PERSUASION WITH NON-LINEAR PREFERENCES

ANTON KOLOTILIN, ROBERTO CORRAO, AND ALEXANDER WOLITZKY

Abstract. In persuasion problems where the receiver's utility is single-peaked in a one-dimensional action, optimal signals are characterized by duality, based on a first-order approach to the receiver's problem. A signal that pools at most two states in each realization is always optimal, and such *pairwise* signals are the only solutions under a non-singularity condition on utilities (the *twist condition*). Our core results provide conditions under which higher actions are induced at more or less extreme pairs of states, so that the induced action is *single-dipped* or *single-peaked* on each set of nested pairs of states. We also provide conditions for the optimality of either full disclosure or *negative assortative disclosure*, where signal realizations can be ordered from least to most extreme. Methodologically, our proofs rely on a novel complementary slackness theorem for persuasion problems.

JEL Classification: C78, D82, D83

Keywords: persuasion, information design, duality, optimal transport, first-order approach, pairwise signals, twist condition, single-dipped disclosure, negative assortative disclosure, complementary slackness

---

Date: 14th February 2023.

Kolotilin: School of Economics, UNSW Business School.
Corrao and *Wolitzky*: Department of Economics, MIT.
This paper was previously circulated with the title "Persuasion as Matching." It supercedes the earlier paper "Assortative Information Disclosure" by Kolotilin and Wolitzky. For helpful comments and suggestions, we thank Jakša Cvitanić, Piotr Dworczak, Jeffrey Ely, Drew Fudenberg, Emir Kamenica, Elliot Lipnowski, Stephen Morris, Paula Onuchic, Eran Shmaya, and Andriy Zapechelnyuk, as well as many seminar participants. We thank Daniel Clark and Yucheng Shang for excellent research assistance. Anton Kolotilin gratefully acknowledges support from the Australian Research Council Discovery Early Career Research Award DE160100964 and from MIT Sloan's Program on Innovation in Markets and Organizations. Alexander Wolitzky gratefully acknowledges support from NSF CAREER Award 1555071 and Sloan Foundation Fellowship 2017-9633.

1. Introduction

Following the seminal papers of Rayo and Segal (2010) and Kamenica and Gentzkow (2011), the past decade has witnessed an explosion of interest in the design of optimal information disclosure policies, or Bayesian persuasion. While significant progress has been made in the special case where the sender's and receiver's utilities are linear in the unknown state (e.g., Gentzkow and Kamenica 2016, Kolotilin, Mylovanov, Zapechelnyuk, and Li 2017, Kolotilin 2018, Dworczak and Martini 2019, Kleiner, Moldovanu, and Strack 2021)—so that a distribution over states can be summarized by its mean—general results beyond this simple case remain scarce. The literature to date thus has little to say about the qualitative implications of economically natural curvature properties of utilities, or about the robustness of optimal disclosure patterns uncovered in the linear case when utilities are non-linear.

This paper studies persuasion with non-linear preferences. We consider a general persuasion problem with one sender and one receiver, where the receiver's action and the state of the world are both one-dimensional. We assume that the sender always prefers higher actions, the receiver prefers higher actions at higher states, and the receiver's expected utility is single-peaked in his action for any belief about the state. In this model, the receiver's action is optimal iff his expected marginal utility from increasing his action equals zero: that is, iff the receiver's first-order condition holds. The validity of this *first-order approach* is key for tractability. We provide three types of results.

First, we show that it is always without loss to focus on *pairwise signals*, where each induced posterior belief has at most binary support. Moreover, under a non-singularity condition on the sender's and receiver's utilities—which we call the *twist condition*—every optimal signal is pairwise. This result implies that, for example, no-disclosure is generically suboptimal whenever the support of the prior contains three or more states.[1]

Second, and most importantly, we ask when it is optimal for the sender to induce higher actions at more or less extreme states. That is, if the sender pools two states $\theta_1 < \theta_3$ to induce some action a, should the actions induced at intervening states $\theta_2 \in (\theta_1, \theta_3)$ be lower than a—in which case we say that disclosure is *single-dipped*, as more extreme states induce higher actions—or higher than a—in which case we say

[1] Our result also subsumes some known special cases where optimal signals are pairwise (e.g., Rayo and Segal 2010, Alonso and Câmara 2016, Zhang and Zhou 2016).

that disclosure is *single-peaked*? This question turns out to be key for understanding optimal disclosure patterns with non-linear preferences. Our core results provide general conditions for the optimality of single-dipped disclosure (and, similarly, single-peaked disclosure), which are all based on a very simple idea. If disclosure is not single-dipped, then there must exist a *single-peaked triple*: a pair of pooled states $\{\theta_1, \theta_3\}$ and an intervening state $\theta_2 \in (\theta_1, \theta_3)$ such that the induced action at θ_2 (say, action a_2) is greater than the induced action at $\{\theta_1, \theta_3\}$ (say, action a_1). Our conditions ensure that any single-peaked triple can be profitably perturbed in the direction of single-dippedness by shifting weight on θ_1 and θ_3 from a_1 to a_2, while shifting weight on θ_2 in the opposite direction. The conditions are a bit complicated in the general model, but they are very simple in leading special cases. In particular, if the receiver's optimal action equals the posterior mean state (the *linear receiver case*), then single-dipped disclosure is optimal if the sender's marginal utility is convex in the state; and if the sender's utility is state-independent (the *state-independent sender case*), then single-dipped disclosure is optimal if the cross-partial of the receiver's utility is log-supermodular. We also establish a notable theoretical implication of single-dippedness/-peakedness: under a regularity condition, whenever a strict version of this property holds, the optimal outcome is unique.

Third, we provide conditions for the optimality of either full disclosure, where the state is always disclosed, or (more interestingly) *negative assortative disclosure*, where the states are paired in a negatively assortative manner, so that signal realizations can be ordered from least to most extreme, and only a single state "in the middle" is disclosed. Intuitively, full disclosure and negative assortative disclosure represent the extremes of maximum disclosure (disclosing all states) and minimal disclosure (disclosing only one state). There is a unique full disclosure outcome, but there are many negative assortative disclosure outcomes, depending on the weights on the states in each pair. We further characterize the optimal negative assortative disclosure outcome as the solution of a pair of ordinary differential equations, and provide examples where these equations admit an explicit solution. Notably, negative assortative disclosure is optimal whenever our conditions for the optimality of single-dipped/-peaked disclosure are satisfied and in addition the sender would rather pool any pair of states (with some non-degenerate weights) rather than separating them.

Methodologically, our proofs rely on a novel complementary slackness theorem for persuasion problems, which we present in Appendix B. This theorem is the key to obtaining all of our substantive results, and may be useful in analyzing optimal solutions in other persuasion problems and related models.[2]

Our model and results generalize a great deal of prior literature; we give references throughout the paper, but some of the key prior works that address persuasion models with non-linear preferences (or that can be judiciously mapped to such models) are Rayo and Segal (2010), Friedman and Holden (2008), Alonso and Câmara (2016), Beiglböck and Juillet (2016), Zhang and Zhou (2016), Goldstein and Leitner (2018), and Guo and Shmaya (2019). In light of our analysis, some of the main results in these papers can be viewed as showing that single-dipped or single-peaked disclosure is optimal in some special settings. For instance, Friedman and Holden's (2008) "matching slices" gerrymandering solution, where a gerrymanderer creates electoral districts that pool extreme supporters with similarly extreme opponents, and wins those districts with the most extreme supporters and opponents with the highest probability, is an example of single-dipped disclosure. Goldstein and Leitner's (2018) non-monotone stress tests, where a regulator designs a test that pools the weakest banks that it wants to receive funding with the strongest banks, pools slightly less weak banks with slightly less strong banks, and so on, such that the weakest and strongest banks receive the highest funding, is another such example. On the other hand, Guo and Shmaya's (2019) "nested intervals" disclosure rule, where a designer pools favorable states with similarly unfavorable states, and persuades the receiver to take her preferred action with higher probability at more moderate states, is an example of single-peaked disclosure.

Technically, we rely on linear programming duality and connections to optimal transport. For duality, we build on Kolotilin (2018), which introduces the first-order approach to persuasion and the corresponding strong duality result. The most related strand of the optimal transport literature is that on martingale optimal transport (e.g., Beiglböck, Henry-Labordere, and Penkner 2013, Galichon, Henry-Labordere, and Touzi 2014, Beiglböck and Juillet 2016), which we discuss in Section 4.[3]

[2]Many of our results hold whether states are discrete or continuous; they can be proved without the technical apparatus of the appendix in the discrete case, but not the continuous case. Other results—such as those pertaining to uniqueness or differential equation characterizations—only hold in the continuous case, where the full apparatus is required.

[3]A few recent papers apply optimal transport to persuasion, but these works are not very related to ours either methodologically or substantively. Perez-Richet and Skreta (2022) and Lin and Liu (2022) consider limited sender commitment; Arieli, Babichenko, and Sandomirskiy (2022) and Smolin and Yamashita (2022) consider persuasion with multiple receivers; Malamud and Schrimpf (2022) focus on the question of when optimal signals partition a multidimensional state space.

2. Framework

2.1. **Model.** We consider a standard persuasion problem, where a sender chooses a signal to reveal information to a receiver, who then takes an action. The sender's utility $V(a,\theta)$ and the receiver's utility $U(a,\theta)$ depend on the receiver's action $a \in A := [0,1]$ and the state of the world $\theta \in \overline{\Theta} := [0,1]$. The sender and receiver share a common prior $\phi \in \Delta(\overline{\Theta})$, with support $\Theta := \mathrm{supp}(\phi)$.[4] A *signal* $\tau \in \Delta(\Delta(\Theta))$ is a distribution over posterior beliefs $\mu \in \Delta(\Theta)$ such that the average posterior equals the prior: $\int \mu \mathrm{d}\tau = \phi$ (Aumann and Maschler 1995, Kamenica and Gentzkow 2011). An *outcome* $\pi \in \Delta(A \times \Theta)$ is a joint distribution over actions and states.

We impose four standard assumptions on preferences, which are similar to those in canonical unidimensional models of communication such as signaling (Spence 1973), cheap talk (Crawford and Sobel 1982), and hard information disclosure (Seidmann and Winter 1997).[5] First, utilities are smooth.

Assumption 1. $V(a,\theta)$ and $U(a,\theta)$ are four times differentiable.

We denote the sender's and receiver's marginal utilities in a by

$$v(a,\theta) := \frac{\partial V(a,\theta)}{\partial a} \quad \text{and} \quad u(a,\theta) := \frac{\partial U(a,\theta)}{\partial a},$$

and denote further partial derivatives with subscripts: e.g., $u_a(a,\theta) = \partial u(a,\theta)/\partial a$.

Second, the receiver's expected utility is single-peaked in his action for any posterior belief.

Assumption 2. $U(a,\theta)$ satisfies *strict aggregate quasi-concavity* in a: for all posteriors $\mu \in \Delta(\Theta)$,

$$\int u(a,\theta)\mathrm{d}\mu = 0 \implies \int u_a(a,\theta)\mathrm{d}\mu < 0.$$

Quah and Strulovici (2012) and Choi and Smith (2017) characterized a weak version of aggregate quasi-concavity in terms of primitive conditions on u. We provide an analogous characterization of strict aggregate quasi-concavity in Appendix A. A sufficient condition is that $u_a(a,\theta) < 0$ for all (a,θ), so that U is strictly concave in a. In

[4]Throughout, for any compact metric space X, $\Delta(X)$ denotes the set of Borel probability measures on X, endowed with the weak* topology. For any $\eta \in \Delta(X)$, its support $\mathrm{supp}(\eta)$ is the smallest compact set of measure one.

[5]Not all of these assumptions are necessary for all of our results. We clarify which assumptions are needed for each result in the appendix.

fact, Appendix A shows that strict aggregate quasi-concavity is equivalent to strict concavity up to a normalization.

Third, the receiver's optimal action satisfies an interiority condition.[6]

Assumption 3. $\min_{\theta\in\overline{\Theta}} u(0,\theta) = \max_{\theta\in\overline{\Theta}} u(1,\theta) = 0$.

