efficient and equilibrium outcome (by Theorem 9). Then there is no dynamic inconsistency for $\delta < \frac{2}{3}$. Thus, we only need to consider the case $\delta \geq \frac{2}{3}$, which means $\bar{w}(C_1^{\delta y}) \geq \bar{w}(C_2^y)$, and $\kappa = y\big(\bar{w}(C_1^{\delta y})\big)/y\big(\bar{w}(C_2^y)\big) \geq 1$. As before, $\tilde{\delta}^t$ is given by

$$\int_0^{\frac{y(\bar{w}(C_2^y))}{3}} \big(y(\bar{w}(C_2^y)) - 3w\big) dF(w) = \int_0^{\frac{\tilde{\delta}^t y(\bar{w}(C_1^{\delta y}))}{2}} \big(\tilde{\delta}^t y(\bar{w}(C_1^{\delta y})) - 2w\big) dF(w)$$
$$\Leftrightarrow \int_0^{\frac{y(\bar{w}(C_2^y))}{3}} \big(y(\bar{w}(C_2^y)) - 3w\big) dF(w) = \int_0^{\tilde{\delta}^t \kappa \cdot \frac{y(\bar{w}(C_2^y))}{2}} \Big(\tilde{\delta}^t \kappa \cdot y(\bar{w}(C_2^y)) - 2w\Big) dF(w),$$

which means $\tilde{\delta}^t \kappa = \tilde{\delta}$. Thus $\tilde{\delta}^t = \tilde{\delta}/\kappa \leq \tilde{\delta}$. As for $\hat{\delta}^t$, which is given by

$$\int_0^{\frac{\hat{\delta}^t y(\bar{w}(C_2^y))}{2}} \big(y(\bar{w}(C_2^y)) - 2w\big) dF(w) = \int_0^{\frac{\hat{\delta}^t y(\bar{w}(C_2^y))}{2}} \big(\hat{\delta}^t y(\bar{w}(C_2^y)) - w\big) dF(w),$$

we have that $\hat{\delta}^t \equiv \hat{\delta}$.

If $\tilde{\delta} > \kappa\hat{\delta}$, then $\tilde{\delta}^t > \hat{\delta}^t$, which means that the collective decision is present-biased and dynamic preference reversals occurs for $\delta \in (\hat{\delta}^t, \tilde{\delta}^t)$. If $\tilde{\delta} = \kappa\hat{\delta}$, the range of δ which induces preference reversals completely disappears. If $\tilde{\delta} < \kappa\hat{\delta}$, then $\tilde{\delta}^t < \hat{\delta}^t$, which means that the collective decision is future-biased. That is, for $\delta \in (\tilde{\delta}^t, \hat{\delta}^t)$, the principal initially prefers earlier consumption ($C_1^{\delta y}$) but changes to later consumption (C_2^y) when the decision time is postponed.

□

D.2 Other Omitted Results

D.2.1 Supplementary lemmas

Lemma 19. When δ is common knowledge, the pure-strategy SPE-NCF is the *unique* subgame perfect Nash equilibrium if stage (i) is extended to be a perfect-information sequential game with arbitrary decision order of agents.

Proof. Pick arbitrary decision order of agents, and relabel them in an ascending manner. Participation order is deterministic and common knowledge. As before, $w(i):[0,1]\mapsto[0,W]$ specifies the outside option for agent i. Each agent's action at time 0 is given by $a_i\in\{0,1\}$, where 0 stands for quitting the group while 1 stands for participating. A history $h(i):[0,1]\mapsto H=\bigcup_{\tau\in[0,1]}\{a:[0,\tau)\mapsto\{0,1\}\}$, is the action profile of all the previous agents when agent i is making a decision. We only need to consider the case where $\delta>\hat{\delta}$. Given any SPE and any i, in the subgame $\Gamma(h_i)$ we have:

(1) If $w(i)\le\frac{y}{3}$, then the agent gets 0 surplus by choosing $a_i=0$, but always gets positive gain[6] by choosing $a_i=1$ no matter what subgame perfect equilibrium would be induced in $\Gamma(h_i,a_i=1)$, i.e. C_2^y or $C_1^{\delta y}$.

(2) If $w(i)>\frac{\delta y}{2}$, the agent quits at time 0 because he always gets negative surplus in subgame perfect equilibrium by choosing $a_i=1$.

(3) If $\frac{y}{3}<w(i)\le\frac{\delta y}{2}$, one can easily check that C_2^y is implemented in subgame perfect equilibrium only when both $\Gamma(h_i,a_i=0)$ and $\Gamma(h_i,a_i=1)$ induce SPE where C_2^y is implemented.[7]

Thus we have that, (i) only $w\in(\frac{y}{3},\frac{\delta y}{2}]$ are strategic, and (ii) C_2^y is implemented in a subgame only when the terminal nodes are all C_2^y. Since $\delta>\hat{\delta}$, the joint participation of $w\in[0,\frac{\delta y}{2}]$ can induce $C_1^{\delta y}$, then $C_1^{\delta y}$ is one of the terminal nodes of the original game. Thus, in any SPE the outcome must be $C_1^{\delta y}$.

Next, we show that the unique SPE is where all $w\in[0,\frac{\delta y}{2}]$ participate at time 0. Suppose there exists some i such that $w(i)\in(\frac{y}{3},\frac{\delta y}{2}]$ and $a_i=0$, then we must have: $C_1^{\delta y}$ is induced in

[6] We assume that when agents are indifferent between participating and quitting the group, they prefer staying in the group to consume the public goods.

[7] We distinguish four possibilities to show the result: (1) if $\Gamma(h_i,a_i=0)\to^{SPE}C_2^y$ and $\Gamma(h_i,a_i=1)\to^{SPE}C_2^y$, then $a_i(h_i)=0$ and C_2^y is implemented; (2) if $\Gamma(h_i,a_i=0)\to^{SPE}C_2^y$ and $\Gamma(h_i,a_i=1)\to^{SPE}C_1^{\delta y}$, then $a_i(h_i)=1$ and $C_1^{\delta y}$ is implemented; (3) if $\Gamma(h_i,a_i=0)\to^{SPE}C_1^{\delta y}$ and $\Gamma(h_i,a_i=1)\to^{SPE}C_2^y$, then $a_i(h_i)=0$ and $C_1^{\delta y}$ is implemented; (4)if $\Gamma(h_i,a_i=0)\to^{SPE}C_1^{\delta y}$ and $\Gamma(h_i,a_i=1)\to^{SPE}C_1^{\delta y}$, then $a_i(h_i)=1$ and $C_1^{\delta y}$ is implemented.

the SPE of $\Gamma(h_i, a_i = 0)$, while C_2^y is induced in the SPE of $\Gamma(h_i, a_i = 1)$. Denote agents' action profile in SPE as $(h_i, a_i = 0, (a_{\tilde{i}})_{\tilde{i}>i})$, which induces the principal to choose $C_1^{\delta y}$. Since agent i also prefers $C_1^{\delta y}$, the action profile $(h_i, a_i = 1, (a_{\tilde{i}})_{\tilde{i}>i})$ also induces the principal to choose $C_1^{\delta y}$. Thus, $C_1^{\delta y}$ is also one of the terminal nodes of subgame $\Gamma(h_i, a_i = 1)$, contradicting that C_2^y is the associated SPE. □

Lemma 20. In the Bayesian persuasion model with a continuum of agents, the pure-strategy SPE-NCF can be uniquely approximated by a sequence of subgame perfect equilibriums of stage-(i) sequential games with finitely many agents.

Proof. We apply the framework established in Fudenberg and Levine (1986). The set of players of the original game, i.e., $I = [0,1]$, is endowed with Lebesgue measure. Consider a sequence of finite sequential games converging to the original game, where the n-th game has a set of players denoted by $I_n = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$. A pure strategy profile for the n-th game is a map $s_n : H_n \mapsto A = \{0,1\}$ from histories to actions, where $H_n = \bigcup_{i \in I_n} H_n^i$ and $H_n^i = \{0,1\}^{\#(i' \prec i)}$, with $\#(i' \prec i)$ representing the number of agents who make decisions before agent i. In the original game a pure strategy profile is a measurable map $s : H \mapsto A = \{0,1\}$. Let S_n (or S) be the space of all maps from H_n (or H) to A. The payoff function of agent i in the n-th game is $\pi_n^i : A \times H_n \mapsto \mathbb{R}$; while the payoff function of agent i in the original game is $\pi^i : A \times H \mapsto \mathbb{R}$. There are natural restriction mappings $r_n : S \mapsto S_n$. For the n-th game, a subgame perfect equilibrium is $\hat{s}_n = r_n(\hat{s}) \in S_n$ such that for any $i \in I_n$, $h_n^i \in H_n^i$ and $a_i \in \{0,1\}$, we have $\pi_n^i(\hat{s}^i, \hat{s}^{-i} \mid h_n^i) \geq \pi_n^i(a_i, \hat{s}^{-i} \mid h_n^i)$. For the original game, a subgame perfect equilibrium is $\hat{s} \in S$ such that for any $i \in I$, $h^i \in H^i$ and $a_i \in \{0,1\}$, we have $\pi^i(\hat{s}^i, \hat{s}^{-i} \mid h^i) \geq \pi^i(a_i, \hat{s}^{-i} \mid h^i)$.