The key implication of Assumptions 1–3 is that for any posterior μ, the receiver's optimal action $a^\star(\mu) := \arg\max_{a\in[0,1]} \int U(a,\theta)\mathrm{d}\mu$ is unique and is characterized by the first-order condition

$$\int u(a^\star(\mu),\theta)\mathrm{d}\mu = 0.$$

Our assumptions thus allow a "first-order approach" to the persuasion problem, similar to the approach of Mirrlees (1999) and Holmström (1979) to the classical moral hazard problem.[7] The conditions under which the first-order approach is valid in the persuasion problem (Assumptions 1–3) are much simpler than those in the classical moral hazard problem (e.g., Rogerson 1985, Jewitt 1988).[8] The first-order approach is valid in many persuasion problems but not all of them: for example, it is not valid in the price discrimination problem of Bergemann, Brooks, and Morris (2015), where the receiver's (seller's) utility is not quasi-concave. We also note that uniqueness of the receiver's optimal action implies that any signal τ induces a unique outcome π_τ through the map $\mu \mapsto a^\star(\mu)$.

Fourth, the sender prefers higher actions, and the receiver's utility is supermodular.

Assumption 4. $v(a,\theta) > 0$ and $u_\theta(a,\theta) > 0$.

Together with Assumptions 1–3, Assumption 4 ensures that for each action a there is a unique state $\theta^\star(a)$ such that $u(a,\theta^\star(a)) = 0$ (i.e., the receiver's optimal action at $\theta^\star(a)$ is a), and that $\theta^\star(a)$ is a strictly increasing continuous function from A onto $\overline{\Theta}$.

[6]The substance of Assumption 3 is that for each θ, there exists a such that $u(a,\theta) = 0$. Note that it can never be optimal for the receiver to take any a such that $u(a,\theta)$ has a constant sign for all θ. We can then remove all such a from A and renormalize A to $[0,1]$, so that Assumption 3 holds.

[7]The first-order approach to persuasion was introduced in Kolotilin (2018).

[8]Intuitively, this is because in classical moral hazard there are contracts that make the agent's problem non-quasi-concave, which must be ruled out, while in persuasion the receiver's problem is often quasi-concave for any posterior.

A common interpretation of the receiver's action $a \in [0,1]$ is that the receiver has a private type and makes a binary choice—say, whether to accept or reject a proposal—and a is the receiver's choice of a cutoff type below which he accepts. This interpretation is especially useful for some special cases of the model, as we see next.[9]

2.2. **Special Cases.** We list some leading special cases of the model, which we return to periodically to illustrate our results. This subsection can be skimmed (or skipped) without loss of continuity.

(1) The *linear case* (Gentzkow and Kamenica 2016): $u(a,\theta) = \theta - a$ and $V(a,\theta) = V(a)$. That is, $a^\star(\mu) = \mathbb{E}_\mu[\theta]$ and V is state-independent. This is the well-studied case where the sender's indirect utility from inducing posterior μ is $V(\mathbb{E}_\mu[\theta])$.

(2) The *linear receiver case* (Beiglböck, Henry-Labordere, and Penkner 2013): $u(a,\theta) = \theta - a$ but V is arbitrary (e.g., possibly state-dependent). Here the receiver's preferences are as in the linear case, while the sender's preferences are general.

(2a) The *separable subcase* (Rayo and Segal 2010): $V(a,\theta) = w(\theta)G(a)$ with $w > 0$ and $G > 0$. An interpretation of this subcase is that the receiver has a private type t with distribution G and accepts a proposal iff $\mathbb{E}_\mu[\theta] \geq t$, and the sender's utility when the proposal is accepted is $w(\theta)$. Rayo and Segal focused on the sub-subcase with the uniform distribution $G(a) = a$.[10]

(2b) The *translation-invariant subcase* (Beiglböck and Juillet 2016): $V(a,\theta) = P(a - \theta)$. An interpretation of this subcase is that the receiver "values" the proposal at $\mathbb{E}_\mu[\theta]$, and the sender's utility depends on the amount by which the proposal is "over-valued," $\mathbb{E}_\mu[\theta] - \theta$. For example, a school may care about the extent to which its students are over- or under-placed. These preferences are similar to those in Goldstein and Leitner's (2018) model of stress tests (see Section 6.3).

---

[9]To spell out this interpretation, let $g(t|\theta)$ be the conditional density of the receiver's type $t \in [0,1]$ given the state $\theta \in [0,1]$. The sender's and receiver's utilities from rejection are normalized to zero. The sender's and receiver's utilities from acceptance are functions $\tilde{v}(t,\theta)$ and $\tilde{u}(t,\theta)$, with $\tilde{u}(t,\theta)g(t|\theta)$ satisfying Assumption 2. For $a \in [0,1]$ (interpreted as the cutoff such that the receiver accepts iff $t \leq a$), we recover our model with $V(a,\theta) = \int_0^a \tilde{v}(t,\theta)g(t|\theta)\mathrm{d}t$ and $U(a,\theta) = \int_0^a \tilde{u}(t,\theta)g(t|\theta)\mathrm{d}t$.

[10]Rayo and Segal (2010) assume that the state (ω,θ) is two-dimensional, and the sender's and receiver's marginal utilities are $v(a,\theta,\omega) = \omega$ and $u(a,\theta) = \theta - a$. They assume that there are finitely many states (θ,ω), so generically the sender's utility can be written as $v(a,\theta) = w(\theta)$. Rayo (2013), Nikandrova and Pancs (2017), and Onuchic and Ray (2022) consider the separable subcase where θ is continuous and (ω,θ) is supported on the graph of $\theta \to w(\theta)$, albeit Rayo and Onuchic and Ray restrict attention to monotone partitions. Tamura (2018), Kramkov and Xu (2022), and Dworczak and Kolotilin (2022) allow more general distributions of $(\omega,\theta) \in \mathbb{R}^2$.

(3) The *state-independent sender case* (Friedman and Holden 2008): $V(a, \theta) = V(a)$ but u is arbitrary. Here the sender's preferences are as in the linear case, while the receiver's preferences are general.

(3a) The *separable subcase*: $u(a, \theta) = I(\theta)(\theta - a)$, with $I > 0$. This subcase extends the linear case by letting the receiver put more weight on some states than others.

(3b) The *translation-invariant subcase*: $u(a, \theta) = T(\theta - a)$, with $T(0) = 0$. An example that fits this subcase is that the sender's utility when the proposal is accepted is 1, and accepting the proposal corresponds to the receiver undertaking a project that can either succeed or fail, where the receiver's payoff is $1 - \kappa$ when the project succeeds and $-\kappa$ when it fails (and 0 when it is not undertaken), with $\kappa \in (0, 1)$. The difficulty of the project is $1 - \theta$, the receiver's ability is $1 - t$, the receiver's "bad luck" ε has distribution J, and the project succeeds iff $1 - \theta \leq 1 - t - \varepsilon$, or equivalently $\varepsilon \leq \theta - t$. This example fits the current subcase with V equal to the distribution of t and $T(\theta - a) = J(\theta - a) - \kappa$.

(3c) The *quantile sub-subcase*: $u(a, \theta) = \mathbf{1}\{\theta \geq a\} - \kappa$, with $\kappa \in (0, 1)$. This subcase corresponds to the previous example with $J(\theta - a) = \mathbf{1}\{\theta \geq a\}$, so the project succeeds iff the receiver's ability exceeds the project's difficulty. While u is now discontinuous, we admit this subcase as a limit of the translation-invariant case. Friedman and Holden (2008) focused on the translation-invariant case where T is a continuous approximation of the step function $\mathbf{1}\{\theta \geq a\} - 1/2$. Yang and Zentefis (2023) also study the quantile sub-subcase.

We explain the connection between our model and Beiglböck, Henry-Labordere, and Penkner (2013), Beiglböck and Juillet (2016), and Friedman and Holden (2008) following Theorem 3, which is the closest point of contact with their results.

2.3. **Primal and Dual Programs.** Duality approaches to persuasion are well-established. Here we formulate the sender's primal and dual problems and the relevant duality theorem for our setup.

The sender's (primal) problem is to choose an outcome $\pi \in \Delta(A \times \Theta)$ to

$$\text{maximize} \quad \int_{A \times \Theta} V(a, \theta) \mathrm{d}\pi(a, \theta) \tag{P}$$

$$\text{subject to} \quad \int_{A \times \widetilde{\Theta}} \mathrm{d}\pi(a, \theta) = \int_{\widetilde{\Theta}} \mathrm{d}\phi(\theta), \quad \text{for all measurable } \widetilde{\Theta} \subset \Theta, \tag{P1}$$

$$\int_{\tilde{A} \times \Theta} u(a, \theta) \mathrm{d}\pi(a, \theta) = 0, \quad \text{for all measurable } \widetilde{A} \subset A. \tag{P2}$$

(P1) is the *feasibility* constraint that the marginal of π on Θ equals the prior, ϕ. (P2) is the *obedience* constraint that the receiver's action is $a^{\star}(\mu)$ at each posterior μ. An outcome π that violates (P2) is inconsistent with optimal play by the receiver, as there exists $\tilde{A} \subset A$ such that the receiver's play is suboptimal conditional on the event $\{a \in \tilde{A}\}$. Conversely, for any outcome π that satisfies (P1) and (P2), if the sender designs a mechanism that draws (a, θ) according to π and recommends action a to the receiver, it is optimal for the receiver to obey the recommendation. We therefore say that an outcome is *implementable* iff it satisfies (P1) and (P2), and *optimal* iff it solves (P).

Problem (P) is related to—but not the same as—the standard optimal transport (Monge-Kantorovich) problem (e.g., Villani 2009). In optimal transport, two marginal distributions are given (e.g., of men and women, or workers and firms), and the problem is to find an optimal joint distribution with the given marginals. In persuasion, the marginal distribution over states is given (by the prior ϕ), and the problem is to find an optimal joint distribution with this marginal (so (P1) holds), where for each action the conditional distribution over states satisfies obedience (so (P2) holds).

The dual problem is to find a continuous function $p : \Theta \to \mathbb{R}$ and a bounded, measurable function $q : A \to \mathbb{R}$ to

$$\text{minimize} \quad \int_{\Theta} p(\theta) \mathrm{d}\phi(\theta) \tag{D}$$

$$\text{subject to} \quad p(\theta) \geq V(a, \theta) + q(a)u(a, \theta), \quad \text{for all } (a, \theta) \in A \times \Theta. \tag{D1}$$

We say that (p, q) is *feasible* iff it satisfies (D1), and *optimal* iff it solves (D). The interpretation of the dual problem is that $p(\theta)$ is the shadow price of state θ; $q(a)$ is the value of relaxing the obedience constraint at action a; and the dual constraint (D1) says that $p(\theta)$ is no less than the sender's value from assigning state θ to any action a, where this value is the sum of the sender's utility, $V(a, \theta)$, and the product of $q(a)$ and the amount by which the obedience constraint at a is relaxed when state θ is assigned to action a, $u(a, \theta)$.

A preliminary result is that strong duality holds: solutions to (P) and (D) exist and give the same value.[11] We say that a continuous price function p *solves* (D) iff there exists a bounded, measurable function q such that (p, q) is a solution to (D).