Suppose the limit of $\{\hat{s}_n\}_{n=1}^{\infty}$, denoted by $\hat{s}$, is not a SPE of the original game, then there exist some i, h^i and $a_i \neq \hat{s}^i(h^i)$ such that $\pi^i(\hat{s}^i, \hat{s}^{-i} \mid h^i) < \pi^i(a_i, \hat{s}^{-i} \mid h^i) - \varepsilon$ for some $\varepsilon > 0$. By definition of agents' utility functions, π^i is uniformly continuous. Then there exists some sufficiently large integer N such that for all $n > N$ we have

$$\sup_{i, a^i, \hat{s}^{-i}, h^i} |\pi^i(a_i, \hat{s}^{-i} \mid h^i) - \pi_n^i(a_i, r_n(\hat{s}^{-i}) \mid r_n(h^i))| < \frac{\varepsilon}{3}.$$

It follows that

$$\begin{aligned} &\pi_n^i(\hat{s}_n^i, \hat{s}_n^{-i} \mid h_n^i) - \pi_n^i(a_i, \hat{s}_n^{-i} \mid h_n^i) < \left(\pi^i(\hat{s}^i, \hat{s}^{-i} \mid h^i) + \frac{\varepsilon}{3}\right) - \left(\pi^i(a_i, \hat{s}^{-i} \mid h^i) - \frac{\varepsilon}{3}\right) \\ =&\pi^i(\hat{s}^i, \hat{s}^{-i} \mid h^i) - \pi^i(a_i, \hat{s}^{-i} \mid h^i) + \frac{2\varepsilon}{3} < -\varepsilon + \frac{2\varepsilon}{3} < 0, \end{aligned}$$

contradicting that $\hat{s}_n$ is SPE of the n-th finite game.

On the other hand, let $\hat{s}$ be the SPE-NCF, if $\{\hat{s}_n\}_{n=1}^{\infty}$ converges to some SPE $\hat{s}' \neq \hat{s}$ with coordination failure, then the distance between $\hat{s}_n$ and $r_n(\hat{s})$ is bounded below from some positive number when n is sufficiently large. At the same time, due to the uniform continuity of π^i, the difference in payoff structure between the original game and the n-th finite game gets arbitrarily small, which means $r_n(\hat{s})$ constitutes another SPE of the n-th game, contradicting the uniqueness of SPE in finite extensive-form game. Thus, $\{\hat{s}_n\}_{n=1}^{\infty}$ must converge to $\hat{s}$.

□

D.2.2 An example on mechanism designs

Assume that the liquidation option (debt plan) is provided by the social planner (grand principal), whose cost of public funds is $C(\cdot) > 0$. Group leader's time-2 resource y and agents' outside options are common knowledge within the group; while the grand principal only knows that w follows a distribution function $F(w)$ with full-support continuous density $f(w)$, and y follows a distribution function $G(y)$ with full-support density $g(y)$. The grand principal's aim is to maximize the expected net surplus by offering the group leader a menu of debt plans.

The mechanism (M, δ) is made public at time 0 and the grand principal can commit to it for the whole time horizon. At time 1 the group leader sends a message $m \in M$ to the grand principal and then is entitled to a borrowing rate $\delta(m)$. We assume that no transfer is allowed, and that the grand principal cannot contract on the amount of resource to be liquidated. As before, we look for subgame perfect equilibrium without coordination failure among agents. Given the mechanism (M, δ), in equilibrium agents' participation at time 0 is $I_1^*(y)$. At time 1, the group leader's equilibrium message is $m^*\left(y, I_1^*(y)\right)$ and the associated borrowing rate is $\delta\left(m^*(y, I_1^*(y))\right)$. By revelation principle, we can consider direct mechanism where the group leader only needs to report time 2's resource.