[11]This result—Lemma 1—is similar to Lemmas 1 and 2 of Kolotilin (2018). We give a more detailed alternative proof that applies under slightly weaker assumptions. The proof in Kolotilin (2018) uses the Banach-Alaoglu theorem, as in linear programming references such as Anderson and Nash (1987),

Lemma 1. *There exists an outcome π that solves (P), there exists a price function p that solves (D), and the values of (P) and (D) are the same: for any solutions π of (P) and p of (D), we have*

$$\int_{A\times\Theta} V(a,\theta)\mathrm{d}\pi(a,\theta) = \int_{\Theta} p(\theta)\mathrm{d}\phi(\theta).$$

Duality yields an equation that plays a key role in our analysis. Intuitively, complementary slackness implies that the support of any outcome π is contained in the set of points (a,θ) that satisfy (D1) with equality. Thus, if it is ever optimal for the sender to induce action a at state θ—i.e., if a maximizes $V(a,\theta)+q(a)u(a,\theta)$—then a must satisfy the first-order condition

$$v(a,\theta)+q(a)u_a(a,\theta)+q'(a)u(a,\theta)=0. \tag{1}$$

We will use this equation frequently.[12]

3. Pairwise Disclosure and the Twist Condition

Our first result is that there always exist optimal signals that never pool more than two states in a single signal realization, and that under an additional condition every optimal signal has this property. This result simplifies persuasion to a generalized matching problem, where the sender chooses what pairs of states to match together, and with what weights.

Formally, a signal τ is *pairwise* if it induces posterior beliefs with at most binary support: $|\operatorname{supp}(\mu)|\leq 2$ for each $\mu\in\operatorname{supp}(\tau)$. For example, with a uniform prior, for any cutoff $\hat\theta\in[0,1]$ the signal that reveals states below the cutoff and pools each pair of states θ and $1+\hat\theta-\theta$ for $\theta\in[\hat\theta,(1+\hat\theta)/2]$ to induce posterior $\mu=\delta_\theta/2+\delta_{1+\hat\theta-\theta}/2$ is pairwise. The special case where $\hat\theta=1$ is full-disclosure, which is also pairwise. In contrast, no-disclosure, where $\tau(\phi)=1$, is not pairwise.[13]

while our proof uses the Arzela-Ascoli theorem, as in optimal transport references such as Villani (2009) and Santambrogio (2015). A key step in the proof (Lemma 24)—which was left somewhat implicit in Kolotilin—is showing that q may be assumed bounded in (D). Dworczak and Kolotilin (2022) prove a related duality result which allows the receiver's action to be multi-dimensional but requires Lipschitz continuity of the sender's indirect utility. Other (less related) duality results for persuasion problems include those of Dworczak and Martini (2019), Dizdar and Kováč (2020), Kramkov and Xu (2022), and Galperti, Levkun, and Perego (2022).

[12] Appendix B addresses technical issues such as defining the set of points (a,θ) where (1) holds, defining the function q in (1) when (D) has multiple solutions, and defining (1) at points where q is not differentiable.

[13] See Figure 2. The "disclose-pair" pattern in Panel d. is reminiscent of this example, but with different weights on the states in each pair.

If the receiver's utility is not quasi-concave, pairwise signals may be suboptimal. For example, suppose the sender rules three castles, one of which is undefended. The state θ—the identity of the undefended castle—is uniformly distributed. Suppose the receiver can attack any two castles, and payoffs are $(-1, +1)$ for the sender and receiver, respectively, if the receiver attacks the undefended castle, and are $(+1, -1)$ otherwise. Then any pairwise signal narrows the set of possibly undefended castles to at most two, so the receiver always wins. But if the sender discloses nothing, the receiver wins only with probability $2/3$.[14]

In contrast, pairwise signals are without loss under Assumptions 1–3. Moreover, equation (1) implies that if it is optimal to induce the same action a at three states θ_1, θ_2, and θ_3, then the vector $(v(a, \theta_1), v(a, \theta_2), v(a, \theta_3))$ must be a linear combination of the vectors $(u(a, \theta_1), u(a, \theta_2), u(a, \theta_3))$ and $(u_a(a, \theta_1), u_a(a, \theta_2), u_a(a, \theta_3))$. This observation gives a condition—which we call the *twist condition*—under which pooling more than two states is suboptimal, so that every optimal signal is pairwise.

Twist Condition *For any action a and any triple of states $\theta_1 < \theta_2 < \theta_3$ such that $\theta_1 < \theta^\star(a) < \theta_3$, we have*

$$|S| := \begin{vmatrix} v(a, \theta_1) & v(a, \theta_2) & v(a, \theta_3) \\ u(a, \theta_1) & u(a, \theta_2) & u(a, \theta_3) \\ u_a(a, \theta_1) & u_a(a, \theta_2) & u_a(a, \theta_3) \end{vmatrix} \neq 0.^{15} \tag{2}$$

We will apply this condition extensively in Section 4.

Theorem 1. *For any signal τ (whether optimal or not), there exists a pairwise signal $\hat{\tau}$ that induces the same outcome. Moreover, if the twist condition holds, then every optimal signal is pairwise.*

The intuition for the first part of the theorem is that for any posterior, there exists a hyperplane passing through it such that all posteriors on the hyperplane induce the same action, and the extreme points of the hyperplane in the simplex have at most binary support. Thus, any posterior that puts weight on more than two states can be split into posteriors with at most binary support without affecting the induced distribution on $A \times \Theta$. Figure 1 illustrates this argument for a posterior with weight on three states.

[14]Pairwise signals are also suboptimal in the price-discrimination problem of Bergemann, Brooks, and Morris (2015).

[15]Here $|\cdot|$ denotes the determinant of a matrix; we also use the same notation for the cardinality of a set.

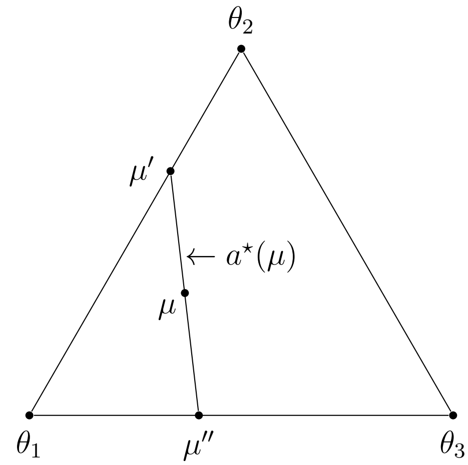

FIGURE 1. Pairwise Signals are Without Loss

Notes: The optimal action at any posterior on the line between μ' and μ'' equals $a^\star(\mu)$, so splitting μ into μ' and μ'' eliminates a non-binary-support posterior without changing the outcome.

The intuition for the second part is that this splitting leaves an extra degree of freedom, which can be profitably exploited under the twist condition. Consider a posterior distribution μ with $\text{supp}(\mu) = \{\theta_1, \theta_2, \theta_3\}$. We can split μ into posterior distributions μ' and μ'' with at most binary support that both induce action $a^\star(\mu)$. For example, suppose that $\text{supp}(\mu') = \{\theta_1, \theta_2\}$ and $\text{supp}(\mu'') = \{\theta_1, \theta_3\}$. Consider a perturbation that moves probability mass $\mathrm{d}p$ on θ_1 from μ' to μ''. This perturbation induces non-zero marginal changes in the receiver's action at μ' and μ''. Under the twist condition, these changes have a non-zero marginal effect on the sender's expected utility, by the implicit function theorem. Therefore, either this perturbation or the reverse perturbation, where $\mathrm{d}p$ is replaced with $-\mathrm{d}p$, is strictly profitable.[16]

Prior results by Rayo and Segal (2010), Alonso and Câmara (2016), and Zhang and Zhou (2016) also give conditions under which all optimal signals are pairwise. Theorem 1 easily implies these earlier results.[17] Note that the twist condition always fails in the linear case (i.e., $|S| = 0$). Hence, in the linear case, Theorem 1 never rules

[16]Formally, the second part of Theorem 1 directly follows from our complementary slackness theorem, Theorem 7.

[17]Proposition 4 in Alonso and Câmara (2016) states that if $u(a, \theta) = \theta - a$ and there do not exist $\zeta \leq 0$ and $\iota \in \mathbb{R}$ such that $v(a, \theta_i) = \zeta\theta_i + \iota$ for $i = 1, 2, 3$, then it is not optimal to induce action a at states θ_1, θ_2, and θ_3. This result is too strong as stated, and it is not correct unless ζ is also allowed to be positive. Theorem 1 implies this corrected version of Alonso and Câmara's result.

out pooling multiple states, and indeed pooling multiple states is often optimal (e.g., Kolotilin, Mylovanov, Zapechelnyuk, and Li 2017).[18]

An immediate corollary of Theorem 1 is that no disclosure is generically suboptimal when there are at least three states, because for a fixed action a a generic vector $(v(a,\theta))_{\theta\in\Theta}$ with $|\Theta| \geq 3$ coordinates cannot be expressed as a linear combination of two fixed vectors $(u(a,\theta))_{\theta\in\Theta}$ and $(u_a(a,\theta))_{\theta\in\Theta}$, as is required by (1).

Corollary 1. *For any prior ϕ with $|\operatorname{supp}(\phi)| \geq 3$ and any receiver utility U, no disclosure is suboptimal for generic sender utility V.*

Given Kamenica and Gentzkow's concavification result, Corollary 1 implies that, for generic direct utilities U and V, the sender's indirect utility is not concave in the posterior when there are more than two states. Also, observe that Corollary 1 allows the case where u and v always have the opposite sign, so the sender's and receiver's ordinal preferences over actions are diametrically opposed. Hence, even in this case no-disclosure is generically suboptimal.

4. Single-Dipped and Single-Peaked Disclosure

The next two sections present our main results, which characterize optimal disclosure patterns. The current section asks when it is optimal for the sender to induce higher actions at more or less extreme states: that is, when optimal outcomes are "single-dipped" or "single-peaked." As we will see, this is a key question, which unifies and generalizes much of what is known about special cases of the persuasion problem with non-linear preferences, as well as related models outside the persuasion literature.[19]

Formally, a triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3) where $\theta_1 < \theta_2 < \theta_3$ is *single-dipped* (*-peaked*) if $a_1 \geq (\leq) a_2$, and is *strictly single-dipped* (*-peaked*) if $a_1 > (<) a_2$. A set $B \subset A \times \Theta$ is *single-dipped* (*-peaked*) if it does not contain a strictly single-peaked (-dipped) triple of points, and is *strictly single-dipped* (*-peaked*) if it does not contain a

[18]Of course, Theorem 1 shows that even when pooling multiple states is optimal, there also exists an optimal pairwise signal, where the "multi-state pool" is split into pairs. Conversely, if multiple posteriors all induce the same action, they can be pooled without affecting the outcome.

[19]Beiglböck and Juillet (2016) argue that single-dippedness/-peakedness are canonical properties analogous to positive/negative assortativity in standard matching models. Mathematically, positive/negative assortativity corresponds to monotonicity in the FOSD order, while single-dippedness/-peakedness corresponds to monotonicity in a variability order that depends on u; when $u(a,\theta) = \theta - a$, this variability order is the usual convex order.

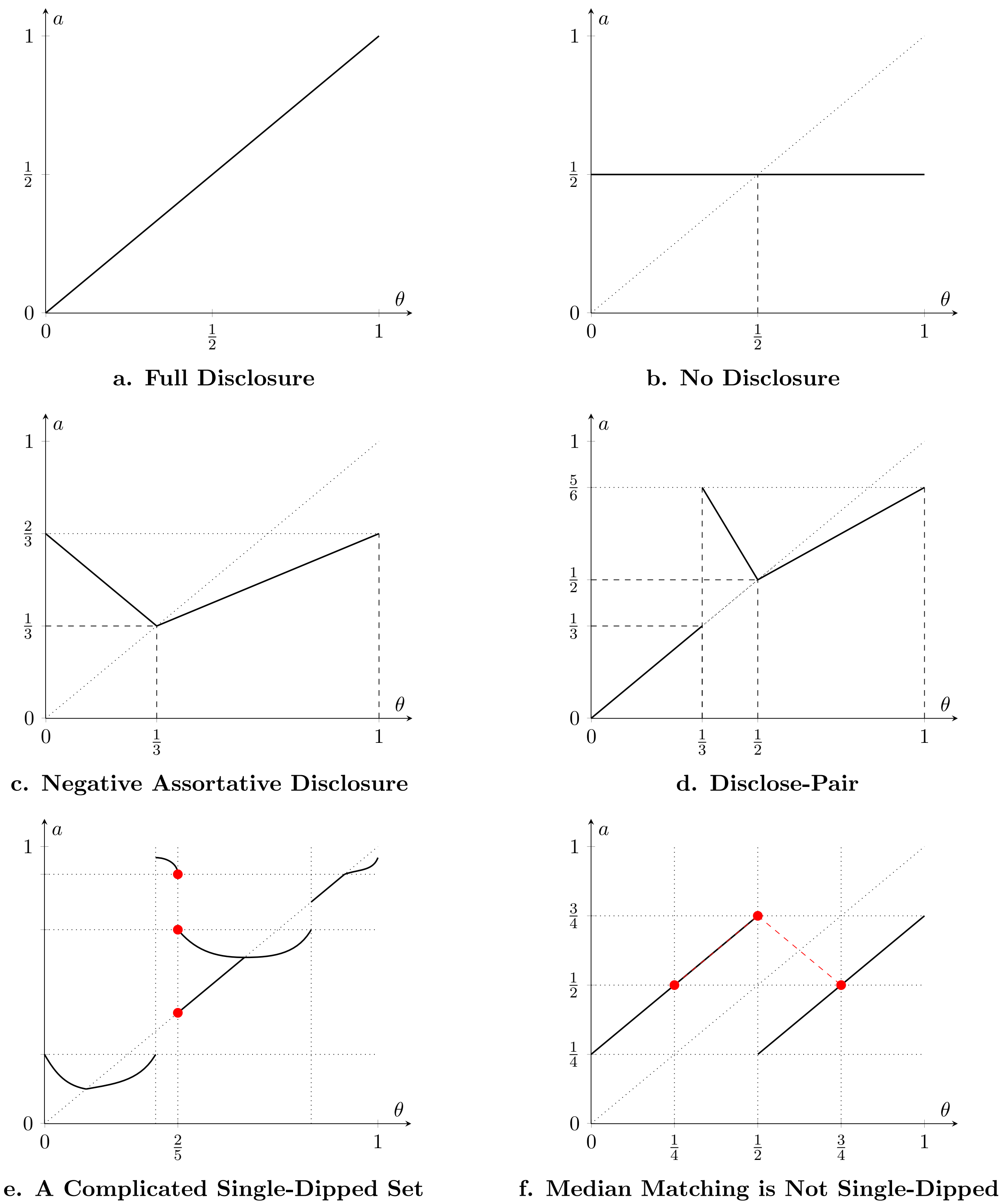

FIGURE 2. Some Single-Dipped Disclosure Patterns

single-peaked (-dipped) triple. Finally, an outcome π is (*strictly*) *single-dipped* if it is concentrated on a (strictly) single-dipped set,[20] and similarly for single-peakedness.[21]

Remark 1. *A strictly single-dipped (-peaked) outcome can be described by the two functions t_1 and t_2 that specify the states $t_1(a)$ and $t_2(a)$ which are pooled together to induce each action a. Specifically, for any implementable strictly single-dipped (-peaked) outcome π, there exist two functions $t_1, t_2 : A \to \overline{\Theta}$ satisfying $t_1(a) \leq \theta^\star(a) \leq t_2(a)$, $t_2(a) \leq t_2(a')$, and $t_1(a') \notin (t_1(a), t_2(a))$ ($t_1(a) \leq t_1(a')$, and $t_2(a) \notin (t_1(a'), t_2(a'))$) for all $a < a'$ such that π is concentrated on the graphs of t_1 and t_2.*[22]

Each panel in Figure 2 illustrates an implementable outcome in the linear receiver case ($u(a, \theta) = \theta - a$). Panel a. is full disclosure, which is trivially strictly single-dipped, as no states are paired. Panel b. is no disclosure, which is single-dipped but not strictly single-dipped. Panels c., d., and e. are all strictly single-dipped. Panel c. is an example of negative assortative disclosure, where state $\theta = 1/3$ is disclosed and the other states are paired with weight $2/3$ on the higher state in each pair. Panel d. shows an outcome where all states below $1/3$ (as well as state $1/2$) are disclosed, and the other states are paired with weight $3/4$ on the higher state in each pair. This "disclose-pair" pattern is a strictly single-dipped analogue of *upper-censorship*, where all states below a cutoff are disclosed, and all states above the cutoff are pooled (e.g., Kolotilin, Mylovanov, and Zapechelnyuk 2022). Upper-censorship is only weakly single-dipped; disclose-pair splits up the pooling region in upper-censorship to obtain strict single-dippedness. Panel e. is a more complicated strictly single-dipped outcome. While strict single-dippedness implies that each action is induced at at most two states, Panel e. shows that more than two actions can be induced at a single state (here, state $2/5$).[23] Panel f. shows "matching across the median" (e.g., Kremer and Maskin 1996), which is not single-dipped, e.g. because it contains the strictly single-peaked triple $\{(1/4, 1/2), (1/2, 3/4), (3/4, 1/2)\}$.

4.1. **Variational Theorem.** The next result captures the core economic logic behind single-dippedness/-peakedness. It is also our key tool for determining when

[20]That is, if there exists a Borel (strictly) single-dipped set B such that $\pi(B) = 1$.

[21]Most of our results for single-dippedness/-peakedness are symmetric, in which case we provide proofs only for the single-dipped case.

[22]This remark follows from Corollary 1.6 and Lemma A.6 of Beiglböck and Juillet (2016). We prove a related result in the first paragraph of the proof of Theorem 8.

[23]Also, while the function t_2 is always monotone under strict single-dippedness, Panel e. shows that the function t_1 can be non-monotone.

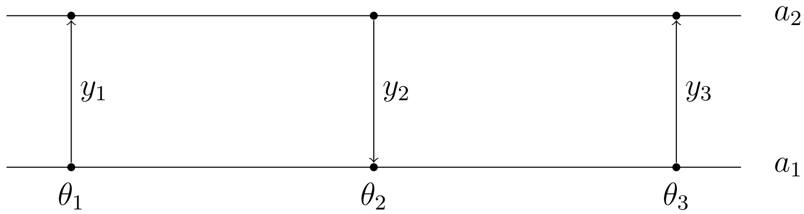

FIGURE 3. A Profitable Perturbation of a Non-Single-Dipped Outcome

Notes: The figure shows a perturbation of an outcome that shifts weights y_1 and y_3 on θ_1 and θ_3 from a_1 to a_2, and shifts weight y_2 on θ_2 from a_2 to a_1. This perturbation is profitable if it increases the receiver's expected marginal utility at a_1 and a_2 and also increases the sender's expected utility for fixed a_1 and a_2.

optimal outcomes are single-dipped/-peaked: we use it to establish our main sufficient condition for single-dipped/-peaked disclosure to be optimal (Theorem 3, in the next subsection), and also use it directly to study some applications in Section 6.[24]

Theorem 2. *Suppose that for any pair of actions $a_1 < (>) a_2$ and any triple of states $\theta_1 < \theta_2 < \theta_3$ such that $\theta_1 \le \theta^\star(a_1) \le \theta_3$, there exists a vector $y \ge 0$ such that $Ry \ge 0$ and $Ry \ne 0$, where*

$$R := \begin{pmatrix} V(a_2,\theta_1) - V(a_1,\theta_1) & -(V(a_2,\theta_2) - V(a_1,\theta_2)) & V(a_2,\theta_3) - V(a_1,\theta_3) \\ -u(a_1,\theta_1) & u(a_1,\theta_2) & -u(a_1,\theta_3) \\ u(a_2,\theta_1) & -u(a_2,\theta_2) & u(a_2,\theta_3) \end{pmatrix}.$$

Then every optimal outcome is single-dipped (-peaked).

The intuition behind Theorem 2 is very simple, and is illustrated in Figure 3. The condition for single-dippedness says that an outcome that assigns positive probability to a strictly single-peaked triple (a_1, θ_1), (a_2, θ_2), (a_1, θ_3) can be improved by re-allocating mass y_1 on θ_1 and mass y_3 on θ_3 from a_1 to a_2, while re-allocating mass y_2 on θ_2 from a_2 to a_1. This re-allocation is profitable for the sender, because the sender's expected utility increases when a_1 and a_2 are held fixed (i.e., the first coordinate of Ry is non-negative); the receiver's marginal utility conditional on being recommended a_1 increases (i.e., the second coordinate of Ry is non-negative), which increases the receiver's action, and hence increases the sender's expected utility; and the receiver's

[24]Theorem 2 provides conditions under which *every* optimal outcome is single-dipped/-peaked. In addition, Lemma 19 in Appendix D.5 provides weaker conditions under which *some* optimal outcome has this property.

marginal utility conditional on being recommended a_2 also increases (i.e., the third coordinate of Ry is non-negative), which again increases the sender's expected utility. Moreover, at least one of these improvements is strict (i.e., $Ry \neq 0$). The same logic applies for any outcome whose support contains a strictly single-peaked triple, even if this triple occurs with 0 probability, except now mass must be re-allocated from small intervals around θ_1, θ_2, and θ_3.[25]

4.2. **Sufficient Conditions for Single-Dipped/-Peaked Disclosure.** We can now give our main sufficient condition on utilities for single-dipped/-peaked disclosure to be optimal. This is a central result of our paper. As we will see, our condition covers several prior models, as well as some new applications.

Theorem 3. *If $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ and $v_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ are increasing (decreasing) in θ for any a and $a_1 \leq (\geq) a_2$, then there exists an optimal single-dipped (-peaked) outcome.*

Moreover, if in addition either $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ or $v_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ is strictly increasing (decreasing) in θ for any a and $a_1 \leq (\geq) a_2$, then every optimal outcome is strictly single-dipped (-peaked).

The proof establishes single-dippedness/-peakedness by constructing perturbations that satisfy the conditions in Theorem 2, and further establishes strict single-dippedness/-peakedness by verifying the twist condition from Theorem 1.

The intuition for Theorem 3 is relatively straightforward in the linear receiver and state-independent sender cases. (See Figure 4.) In the linear receiver case, $u_{a\theta}(a,\theta)/u_\theta(a,\theta) = 0$ and $v_\theta(a_2,\theta)/u_\theta(a_1,\theta) = v_\theta(a_2,\theta)$, so our sufficient conditions for single-dipped disclosure to be optimal are satisfied iff v is convex in θ.[26] To see why, note that for any strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3), the perturbation that moves mass on θ_1 and θ_3 from a_1 to a_2 and moves mass on θ_2 in the opposite direction, so as to hold fixed the receiver's marginal utility conditional on being recommended either action, has the effect of also holding fixed the probability of each recommendation, while spreading out the state conditional on action a_2 and concentrating the state conditional on action a_1. This perturbation is profitable when the difference $V(a_2,\theta) - V(a_1,\theta)$ is convex in θ, which holds whenever v is convex in θ.

[25]Formally, this step crucially relies on our complementary slackness theorem, Theorem 7.

[26]In the separable and translation-invariant subcases, convexity of v simplifies to convexity of w and P', respectively.

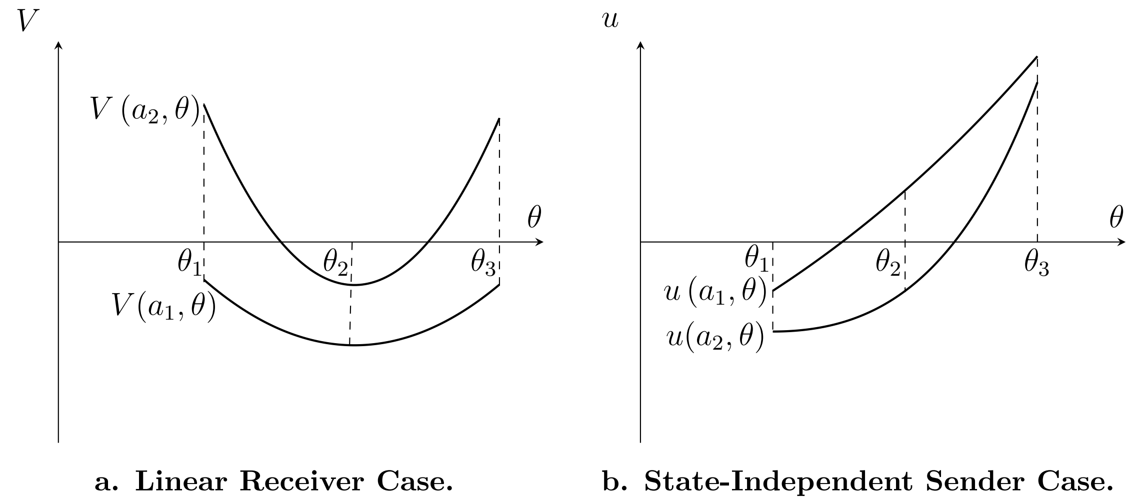

FIGURE 4. The Intuition for Theorem 3 in Two Special Cases

Notes: Panel a. In the linear receiver case, when the sender's utility increment $V(a_2,\theta) - V(a_1,\theta)$ is convex in the state, more extreme states should induce higher actions.
Panel b. In the state-independent sender case, when the receiver's marginal utility $u(a,\theta)$ is more convex in the state at higher actions, more extreme states should induce higher actions.