The benchmark case is that the group leader has the commitment power to convince the agents that her time 1's message will align with her time 0's promise. Group leader's objective function is given by

$$U_{p|0}^{c}(y, \delta) = \begin{cases} \int_0^{\frac{y}{3}} (y - 3w) dF(w) & \text{if } \delta \leq \tilde{\delta}(y) \\ \int_0^{\frac{\delta y}{2}} (\delta y - 2w) dF(w) & \text{if } \delta > \tilde{\delta}(y), \end{cases}$$

where $\tilde{\delta}(y)$, as well as $\hat{\delta}(y)$, is defined as before. Obviously, $U_{p|0}^{c}(y, \delta)$ is increasing on δ, which means group leader will always choose the message inducing the highest δ. Thus it is sufficient to offer a unique δ together with the default option. The objective function of the grand principal when offering δ is given by $O^c(\delta) =$

$$\int_{\tilde{\delta}(y)\geq\delta}\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big)dG(y)+\int_{\tilde{\delta}(y)<\delta}\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)-C(\delta y)\Big)dG(y).$$

The optimal borrowing rate in the benchmark case is given by $\delta^c = \arg\max_{\delta\in[0,1]} O^c(\delta)$.

Time inconsistency issue kicks in when group leader doesn't have such commitment power. At time 1, the group leader's objective function is

$$U^{nc}_{p|1}(y,\delta) = \begin{cases} \int_0^{\frac{y}{3}}(y-2w)dF(w) & \text{if } \delta \leq \hat{\delta}(y) \\ \int_0^{\frac{\delta y}{2}}(\delta y-w)dF(w) & \text{if } \delta > \hat{\delta}(y). \end{cases}$$

Suppose grand principal offers $\delta > \delta'$ in the menu, then for any y, if $\delta > \delta' > \hat{\delta}(y)$ or $\delta > \hat{\delta}(y) \geq \delta'$, then $I_1^* = \{i \in I \mid w(i) \leq \frac{\delta y}{2}\}$ and group leader chooses δ at time 1; if $\hat{\delta}(y) \geq \delta > \delta'$, then $I_1^* = \{i \in I \mid w(i) \leq \frac{y}{3}\}$ and group leader chooses the default option. Thus it is sufficient to offer a unique δ together with the default option. The grand principal's objective function is given by $O^{nc}(\delta) =$

$$\int_{\hat{\delta}(y)\geq\delta}\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big)dG(y)+\int_{\hat{\delta}(y)<\delta}\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)-C(\delta y)\Big)dG(y).$$

The optimal borrowing rate is given by $\delta^{nc} = \arg\max_{\delta\in[0,1]} O^{nc}(\delta)$.

Notice that for any y we have $\hat{\delta}(y) < \tilde{\delta}(y)$, then fixed $\forall\delta$, we get

$$\begin{aligned} O^c(\delta)-O^{nc}(\delta) &= \int_{\hat{\delta}(y)<\delta\leq\tilde{\delta}(y)}\Big(\int_0^{\frac{y}{3}}(y-3w)dF(w)\Big)dG(y) \\ &\quad -\int_{\hat{\delta}(y)<\delta\leq\tilde{\delta}(y)}\Big(\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)-C(\delta y)\Big)dG(y) \\ &= \int_{\hat{\delta}(y)<\delta\leq\tilde{\delta}(y)}\Big(\underbrace{\int_0^{\frac{y}{3}}(y-3w)dF(w)-\int_0^{\frac{\delta y}{2}}(\delta y-2w)dF(w)}_{\geq 0,\ \because \delta\leq\tilde{\delta}(y)}+C(\delta y)\Big)dG(y)\geq 0. \end{aligned}$$

Thus, time inconsistency issue is harmless if there exists $\delta^c \in \arg\max_{\delta\in[0,1]} O^c(\delta)$ such that $\int_{\hat{\delta}(y)<\delta^c\leq\tilde{\delta}(y)} dG(y) = 0$. To prove this, we first have $O^c(\delta^c) = O^{nc}(\delta^c)$. Together with the fact that $O^c(\delta) \geq O^{nc}(\delta)$ for all δ , we have $O^{nc}(\delta^c) = O^c(\delta^c) \geq O^c(\delta) \geq O^{nc}(\delta)$ for all δ, which means $O^{nc}(\delta^c) \geq O^{nc}(\delta^{nc})$. It follows that $O^{nc}(\delta^c) = O^{nc}(\delta^{nc})$. Thus $\delta^c = \delta^{nc}$ and there is no welfare loss caused by time inconsistency issue.