In the state-independent sender case, $v_\theta(a_2,\theta)/u_\theta(a_1,\theta) = 0$, so our sufficient conditions for single-dipped disclosure to be optimal are satisfied iff u_θ is log-supermodular in (a,θ), or equivalently u is more log-convex in θ at higher actions a.[27] To see why, note that for any strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3), the perturbation that moves mass on θ_1 and θ_3 from a_1 to a_2 and moves mass on θ_2 in the opposite direction, so as to hold fixed the receiver's marginal utility conditional on being recommended a_1 as well as the total probability of each recommendation, has the effect of increasing the receiver's marginal utility conditional on being recommended a_2. This follows because, by log-supermodularity of u_θ, for the receiver's expected marginal utility the marginal rate of substitution between "shifting weight from θ_1 to θ_2" and "shifting weight from θ_2 to θ_3" is higher at a_1 than a_2. Finally, when V is state-independent, this perturbation increases the sender's expected utility.[28]

[27]In the separable and translation-invariant subcases, log-supermodularity of u_θ simplifies to $2I'(\theta)^2 \geq I(\theta)I''(\theta)$ and log-concavity of T', respectively.

[28]In the linear receiver and state-independent sender cases, the sufficient conditions for the optimality of strict single-dipped/-peaked disclosure in Theorem 3 are "almost necessary," because the condition $|S| \neq 0$ on $A \times \overline{\Theta}$ implies that $|S|$ has a constant sign on $A \times \overline{\Theta}$, which can be shown to be equivalent

There are close antecedents to the conditions in Theorem 3 for the linear receiver and state-independent sender cases (albeit outside of the persuasion literature). A branch of the optimal transport literature known as *martingale optimal transport* studies the problem of finding an optimal joint distribution of two variables (say, a and θ) with given marginals, subject to the martingale constraint $\mathbb{E}[\theta|a] = a$ for all a. This problem coincides with our linear receiver case, but with an exogenously fixed distribution of the receiver's action. In this literature, Beiglböck and Juillet (2016) introduce the notions of single-dipped/-peaked outcomes under the names "left-curtain/right-curtain couplings," and show that these outcomes are optimal when the planner's (sender's) marginal utility is convex in θ—a condition referred to in this literature as the "martingale Spence-Mirrlees condition."[29] Earlier, in a model of *partisan gerrymandering*, Friedman and Holden (2008) show that under an "informative signal property," if it is optimal to assign two voter types to the same district, then all voter types in between these two must be assigned to districts with less favorable median voters. Partisan gerrymandering is equivalent to the state-independent sender case (as the map-maker cares only about winning seats, and not directly about the composition of districts), the above property of districting is equivalent to single-dippedness, and the informative signal property is equivalent to log-supermodularity of u_θ.[30] Theorem 3 thus unifies and generalizes these disparate contributions.

We also establish an additional result in Appendix C: under our conditions for strictly single-dipped/-peaked disclosure to be optimal (and an additional regularity condition), the optimal outcome is unique.[31]

---

to strict convexity/concavity of v in the linear receiver case, and to strict log-supermodularity/log-submodularity of u_θ in the state-independent sender case. By Theorem 1, a necessary condition for the optimality of strictly single-dipped/-peaked disclosure is that $|S| \neq 0$ on the restricted domain where $\theta_1 < \theta^\star(a) < \theta_3$.

[29] More precisely, Beiglböck and Juillet (2016) show that the unique optimal outcome is single-dipped in the translation-invariant subcase if P' is strictly convex (Theorem 6.1), and in the separable subcase if w is strictly convex (Theorem 6.3). Theorem 5.1 in Henry-Labordère and Touzi (2016) and Theorem 3.3 in Beiglböck, Henry-Labordère, and Touzi (2017) extend this conclusion to the general linear receiver case where $v(a, \theta)$ is strictly convex in θ. In all these papers, the marginal distribution over actions is fixed.

[30] We further investigate the connection between gerrymandering and persuasion in a companion paper, Kolotilin and Wolitzky (2020).

[31] This result is somewhat akin to *Brenier's theorem* in optimal transport, which shows that the optimal transport plan is unique under a suitable complementarity-type condition, called the twist or generalized Spence-Mirrlees condition (Brenier 1991, Gangbo and McCann 1996; or see Section 1.3 in Santambrogio 2015).

5. Full Disclosure and Negative Assortative Disclosure

While single-dippedness/-peakedness is an important property, many outcomes are single-dipped/-peaked.[32] It is thus also important to go beyond single-dippedness/-peakedness and fully characterize optimal outcomes, when this is tractable. The current section does this for the polar cases of "maximum" and "minimal" disclosure: the former case corresponds to *full disclosure*, where each state is disclosed; while the latter case correponds to *negative assortative disclosure*, where all states are paired in a negatively assortative manner, so all signal realizations can be ordered from least to most extreme. Here our results on full disclosure extend existing results, whereas our results on negative assortative disclosure are entirely novel.

A note on terminology: in Section 4, we considered "assortativity" between states and actions, asking whether higher actions should be induced at more or less extreme states. In the current section, "negative assortative disclosure" refers to assortativity between pairs of states. One can also view full disclosure as capturing "positive assortativity" between states, by matching identical states to form degenerate "pairs."

5.1. **Full Disclosure.** *Full disclosure* is when each state θ induces action $a^\star(\delta_\theta)$.

If for all states θ_1 and θ_2, and all probabilities ρ, the sender prefers to split the posterior $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ into degenerate posteriors δ_{θ_1} and δ_{θ_2}, then the sender prefers full disclosure to any pairwise signal. Since pairwise signals are without loss by Theorem 1, full disclosure is then optimal. Conversely, if the sender strictly prefers not to split $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ into δ_{θ_1} and δ_{θ_2} for some states θ_1 and θ_2 and some probability ρ, then the sender strictly prefers the pairwise signal that differs from full disclosure only in that it pools states θ_1 and θ_2 into μ; so full disclosure is not optimal.[33] Recalling that belief $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ induces action $a^\star(\mu)$ satisfying $\rho u(a^\star(\mu), \theta_1) + (1-\rho)u(a^\star(\mu), \theta_2) = 0$, we obtain the following result.

Theorem 4. *Full disclosure is optimal iff, for all $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ with $\theta_1 < \theta_2$ in Θ and $\rho \in (0,1)$, we have*

$$\rho V(a^\star(\mu), \theta_1) + (1-\rho)V(a^\star(\mu), \theta_2) \leq \rho V(a^\star(\delta_{\theta_1}), \theta_1) + (1-\rho)V(a^\star(\delta_{\theta_2}), \theta_2). \quad (3)$$

Moreover, full disclosure is uniquely optimal if (3) holds with strict inequality for all such μ.

[32] Recall the variety of single-dipped patterns in Figure 2.

[33] This argument is valid when ϕ has finite support. The general case (Theorem 4) uses duality and is adaptated from part (2) of Proposition 1 in Kolotilin (2018); we give a simpler proof based on our complementary slackness theorem (Theorem 7) and also establish uniqueness.

In the linear case, condition (3) holds iff V is convex in a. In the state-independent sender case, condition (3) simplifies as follows:

Corollary 2. *In the state-independent sender case, full disclosure is optimal iff, for all $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ with $\theta_1, \theta_2 \in \Theta$ and $\rho \in (0,1)$, we have*

$$V\left(a^\star\left(\mu\right)\right) \leq \rho V\left(a^\star\left(\delta_{\theta_1}\right)\right) + \left(1-\rho\right) V\left(a^\star\left(\delta_{\theta_2}\right)\right). \tag{4}$$

In a classical one-to-one matching model, Becker (1973) showed that if the utility from matching two types $h\left(\theta_1, \theta_2\right)$ is supermodular, then it is optimal to match like types. Legros and Newman (2002) refer to this extreme form of positive assortative matching as *segregation*. Their Propositions 4 and 9 show that segregation is optimal iff $h\left(\theta_1, \theta_1\right) + h\left(\theta_2, \theta_2\right) \geq 2h\left(\theta_1, \theta_2\right)$ for all θ_1, θ_2 (which is a strictly weaker property than supermodularity). In the context of persuasion, segregation corresponds to full disclosure. Note that if we fix $p = 1/2$ and let $h\left(\theta_1, \theta_2\right) = V\left(a^\star\left(\delta_{\theta_1}/2 + \delta_{\theta_2}/2\right)\right)$, then (4) reduces to Legros and Newman's condition. Intuitively, full disclosure is "less likely" to be optimal in persuasion than segregation is in classical matching, because in persuasion the designer has an extra degree of freedom ρ in designing matches.

In the linear receiver case, there is a simple sufficient condition for (3):

Corollary 2'. *In the linear receiver case, full disclosure is optimal if $V(a,\theta)$ is convex in a and satisfies $V(\theta_1, \theta_2) + V(\theta_2, \theta_1) \leq V(\theta_1, \theta_1) + V(\theta_2, \theta_2)$ for all $\theta_1, \theta_2 \in \Theta$.*

A sufficient condition for $V(\theta_1, \theta_2) + V(\theta_2, \theta_1) \leq V(\theta_1, \theta_1) + V(\theta_2, \theta_2)$ is supermodularity of V: for all $\theta_1 < \theta_2$ and $a_1 < a_2$, $V(a_1, \theta_1) + V(a_2, \theta_2) \geq V(a_1, \theta_2) + V(a_2, \theta_1)$. Thus, in the linear receiver case, full disclosure is optimal whenever the sender's utility is convex in a and supermodular in (a, θ). This sufficient condition for full disclosure generalizes that given by Rayo and Segal (2010) for the separable subcase.[34]

When the prior has full support and the the twist condition holds, full disclosure is uniquely optimal whenever it is optimal. To see the intuition, suppose full disclosure is optimal, and suppose there is another optimal signal that pools some states θ_1 and θ_2 to induce an action a. Then the signal that discloses all other states while pooling θ_1 and θ_2 to induce a is also optimal. But then the signal that discloses all

[34]Their condition is that w is increasing in θ and G is convex in a, where $V(a,\theta) = w(\theta)G(a)$. In the sub-subcase with $G(a) = a$, (3) holds iff w is increasing in θ, because (3) simplifies to $\rho(1-\rho)(w(\theta_2) - w(\theta_1))(\theta_2 - \theta_1) \geq 0$.

other states while pooling θ_1, θ_2, and the third state $\theta^\star(a) \neq \theta_1, \theta_2$ to induce a is also optimal, and this signal is not pairwise.[35]

Theorem 5. *Assume that* $\Theta = [0,1]$ *and the twist condition holds (e.g., the sufficient condition in Theorem 3 holds). If full disclosure is optimal, then it is uniquely optimal.*

5.2. **Negative Assortative Disclosure.** An implementable outcome π is *single-dipped (-peaked) negative assortative* if it is concentrated on the graphs of two functions $t_1, t_2 : [\underline{a}, \overline{a}] \to \Theta$ for some $\underline{a} < \overline{a}$, where $t_1(a) \leq \theta^\star(a) \leq t_2(a)$ for all a, t_1 is decreasing (increasing), and t_2 is increasing (decreasing).

The main result of this section is that if strictly single-dipped (-peaked) disclosure is optimal and the sender strictly prefers to pool any two states, then single-dipped (-peaked) negative assortative disclosure is optimal. Moreover, if the prior has a density, then the optimal outcome is unique (by Theorem 8 in Appendix C) and is characterized as the solution to a system of two ordinary differential equations.

To see the intuition, note that if strictly single-dipped disclosure is optimal, then any two pairs of pooled states $\{\theta_1, \theta_3\}$ and $\{\theta_1', \theta_3'\}$ with (without loss) $\theta_1 < \theta_3$, $\theta_1' < \theta_3'$, and $\theta_1 \leq \theta_1'$, must be either ordered (i.e., $\theta_1 < \theta_3 \leq \theta_1' < \theta_3'$) or nested (i.e., $\theta_1 \leq \theta_1' < \theta_3' \leq \theta_3$). This follows because if the pairs overlap (i.e., $\theta_1 < \theta_1' < \theta_3 < \theta_3'$), then either $(\theta_1, \theta_1', \theta_3)$ or $(\theta_1', \theta_3, \theta_3')$, together with the corresponding actions, would form a single-peaked triple. Hence, for any pair of pooled states $\{\theta_1, \theta_3\}$, there must exist a disclosed state $\theta_2 \in (\theta_1, \theta_3)$: intuitively, there must exist pairs of pooled states in the interval (θ_1, θ_3) that are closer and closer together, until the pair degenerates into a single disclosed state. Therefore, if any two pairs of pooled states $\{\theta_1, \theta_3\}$ and $\{\theta_1', \theta_3'\}$ are ordered, there would exist two distinct disclosed states $\theta_2 \in (\theta_1, \theta_3)$ and $\theta_2' \in (\theta_1', \theta_3')$. But if the sender strictly prefers to pool any two states, this is impossible. Finally, if pairs of pooled states cannot overlap or be ordered, the only remaining possibility is that all pairs of pooled states are nested: that is, disclosure is negative assortative.[36]

Theorem 6. *Assume that the condition for strict single-dippedness (-peakedness) given in Theorem 3 holds and that the prior* ϕ *has positive density on* $[0,1]$*. If for all*

[35]The proof relies on complementary slackness.

[36]In this argument, the existence of the two disclosed states relies on the assumption that $\mathrm{supp}(\phi) = [0,1]$. The formal proof again relies on complementary slackness.

$\theta_1 < \theta_2$ there exists $\rho \in (0,1)$ such that

$$\rho V(a^\star(\mu),\theta_1) + (1-\rho)V(a^\star(\mu),\theta_2) > \rho V(a^\star(\delta_{\theta_1})),\theta_1) + (1-\rho)V(a^\star(\delta_{\theta_2}),\theta_2), \quad (5)$$

with $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$, then the unique optimal outcome is single-dipped (-peaked) negative assortative, and is described by the functions t_1 and t_2 that solve the system of differential equations,

$$u(a,t_1(a))(-\mathrm{d}\phi([0,t_1(a)])) + u(a,t_2(a))\mathrm{d}\phi([0,t_2(a)]) = 0, \quad (6)$$

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{v(a,t_1(a))u(a,t_2(a)) - v(a,t_2(a))u(a,t_1(a))}{u(a,t_1(a))u_a(a,t_2(a)) - u(a,t_2(a))u_a(a,t_1(a))}\right)\\ &\quad = \frac{v(a,t_1(a))u_a(a,t_2(a)) - v(a,t_2(a))u_a(a,t_1(a))}{u_a(a,t_1(a))u(a,t_2(a)) - u_a(a,t_2(a))u(a,t_1(a))},\end{aligned} \quad (7)$$

for all $a \in (\underline{a}, \overline{a}]$, with the boundary conditions

$$\begin{aligned}&(t_1(\overline{a}), t_1(\underline{a}), t_2(\underline{a}), t_2(\overline{a})) = (0, \theta^\star(\underline{a}), \theta^\star(\underline{a}), 1)\\ &((t_1(\underline{a}), t_1(\overline{a}), t_2(\overline{a}), t_2(\underline{a})) = (0, \theta^\star(\overline{a}), \theta^\star(\overline{a}), 1)).\end{aligned} \quad (8)$$

Similarly to equation (3) in the previous subsection, equation (5) simplifies in special cases. In the linear case, (5) holds iff V is strictly concave in a.[37] In the state-independent sender case, it holds iff $V(a^\star(\mu)) > \rho V(a^\star(\delta_{\theta_1})) + (1-\rho)V(a^\star(\delta_{\theta_2}))$. In the linear receiver case, it holds if $V(a,\theta)$ is concave in a and satisfies $V(\theta_1,\theta_2) + V(\theta_2,\theta_1) > V(\theta_1,\theta_1) + V(\theta_2,\theta_2)$ for all $\theta_1 < \theta_2$; a sufficient condition for the latter property is strict submodularity of V. These conditions generalize the sufficient condition for pooling given by Rayo and Segal (2010) for the separable subcase.[38]

To understand the differential equations, note that if t_1 and t_2 are differentiable then (6) can be written as

$$u(a,t_1(a))f(t_1(a))t_1'(a) = u(a,t_2(a))f(t_2(a))t_2'(a).$$

This is the obedience constraint conditional on recommendation a, as the posterior conditional on recommendation a is

$$\pi_a = \frac{-f(t_1(a))t_1'(a)}{-f(t_1(a))t_1'(a) + f(t_2(a))t_2'(a)}\delta_{t_1(a)} + \frac{f(t_2(a))t_2'(a)}{-f(t_1(a))t_1'(a) + f(t_2(a))t_2'(a)}\delta_{t_2(a)}.$$ [39]

[37] In the linear case, V is strictly concave iff no-disclosure is uniquely optimal for all priors, by Corollary 1 in Kolotilin, Mylovanov, and Zapechelnyuk (2022).

[38] Their condition is that w is strictly decreasing in θ and G is concave in a, where $V(a,\theta) = w(\theta)G(a)$. In the sub-subcase with $G(a) = a$, (5) holds iff w is strictly decreasing in θ.

[39] This equation is a version of the Monge-Ampere equation in optimal transport (e.g., Section 1.7.6 in Santambrogio 2015).

In addition, (7) results from solving the system of equations (from the sender's FOC, (1)),

$$v(a, t_1(a)) + q(a)u_a(a, t_1(a)) + q'(a)u(a, t_1(a)) = 0,$$
$$v(a, t_2(a)) + q(a)u_a(a, t_2(a)) + q'(a)u(a, t_2(a)) = 0,$$

for $q(a)$ and $q'(a)$, and recalling that q' is the derivative of q. Finally, the boundary condition (8) for the single-dipped case follows because the lowest induced action $\underline{a}$ is induced at the disclosed state $\theta^\star(\underline{a}) = t_1(\underline{a}) = t_2(\underline{a})$, and the highest induced action $\overline{a}$ is induced at states $0 = t_1(\overline{a})$ and $1 = t_2(\overline{a})$. The boundary condition for the single-peaked case is analogous.[40]

We can also give primitive conditions on V and u for (5) to hold, and hence for the unique optimal outcome to be negative assortative.

Corollary 3. *Assume that the condition for strict single-dippedness (-peakedness) given in Theorem 3 holds. Then for all $\theta_1 < \theta_2$ there exists $\rho \in (0,1)$ such that (5) holds iff, for all $a \in A$,*

$$v_a(a, \theta^\star(a)) \leq \frac{v(a,\theta^\star(a))u_{aa}(a,\theta^\star(a))}{u_a(a,\theta^\star(a))} + 2\frac{v_\theta(a,\theta^\star(a))u_a(a,\theta^\star(a)) - v(a,\theta^\star(a))u_{a\theta}(a,\theta^\star(a))}{u_\theta(a,\theta^\star(a))}. \tag{9}$$

Equation (9) is a local necessary condition for (5): if (9) fails, then (5) also fails for $\theta_1 < \theta_2$ sufficiently close to $\theta^\star(a)$. When the condition for strict single-dippedness (-peakedness) holds, this local necessary condition turns out to be globally sufficient for (5). Equation (9) simplifies dramatically in some special cases. In the linear receiver case, (9) simplifies to $v_a(a,a) + 2v_\theta(a,a) \leq 0$; in the translation-invariant subcase of the linear receiver case, this simplifies further to $P''(0) \geq 0$. In the separable (resp., translation-invariant) subcase of the state-independent sender case, (9) simplifies to $v_a(a)/v(a) \leq 2I'(a)/I(a)$ (resp., $v_a(a)/v(a) \leq T''(0)/T'(0)$).

We give some examples of optimal single-dipped negative assortative disclosure.[41]

Example 1. Consider the linear receiver case with $A = \Theta = [1/e, e]$, $f(\theta) = 1/(2\theta)$, and $V(a,\theta) = a/\theta$. We claim that the unique optimal outcome matches each state

[40]In the linear receiver case, (7) simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(v(a,t_1(a))\frac{t_2(a)-a}{t_2(a)-t_1(a)} + v(a,t_2(a))\frac{a-t_1(a)}{t_2(a)-t_1(a)}\right) = -\frac{v(a,t_2(a)) - v(a,t_1(a))}{t_2(a)-t_1(a)}.$$

Geometrically, this says that the slope of the curve $a \to \mathbb{E}_\pi[v(a,\theta)|a]$ is equal to the negative of the slope of the secant passing through the points $(t_1(a), v(a,t_1(a)))$ and $(t_2(a), v(a,t_2(a)))$. Nikandrova and Pancs (2017) derive this condition for the separable sub-subcase with $v(a,\theta) = w(\theta)$.

[41]In Examples 1 and 3, Θ and A are compact intervals, which can be rescaled to the unit interval.

$\theta_1 \in [1/e, 1]$ with state $\theta_2 = 1/\theta_1$ with equal weights, so that the induced action is $a = \theta_1/2 + 1/(2\theta_1)$. Thus, $t_1(a) = a - \sqrt{a^2-1}$, and $t_2 = a + \sqrt{a^2-1}$ for all $a \in [1, e/2 + 1/(2e)]$.

Indeed, by Theorem 3, the optimal outcome is strictly single-dipped, since $w(\theta) = 1/\theta$ is strictly convex. By Corollary 3, (5) holds, since $w' < 0$. Hence, by Theorem 6, the optimal outcome is single-dipped negative assortative and satisfies (6)–(8). Now, for $\theta_2 = 1/\theta_1$ and $a = \theta_1/2 + 1/(2\theta_1)$, (6) holds because

$$u(a,\theta_2) = \left(\frac{1}{2\theta_1} - \frac{\theta_1}{2}\right) = -\left(\frac{\theta_1}{2} - \frac{1}{2\theta_1}\right) = -u(a,\theta_1),$$
$$f(\theta_2)\frac{\mathrm{d}\theta_2}{\mathrm{d}a} = \frac{1}{2/\theta_1}\left(-\frac{1}{\theta_1^2}\frac{\mathrm{d}\theta_1}{\mathrm{d}a}\right) = -\frac{1}{2\theta_1}\frac{\mathrm{d}\theta_1}{\mathrm{d}a} = -f(\theta_1)\frac{\mathrm{d}\theta_1}{\mathrm{d}a},$$

(7) holds because

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(w(\theta_1)\frac{1}{2} + w(\theta_2)\frac{1}{2}\right) = \frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{1}{2\theta_1} + \frac{\theta_1}{2}\right) = \frac{\mathrm{d}}{\mathrm{d}a}a = 1,$$
$$\frac{w(\theta_2) - w(\theta_1)}{\theta_2 - \theta_1} = \frac{\theta_1 - 1/\theta_1}{1/\theta_1 - \theta_1} = -1,$$

and (8) holds because $1/(1/e) = e$ and $1/1 = 1$.[42]

Example 2 (Quantile Persuasion)**.** Consider the quantile sub-subcase of the state-independent sender case, where $u(a,\theta) = \mathbf{1}\{\theta \geq a\} - \kappa$ with $\kappa \in (0,1)$. Let ϕ have a density on $[0,1]$. Assuming that the receiver breaks ties in favor of the sender, we obtain that, for $\theta_1 < \theta_2$,

$$a^\star(\rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}) = \begin{cases} \theta_2, & \rho \leq 1-\kappa, \\ \theta_1, & \rho > 1-\kappa. \end{cases}$$

Note that (5) always holds for $\rho \in (0, 1-\kappa)$. We claim that there exists an optimal single-dipped negative assortative outcome π where the marginal distribution over actions α_π satisfies $\alpha_\pi([a,1]) = \phi([a,1])/\kappa$, and the posterior conditional on any recommendation $a \in [\underline{a}, 1]$ is $(1-\kappa)\delta_{t_1(a)} + \kappa\delta_{t_2(a)}$, where $t_2(a) = a$, $t_1(a)$ solves $\kappa\phi([0, t_1(a)]) = (1-\kappa)\phi([a,1])$, and $\underline{a}$ solves $\kappa\phi([0,\underline{a}]) = (1-\kappa)\phi([\underline{a},1])$.[43] A notable feature of this outcome is that, with the informed receiver interpretation, it

[42]We can instead solve this example by using complementary slackness (Theorem 7) directly, because, for $q(a) = a$, the function $V(a,\theta) + q(a)u(a,\theta) = a/\theta + a(\theta - a)$ is maximized at $a = \theta/2 + 1/(2\theta)$ for all $\theta \in [1/e, e]$.