Notice that $\frac{2}{3} < \tilde{\delta}(y) \leq 1$ for all y, then immediately time inconsistency has no effect on efficiency if $\delta^c = 1$ or $\delta^c \leq \frac{2}{3}$. More general, let $O^c := O^c(\delta^c)$ and $O^{nc} := O^{nc}(\delta^{nc})$, the following proposition characterizes when efficiency is undermined by the presence of present-biased collective decision.

Proposition 11. Social welfare is harmed by dynamic inconsistency, i.e. $O^c > O^{nc}$, if and only if for all $\delta^c \in \arg\max_\delta O^c(\delta)$ we have $\inf_\delta \tilde{\delta}(y) < \delta^c < \sup_\delta \tilde{\delta}(y)$.

Proof. (*Sufficiency*) By assumption on $F(w)$, we have that $\tilde{\delta}(y)$ and $\hat{\delta}(y)$ are both continuous with respect to y. Pick $\forall\delta^c \in \arg\max_\delta O^c(\delta)$, by continuity of $\tilde{\delta}(y)$, there exist some y^c such that: (1) $\tilde{\delta}(y^c) = \delta^c$, and (2) for all $y' \in (y^c, y^c+\varepsilon_1)$ we have $\tilde{\delta}(y') \in (\delta^c, \delta^c+\tau_1)$. Let τ be some strictly positive number such that $\tau < \tilde{\delta}(y^c) - \hat{\delta}(y^c)$. By continuity of $\hat{\delta}(y)$, there exists $\varepsilon_2 > 0$ such that for all $y' \in (y^c, y^c+\varepsilon_2)$ we have $\hat{\delta}(y') < \hat{\delta}(y^c) + \tau < \tilde{\delta}(y^c)$. Define $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$, then for all $y' \in (y^c, y^c+\varepsilon)$ we have $\hat{\delta}(y') < \tilde{\delta}(y^c) = \delta^c < \tilde{\delta}(y')$. Since $G(y)$ is assumed to have full-support density, we have $O^c(\delta^c) - O^{nc}(\delta^c) \geq \int_{\hat{\delta}(y)<\delta^c\leq\tilde{\delta}(y)} C(\delta^c y)dG(y) > \int_{y^c}^{y^c+\varepsilon} C(\delta^c y)dG(y) > 0$. On the other hand, for any $\delta' \notin \arg\max_\delta O^c(\delta)$, we have $O^c > O^c(\delta') \geq O^{nc}(\delta')$. Thus we have $O^c > O^{nc}(\delta)$ for all $\delta \in [0,1]$. It follows that $O^c > O^{nc}$.

(*Necessity*) By definition, $O^c(\delta^c) \geq O^c(\delta^c+\varepsilon)$ and $O^c(\delta^c) \geq O^c(\delta^c-\varepsilon)$ for all $\varepsilon > 0$. Plug the expression of $O^c(\cdot)$ into these two inequalities, and omit all higher-order infinitesimal terms, then we get

$$\begin{aligned}
O^c(\delta^c) - O^c(\delta^c+\varepsilon) \simeq &\int_{\delta^c<\tilde{\delta}(y)\leq\delta^c+\varepsilon} C(\delta^c y)dG(y) \\
&-\varepsilon\cdot\int_{\tilde{\delta}(y)\leq\delta^c}\Big(F(\frac{\delta^c y}{2})y - C'(\delta^c y)y\Big)dG(y) \geq 0, \\
O^c(\delta^c) - O^c(\delta^c-\varepsilon) \simeq &-\int_{\delta^c-\varepsilon<\tilde{\delta}(y)\leq\delta^c} C(\delta^c y)dG(y) \\
&+\varepsilon\cdot\int_{\tilde{\delta}(y)\leq\delta^c}\Big(F(\frac{\delta^c y}{2})y - C'(\delta^c y)y\Big)dG(y) \geq 0.
\end{aligned}$$

Suppose $\exists\delta^c \in \arg\max_\delta O^c(\delta)$ such that $\delta^c = \sup_\delta \tilde{\delta}(y)$, then $\tilde{\delta}(y) \leq \delta^c$ for all y, which means $\int_{\delta^c<\tilde{\delta}(y)\leq\delta^c+\varepsilon} C(\delta^c y)dG(y) = 0$ and $\int_{\delta^c-\varepsilon<\tilde{\delta}(y)\leq\delta^c} C(\delta^c y)dG(y) > 0$. Then

$$\begin{aligned}
O^c(\delta^c+\varepsilon) - O^c(\delta^c) \simeq &\varepsilon\cdot\int_{\tilde{\delta}(y)\leq\delta^c}\Big(F(\frac{\delta^c y}{2})y - C'(\delta^c y)y\Big)dG(y) \\
\geq &\int_{\delta^c-\varepsilon<\tilde{\delta}(y)\leq\delta^c} C(\delta^c y)dG(y) > 0,
\end{aligned}$$

which means the grand principal can strictly improve the total welfare by increasing δ. Thus $\sup_\delta \tilde{\delta}(y)$ cannot be the maximizer of $O^c(\delta)$.

If $\exists \delta^c \in \arg\max_{\delta} O^c(\delta)$ such that $\delta^c > \sup_{\delta} \tilde{\delta}(y)$, then $\int_{\hat{\delta}(y)<\delta^c\leq\tilde{\delta}(y)} dG(y) = 0$. It follows that $\delta^c \in \arg\max_{\delta} O^{nc}(\delta)$ and $O^c = O^{nc}$.

If $\exists \delta^c \in \arg\max_{\delta} O^c(\delta)$ such that $\delta^c \leq \inf_{\delta} \tilde{\delta}(y)$, then in the benchmark case, for all y the group leader is offered the default option. Thus it is equivalent to set $\delta^c = 0$, which means $0 \in \arg\max_{\delta} O^c(\delta)$. It follows that $0 \in \arg\max_{\delta} O^{nc}(\delta)$ and $O^c = O^{nc}$. $\square$

Bibliography

ALONSO, R., W. DESSEIN, AND N. MATOUSCHEK (2008): "When Does Coordination Require Centralization?," *American Economic Review*, 98(1), 145–79.

ANDERSON, S. P., AND R. RENAULT (2006): "Advertising content," *The American Economic Review*, 96(1), 93–113.

ANGELETOS, G.-M., AND A. PAVAN (2007): "Efficient use of information and social value of information," *Econometrica*, 75(4), 1103–1142.

BECKER, R. A. (1992): "Cooperative capital accumulation games and the core," in *Economic Theory and International Trade*, pp. 1–27. Springer.

BERGEMANN, D., AND S. MORRIS (2005): "Robust mechanism design," *Econometrica*, 73(6), 1771–1813.

——— (2009b): "Robust implementation in direct mechanisms," *The Review of Economic Studies*, 76(4), 1175–1204.

BERGEMANN, D., AND M. PESENDORFER (2007): "Information structures in optimal auctions," *Journal of economic theory*, 137(1), 580–609.

BISIN, A., A. LIZZERI, AND L. YARIV (2015): "Government policy with time inconsistent voters," *The American Economic Review*, 105(6), 1711–1737.

BLEICHRODT, H., K. I. ROHDE, AND P. P. WAKKER (2009): "Non-hyperbolic time inconsistency," *Games and Economic Behavior*, 66(1), 27–38.

BÖRGERS, T. (2013): "(No) Foundations of Dominant-Strategy Mechanisms: A Comment on Chung and Ely (2007)," mimeo, University of Michigan.

BÖRGERS, T., A. HERNANDO-VECIANA, AND D. KRÄHMER (2013): "When are signals complements or substitutes?," *Journal of Economic Theory*, 148(1), 165–195.

BRANDENBURGER, A., AND E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

CARROLL, G. (2017): "Robustness and separation in multidimensional screening," *Econometrica*, 85(2), 453–488.

CHEN, Y.-C., AND J. LI (2016): "Revisiting the Foundations of Dominant-Strategy Mechanisms," mimeo, National University of Singapore.

CHUNG, K.-S., AND J. C. ELY (2007): "Foundations of dominant-strategy mechanisms," *The Review of Economic Studies*, 74(2), 447–476.