[43]See Appendix F.7 for the proof.

would remain optimal even if the sender knew the receiver's type and could condition disclosure on it.

Example 3 (A Stochastic Optimal Signal[44])**.** In the following example, for some priors negative assortative disclosure is optimal; and for other priors (perhaps surprisingly), the unique optimal signal randomizes conditional on the state, even though the state is atomless.

Consider the translation-invariant subcase of the state-independent sender case. Let $A = \Theta = [-1, 3]$, let ϕ have a density f with $f(-a) \geq 3f(3a)$ for all $a \in (0, 1]$, let $u(a, \theta) = T(\theta - a)$ with $T(0) = 0$ and strictly log-concave T', and let $V(a, \theta) = T(2a)$. With the informed receiver interpretation, this captures a case where, for example, $\kappa = 1/2$, the distribution of ε is $N(0, \sigma^2)$, and the distribution of t is $N(0, (\sigma/2)^2)$.[45]

We claim that

$$t_1(a) = \begin{cases} a, & a \in [-1, 0], \\ -a, & a \in (0, 1], \end{cases} \quad \text{and} \quad t_2(a) = \begin{cases} a, & a \in [-1, 0), \\ 3a, & a \in (0, 1], \end{cases}$$

so that the posterior conditional on any recommendation $a \in [-1, 1]$ is $\delta_{t_1(a)}/2 + \delta_{t_2(a)}/2$, and the marginal distribution over actions has density h given by

$$h(a) = \begin{cases} 6f(3a), & a \in (0, 1], \\ f(-a) - 3f(3a), & a \in [-1, 0). \end{cases}$$

Note that the unique optimal outcome is single-dipped negative assortative iff $f(-a) = 3f(3a)$ for all $a \in (0, 1]$. In contrast, if $f(-a) > 3f(3a)$ for all $a \in (0, 1]$, then each state $\theta \in [-1, 0)$ is mixed between recommendations $a = \theta$ and $a = -\theta$.[46] See Figure 5.

6. Applications and Extensions

We now show how our analysis covers several well-known applications, where single-dipped or single-peaked disclosure is optimal.[47]

---

[44] This example is an adaptation of Example 2 in Kolotilin and Wolitzky (2020).

[45] By symmetry and strict log-concavity of T', $v_a(a)/v(a) = 2T''(a)/T'(a) > (<) T''(0)/T'(0) = 0$ for $0 > (<) a$, showing that (9) fails for $a < 0$, and thus Theorem 6 does not apply.

[46] See Appendix F.9 for the proof.

[47] We have already described how our results generalize those of Friedman and Holden (2008) in gerrymandering and Beiglböck and Juillet (2016) and others in martingale optimal transport. The applications in this section also illustrate some technical points. Section 6.1 illustrates how directly applying Theorem 2 can yield weaker sufficient conditions for the optimality of single-dipped/-peaked

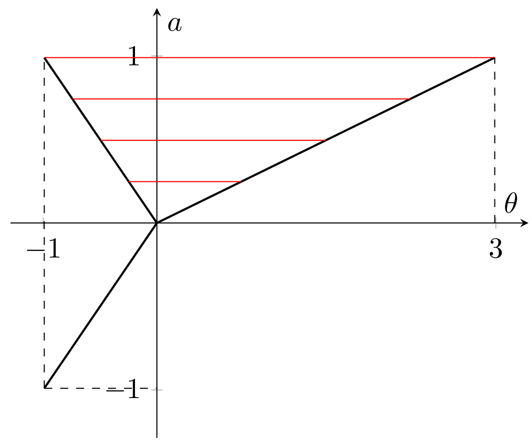

Figure 5. The Optimal Outcome in Example 3

Notes: The support of the optimal outcome is contained in the three black line segments. The red line segments indicate pairs of states that may be pooled at the optimal outcome. If the prior density satisfies $f(-a) > 3f(3a)$ for all $a \in (0, 1]$, the support of the optimal outcome *equals* the three black line segments. In this case, for each state $\theta < 0$, the optimal signal randomizes between disclosing θ (and inducing action θ) and pooling θ with state -3θ (and inducing action $-\theta$).

6.1. **Contests.** Zhang and Zhou (2016) study information disclosure in contests. In their model, two contestants, A and B, compete for a prize by exerting efforts x_A and x_B. The probability that contestant $i = A, B$ wins is $x_i/(x_A + x_B)$. Everyone knows contestant A's value $v_A = 1$. Contestant B's value v_B is known to contestant B and the designer. The sender designs a signal about v_B to maximize expected total effort.

It is convenient to parameterize $\theta = 1/\sqrt{v_B}$ and $a = \sqrt{x_A}$. With this parameterization, Zhang and Zhou's Proposition 1 shows that, given a posterior μ, contestant A exerts effort $x_A^\star = a^\star(\mu)^2$ determined by $\mathbb{E}_\mu\left[\theta - (1+\theta^2)\, a^\star(\mu)\right] = 0$, and contestant B (who knows θ) exerts effort $x^\star(\theta) = a^\star(\mu)/\theta - a^\star(\mu)^2$, so the sender's expected utility is $x_A^\star + \mathbb{E}_\mu\left[x^\star(v_B)\right] = \mathbb{E}_\mu\left[a^\star(\mu)/\theta\right]$. We thus recover our model with $V(a,\theta) = a/\theta$ and $u(a,\theta) = \theta - (1+\theta^2)a$.

Zhang and Zhou give results on optimality of pairwise disclosure, full-disclosure, and no-disclosure. Our approach easily yields the following result, which additionally gives

disclosure than those in Theorem 3. Sections 6.2 and 6.3 illustrate how our analysis extends when some of our assumptions are violated: in Section 6.2, Assumption 3 fails, so the receiver's optimal action may be at the boundary and thus violate the first-order condition; in Section 6.3, Assumption 4 fails, as the sender only weakly prefers higher actions.

conditions for optimality of single-dipped/-peaked disclosure and negative assortative disclosure (which were not considered by Zhang and Zhou).

Proposition 1. *In Zhang and Zhou's contest model where the prior ϕ has a positive density on $\Theta = [\underline{\theta}, \overline{\theta}]$, where $0 < \underline{\theta} < \overline{\theta}$, if $\underline{\theta} \geq 1$ then the unique optimal outcome is full disclosure; and if $\overline{\theta} \leq 1/\sqrt{3}$ ($1/\sqrt{3} \leq \underline{\theta} < \underline{\theta} \leq 1$) then the unique optimal outcome is single-dipped (-peaked) negative assortative disclosure.*

The proof of single-dippedness/-peakedness uses Theorem 2 with a perturbation that fixes both actions. In contrast, directly applying Theorem 3 would yield only the weaker result that single-peaked negative assortative disclosure is optimal if $1/\sqrt{2} \leq \underline{\theta} < \overline{\theta} < 1$.[48]

6.2. **Affiliated Information.** Guo and Shmaya (2019) consider a persuasion model with a privately informed receiver, where it is commonly known that the receiver wishes to accept a proposal iff θ exceeds a threshold θ_0, and the receiver's type t is his private signal of θ. Letting $G(t|\theta)$ denote the distribution of t conditional on θ, with corresponding density $g(t|\theta)$, this setup maps to our model with $V(a, \theta) = G(a|\theta)$, $u(a, \theta) = (\theta - \theta_0)g(a|\theta)$, and $g(t|\theta)$ strictly log-submodular in (t, θ).[49,50] These preferences satisfy Assumptions 1 and 2 (see Lemma 2), but not Assumption 3, as $u(a, \theta) > 0$ for all a when $\theta > \theta_0$. Nonetheless, assuming that the receiver breaks ties in the sender's favor, we have $a^\star(\mu) = \max\{a : \int_\Theta u(a, \theta)\mathrm{d}\mu \geq 0\}$.

Let us take for granted that Theorem 2 holds even though Assumption 3 is violated (e.g., this is clearly true with a discrete prior). Applying Theorem 2 with a perturbation that fixes one action while increasing the other action and the sender's expected utility (for fixed actions), we obtain the following result, which reproduces Guo and Shmaya's main qualitative insight.[51]

Proposition 2. *In Guo and Shmaya's model of persuading a privately informed receiver, every optimal outcome is single-peaked.*

[48]To see this, suppose $\overline{\theta} < 1$. Then $u_\theta(a, \theta) = 1 - 2\theta a > 0$ for $a \leq \overline{\theta}/(1 + \overline{\theta}^2) = \max A$. Moreover, $u_{a\theta}(a, \theta)/u_\theta(a, \theta) = -2\theta/(1 - 2\theta a)$ is always decreasing in θ, while $v_\theta(a_2, \theta)/u_\theta(a_1, \theta) = -1/(\theta^2 - 2\theta^3 a_1)$ is decreasing in θ iff $3\underline{\theta} \min A = 3\underline{\theta}^2/(1 + \underline{\theta}^2) \geq 1$, or equivalently $\underline{\theta} \geq 1/\sqrt{2}$.

[49]The ordering convention here is that high t is bad news about θ. This ordering is opposite to Guo and Shmaya's, but follows our convention that the receiver accepts for types below a cutoff.

[50]Inostroza and Pavan (2022) study robust stress test design in a setting with multiple receivers with coordination motives. As they note, the single-receiver version of their model is a special case of Guo and Shmaya (2019).

[51]When the prior has positive density on $[0, 1]$, Guo and Shmaya's Theorem 3.1 additionally implies that the optimal outcome is single-peaked negative assortative.

6.3. **Stress Tests.** Goldstein and Leitner (2018) consider a model of optimal stress tests. The sender is a bank regulator and the receiver is a perfectly competitive market. The bank has an asset that yields a random cash flow. The asset's quality is θ, which is observed by the bank and the regulator but not the market, and is normalized to equal the asset's expected cash flow.[52] The regulator designs a test to reveal information about θ. After observing the test result, the market offers a competitive price a for the asset. Finally, the bank decides whether to keep the asset and receive the random cash flow, or sell it at price a. Letting z denote the bank's final cash holding (equal to either the random cash flow or a), the bank's payoff equals $z + \mathbf{1}\{z \geq \theta_0\}$, where θ_0 is a constant. An interpretation is that the bank faces a run if its cash holding falls below θ_0. The regulator designs the test to maximize expected social welfare, or equivalently to minimize the probability of a run.