DASGUPTA, P., AND E. MASKIN (2000): "Efficient auctions," *Quarterly Journal of Economics*, 115(2), 341–388.

DOW, J., AND G. GORTON (1993): "Trading, communication and the response of asset prices to news," *The Economic Journal*, 103(418), 639–646.

DWORK, C. (2006): "Differential Privacy," in *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*, vol. 4052, pp. 1–12, Venice, Italy. Springer Verlag.

DWORK, C., F. MCSHERRY, K. NISSIM, AND A. SMITH (2006): "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference (TCC)*, vol. 3876, pp. 265–284. Springer.

ELIAZ, K. (2002): "Fault tolerant implementation," *The Review of Economic Studies*, 69(3), 589–610.

ERLINGSSON, Ú., V. PIHUR, AND A. KOROLOVA (2014): "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067. ACM.

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

FADEN, A. M. (1985): "The existence of regular conditional probabilities: necessary and sufficient conditions," *The Annals of Probability*, pp. 288–298.

FELDSTEIN, M. S. (1964): "The social time preference discount rate in cost benefit analysis," *The Economic Journal*, 74(294), 360–379.

FISHBURN, P. C., AND A. RUBINSTEIN (1982): “Time preference,” *International economic review*, pp. 677–694.

FUDENBERG, D., AND D. K. LEVINE (1986): “Finite player approximations to a continuum of players,” *A Long-Run Collaboration on Long-Run Games*, p. 59.

——— (1988): “Open-loop and closed-loop equilibria in dynamic games with many players,” *Journal of Economic Theory*, 44(1), 1–18.

——— (2006): “A dual-self model of impulse control,” *The American Economic Review*, pp. 1449–1476.

GALPERTI, S. (2015): “Commitment, flexibility, and optimal screening of time inconsistency,” *Econometrica*, 83(4), 1425–1465.

GANUZA, J.-J., AND J. S. PENALVA (2010): “Signal orderings based on dispersion and the supply of private information in auctions,” *Econometrica*, 78(3), 1007–1030.

GOLDBERG, A. V., J. D. HARTLINE, A. R. KARLIN, M. SAKS, AND A. WRIGHT (2006): “Competitive auctions,” *Games and Economic Behavior*, 55(2), 242–269.

GOLLIER, C., AND R. ZECKHAUSER (2005): “Aggregation of heterogeneous time preferences,” *Journal of Political Economy*, 113(4), 878–896.

GRAY, R. M., ET AL. (2006): “Toeplitz and circulant matrices: A review,” *Foundations and Trends® in Communications and Information Theory*, 2(3), 155–239.

GREEN, E. J. (1984): “Continuum and finite-player noncooperative models of competition,” *Econometrica: Journal of the Econometric Society*, pp. 975–993.

HAGHPANAH, N., AND J. HARTLINE (2014): “Multi-dimensional Virtual Values and Second-degree Price Discrimination,” *arXiv preprint arXiv:1404.1341*.

HALEVY, Y. (2015): “Time consistency: Stationarity and time invariance,” *Econometrica*, 83(1), 335–352.

JACKSON, M. O., AND L. YARIV (2014): “Present bias and collective dynamic choice in the lab,” *The American Economic Review*, 104(12), 4184–4204.

——— (2015): “Collective dynamic choice: the necessity of time inconsistency,” *American Economic Journal: Microeconomics*, 7(4), 150–178.

JEHIEL, P., M. MEYER-TER VEHN, B. MOLDOVANU, AND W. R. ZAME (2006): "The limits of ex post implementation," *Econometrica*, 74(3), 585–610.

KADAN, O., P. J. RENY, AND J. M. SWINKELS (2017): "Existence of Optimal Mechanisms in Principal-Agent Problems," *Econometrica*, 85(3), 769–823.

KALAI, A. T., E. KALAI, E. LEHRER, AND D. SAMET (2010): "A commitment folk theorem," *Games and Economic Behavior*, 69(1), 127–137.

KAMENICA, E., AND M. GENTZKOW (2011): "Bayesian persuasion," *The American Economic Review*, 101(6), 2590–2615.

KOCHERLAKOTA, N. R. (1996): "Implications of efficient risk sharing without commitment," *The Review of Economic Studies*, 63(4), 595–609.

KOLOTILIN, A., M. LI, T. MYLOVANOV, AND A. ZAPECHELNYUK (2015): "Persuasion of a privately informed receiver," *Working paper*.

KOLOTILIN, A., T. MYLOVANOV, A. ZAPECHELNYUK, AND M. LI (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85(6), 1949–1964.

KOOPMANS, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica: Journal of the Econometric Society*, pp. 287–309.

KRÄHMER, D. (2017): "Correlated information structures and optimal auctions," *Working paper*.

KYDLAND, F. E., AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *The journal of political Economy*, pp. 473–491.

LAIBSON, D. (1997): "Golden eggs and hyperbolic discounting," *The Quarterly Journal of Economics*, pp. 443–477.

LEWIS, T. R., AND D. E. SAPPINGTON (1994): "Supplying information to facilitate price discrimination," *International Economic Review*, pp. 309–327.

LI, H., AND X. SHI (2017): "Discriminatory information disclosure," *American Economic Review*, 107(11), 3363–85.

LIU, Q. (2015): "Correlation and common priors in games with incomplete information," *Journal of Economic Theory*, 157, 49–75.

LOEWENSTEIN, G., AND D. PRELEC (1992): "Anomalies in intertemporal choice: Evidence and an interpretation," *The Quarterly Journal of Economics*, pp. 573–597.

MARGLIN, S. A. (1963): "The social rate of discount and the optimal rate of investment," *The Quarterly Journal of Economics*, pp. 95–111.

MASKIN, E., AND J. TIROLE (1990): "The principal-agent relationship with an informed principal: The case of private values," *Econometrica: Journal of the Econometric Society*, pp. 379–409.

——— (1992): "The principal-agent relationship with an informed principal, II: Common values," *Econometrica: Journal of the Econometric Society*, pp. 1–42.

MERTENS, J.-F., AND S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

MILGROM, P. R., AND R. J. WEBER (1982): "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, pp. 1089–1122.

MORRIS, S., AND H. S. SHIN (2002): "Social value of public information," *The American Economic Review*, 92(5), 1521–1534.

MUSSA, M., AND S. ROSEN (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

MYERSON, R. B. (1983): "Mechanism design by an informed principal," *Econometrica: Journal of the Econometric Society*, pp. 1767–1797.

MYLOVANOV, T., AND T. TRÖGER (2012): "Informed-principal problems in environments with generalized private values," *Theoretical Economics*, 7(3), 465–488.

O'DONOGHUE, T., AND M. RABIN (1999): "Doing it now or later," *American Economic Review*, pp. 103–124.

OTTAVIANI, M., AND A. PRAT (2001): "The value of public information in monopoly," *E-conometrica*, 69(6), 1673–1683.

PETERS, M., AND C. TRONCOSO-VALVERDE (2013): "A folk theorem for competing mechanisms," *Journal of Economic Theory*, 148(3), 953–973.

POSTLEWAITE, A., AND D. SCHMEIDLER (1986): "Implementation in differential information economies," *Journal of Economic Theory*, 39(1), 14–33.

RAYO, L., AND I. SEGAL (2010): "Optimal information disclosure," *Journal of political Economy*, 118(5), 949–987.

RENAULT, J., L. RENOU, AND T. TOMALA (2014): "Secure message transmission on directed networks," *Games and Economic Behavior*, 85, 1–18.

RENOU, L., AND T. TOMALA (2012): "Mechanism design and communication networks," *Theoretical Economics*, 7(3), 489–533.

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, pp. 783–826.

SEGAL, I. (2003): "Optimal pricing mechanisms with unknown demand," *The American economic review*, 93(3), 509–529.

SIMS, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50(3), 665–690.

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

THOMAS, J., AND T. WORRALL (1988): "Self-enforcing wage contracts," *The Review of Economic Studies*, 55(4), 541–554.

WILSON, R. (1985): "Incentive efficiency of double auctions," *Econometrica: Journal of the Econometric Society*, pp. 1101–1115.

YAMASHITA, T. (2015): "Implementation in Weakly Undominated Strategies: Optimality of Second-Price Auction and Posted-Price Mechanism," *Review of Economic Studies*, 82(3), 1223–1246.

——— (2016): "Optimal public information disclosure by mechanism designer," *Working paper*.

——— (2017): "Optimal public information disclosure by mechanism designer," Unpublished Manuscript, Toulouse School of Economics.