Goldstein and Leitner show that a bank with a type-θ asset is willing to sell at a price a iff a exceeds a reservation price $\tilde{\sigma}(\theta)$ that satisfies $\tilde{\sigma}(\theta) > \theta$ if $\theta < \theta_0$, $\tilde{\sigma}(\theta) < \theta$ if $\theta > \theta_0$, and $\tilde{\sigma}'(\theta) \geq 0$. Intuitively, if $\theta < \theta_0$ then the bank demands a premium to forego the chance that a lucky cash flow shock pushes its holdings above θ_0, while if $\theta > \theta_0$ then the bank desires insurance against bad cash flow shocks that push its holdings below θ_0. However, the value of the regulator's problem is unaffected if the reservation price is re-defined as $\sigma(\theta) = \theta$ if $\theta \leq \theta_0$ and $\sigma(\theta) = \tilde{\sigma}(\theta)$ if $\theta > \theta_0$, because it is suboptimal for the regulator to induce a bank to sell at a price below θ_0. It is more convenient to work with the normalized reservation price $\sigma(\theta)$.

It is also convenient to restrict attention to tests that, for each θ, either induce the bank to sell or fully disclose the bank's value: this is without loss because pooling two asset types that do not sell is weakly worse than disclosing these types. For such a test, the price induced by any posterior μ is $a^\star(\mu) = \mathbb{E}_\mu[\theta]$, so we are in the linear receiver case. We can capture the requirement that the bank always sells if $a \neq \theta$ by setting $V(a, \theta) = -\infty$ if $a < \sigma(\theta)$. Finally, letting $w(\theta) > 0$ equal the social gain when a bank sells a type-θ asset at a price above θ_0 (which equals the probability that a type-θ asset yields a cash flow below θ_0), we obtain the linear receiver case of our model with

$$V(a, \theta) = \begin{cases} w(\theta)\mathbf{1}\{a \geq \theta_0\}, & \text{if } a \geq \sigma(\theta), \\ -\infty, & \text{otherwise.} \end{cases}$$

[52] This is the model in Section 5 of their paper, where the bank observes θ.

Note that V violates Assumptions 1 and 4, as it is discontinuous and only weakly increasing in a. Nonetheless, if we assume a discrete prior (as do Goldstein and Leitner), we recover their main qualitative insight.

Proposition 3. *In Goldstein and Leitner's stress test model with a discrete prior, there exists an optimal single-dipped outcome.*

To prove the proposition, we use a perturbation that fixes both actions. Since V is only weakly increasing, this perturbation now only weakly increases the sender's expected utility. Nonetheless, when Θ is finite, repeatedly applying such perturbations eventually yields a single-dipped outcome. We also note that, as Goldstein and Leitner show, if $\mathbb{E}_\phi[\theta] < \theta_0$—so that no-disclosure does not attain the sender's first-best outcome—then every optimal outcome is single-dipped.[53]

7. Conclusion

This paper has developed a first-order approach to persuasion with non-linear preferences. Our substantive results provide conditions for all optimal signals to be pairwise, for higher actions to be induced at more or less extreme states, and for full or negative assortative disclosure to be optimal. In some cases, we can characterize optimal signals as the solution to a pair of ordinary differential equations, or even solve them in closed form. Methodologically, we develop a novel complementary slackness theorem based on connections to optimal transport (Theorem 7 in Appendix B), which forms the basis of all of our proofs, and may also be useful in related contexts.

We mention a few open issues. First, while the persuasion literature has made progress by allowing unrestricted disclosure policies, the pairwise signals that we have highlighted are not always realistic. (For example, in reality it is probably not feasible to design a stress test that pools only the weakest and strongest banks.) An alternative, complementary approach is to restrict the sender to partitioning the state space into intervals, as in Rayo (2013) or Onuchic and Ray (2022). An interesting observation is that, at least in the separable subcase of our model considered by Rayo

[53] A related model by Garcia and Tsur (2021) studies optimal information disclosure to facilitate trade in an insurance market with adverse selection. Their model can be mapped to the linear receiver case with $V(a,\theta) = \nu(a)$ if $a \geq \sigma(\theta)$ and $V(a,\theta) = -\infty$ otherwise, where $\nu(a)$ is a strictly increasing, strictly concave function, and σ is a continuous, strictly increasing function that satisfies $\sigma(\theta) < \theta$. Considering a similar perturbation as in Goldstein and Leitner shows that single-dipped negative assortative disclosure is optimal in their model. We also mention Leitner and Williams (2022), where a bank regulator discloses information about the design of a stress test to induce banks to make socially desirable investments. In this model, single-peaked disclosure is optimal.

and Onuchic and Ray, our condition (5) is equivalent to the condition that complete pooling is uniquely optimal among monotone partitions for all prior distributions. This suggests that, under our conditions for the optimality of single-dipped/-peaked disclosure, negative assortative disclosure might be the optimal unrestricted disclosure policy for all priors iff no-disclosure is the optimal monotone policy for all priors. More generally, analyzing the relationship between the optimal pairwise signals we have characterized and simpler signals such as monotone partitions is an important direction for future research.

Second, in the informed receiver interpretation of our model mentioned in Section 2, our analysis pertains to disclosure mechanisms that do not first elicit the receiver's type, or *public persuasion* in the language of Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017). Public persuasion turns out to be without loss in Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), as well as in Guo and Shmaya (2019). It would be interesting to investigate conditions for the optimality of public persuasion in our more general model, and in particular to see how they relate to our conditions for the optimality of full or negative assortative disclosure.

Finally, our model could be generalized to allow multidimensional states or actions. We suspect that our results on duality (Lemma 1), pairwise signals (Theorem 1), and complementary slackness (Theorem 7 in Appendix B) generalize up to some technicalities.[54] Generalizing our other results would require a more general notion of single-dippedness/-peakedness. With a unidimensional action and a multidimensional state, one can still define a notion of single-dippedness as inducing higher actions at more extreme states; with multidimensional actions, the appropriate generalization is unclear.[55] For results on multidimensional persuasion focusing on the linear case, see Dworczak and Kolotilin (2022).

References

Aliprantis, C. D., and K. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

Alonso, R., and O. Câmara (2016): "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, 165, 672–706.

[54]For example, our proof of Theorem 7 is facilitated by the existence of a bijection between actions a and states $\theta^\star(a)$ such that $u(a, \theta^\star(a)) = 0$, cf. Assumption 4.

[55]Possibly relevant recent work on multidimensional martingale optimal transport includes Ghoussoub, Kim, and Lim (2019) and De March and Touzi (2019).

Ambrosio, L., E. Brué, and D. Semola (2021): *Lectures on Optimal Transport.* Springer.

Anderson, E. J., and P. Nash (1987): *Linear Programming in Infinite-Dimensional Space.* John Wiley & Sons, New York.

Arieli, I., Y. Babichenko, and F. Sandomirskiy (2022): "Persuasion as Transportation," *Working paper.*

Aumann, R. J., and M. Maschler (1995): *Repeated Games with Incomplete Information.* MIT press.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Beiglböck, M., P. Henry-Labordere, and F. Penkner (2013): "Model-Independent Bounds for Option Prices–a Mass Transport Approach," *Finance and Stochastics*, 17(3), 477–501.

Beiglböck, M., P. Henry-Labordère, and N. Touzi (2017): "Monotone Martingale Transport Plans and Skorokhod Embedding," *Stochastic Processes and their Applications*, 127(9), 3005–3013.

Beiglböck, M., and N. Juillet (2016): "On a Problem of Optimal Transport under Marginal Martingale Constraints," *Annals of Probability*, 44(1), 42–106.

Beiglböck, M., M. Nutz, and N. Touzi (2017): "Complete Duality for Martingale Optimal Transport on the Line," *Annals of Probability*, 45(5), 3038–3074.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

Brenier, Y. (1991): "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions," *Communications on Pure and Applied Mathematics*, 44(4), 375–417.

Choi, M., and L. Smith (2017): "Ordinal Aggregation Results via Karlin's Variation Diminishing Property," *Journal of Economic Theory*, 168, 1–11.

Crawford, V., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

De March, H., and N. Touzi (2019): "Irreducible Convex Paving for Decomposition of Multidimensional Martingale Transport Plans," *Annals of Probability*, 47(3), 1726–1774.

Dizdar, D., and E. Kováč (2020): "A Simple Proof of Strong Duality in the Linear Persuasion Problem," *Games and Economic Behavior*, 122, 407–412.

Dworczak, P., and A. Kolotilin (2022): "The Persuasion Duality," *Working paper.*

Dworczak, P., and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127(5), 1993–2048.

Friedman, J. N., and R. T. Holden (2008): "Optimal Gerrymandering: Sometimes Pack, but Never Crack," *American Economic Review*, 98(1), 113–44.

Gale, D. (1989): *The Theory of Linear Economic Models.* University of Chicago press.

Galichon, A., P. Henry-Labordere, and N. Touzi (2014): "A Stochastic Control Approach to No-Arbitrage Bounds Given marginals, with an Application to Lookback Options," *Annals of Applied Probability*, 24(1), 312–336.

Galperti, S., A. Levkun, and J. Perego (2022): "The Value of Data Records," *Review of Economic Studies*, Forthcoming.

Gangbo, W., and R. McCann (1996): "The Geometry of Optimal Transportation," *Acta Mathematica*, 177(2), 113–161.

Garcia, D., and M. Tsur (2021): "Information Design in Competitive Insurance Markets," *Journal of Economic Theory*, 191, 105–160.

Gentzkow, M., and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review, Papers & Proceedings*, 106, 597–601.

Ghoussoub, N., Y.-H. Kim, and T. Lim (2019): "Structure of Optimal Martingale Transport Plans in General Dimensions," *Annals of Probability*, 47(1), 109–164.

Goldstein, I., and Y. Leitner (2018): "Stress Tests and Information Disclosure," *Journal of Economic Theory*, 177, 34–69.

Guo, Y., and E. Shmaya (2019): "The Interval Structure of Optimal Disclosure," *Econometrica*, 87(2), 653–675.

Henry-Labordère, P., and N. Touzi (2016): "An Explicit Martingale Version of the One-Dimensional Brenier Theorem," *Finance and Stochastics*, 20(3), 635–668.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10(1), 74–91.

Inostroza, N., and A. Pavan (2022): "Adversarial Coordination and Public Information Design," *Working paper.*

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, pp. 1177–1190.

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

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89(4), 1557–1593.

Kolotilin, A. (2018): "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, 13, 607–636.

Kolotilin, A., T. Mylovanov, and A. Zapechelnyuk (2022): "Censorship as Optimal Persuasion," *Theoretical Economics*, 17(2), 561–585.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a Privately Informed Receiver," *Econometrica*, 85, 1949–1964.

Kolotilin, A., and A. Wolitzky (2020): "The Economics of Partisan Gerrymandering," *Working paper*.

Kramkov, D., and Y. Xu (2022): "An Optimal Transport Problem with Backward Martingale Constraints Motivated by Insider Trading," *Annals of Applied Probability*, 32(1), 294–326.

Kremer, M., and E. Maskin (1996): "Wage Inequality and Segregation by Skill," *Working paper*.

Legros, P., and A. F. Newman (2002): "Monotone Matching in Perfect and Imperfect Worlds," *Review of Economic Studies*, 69(4), 925–942.

Leitner, Y., and B. Williams (2022): "Model Secrecy and Stress Tests," *Journal of Finance*, Forthcoming.

Lin, X., and C. Liu (2022): "Credible Persuasion," *Working paper*.

Malamud, S., and A. Schrimpf (2022): "Persuasion by Dimension Reduction," *Working paper*.

Mirrlees, J. A. (1999): "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *Review of Economic Studies*, 66(1), 3–21.

Nikandrova, A., and R. Pancs (2017): "Conjugate Information Disclosure in an Auction with Learning," *Journal of Economic Theory*, 171, 174–212.

Onuchic, P., and D. Ray (2022): "Conveying Value via Categories," *Theoretical Economics*, Forthcoming.

Perez-Richet, E., and V. Skreta (2022): "Test Design Under Falsification," *Econometrica*, 90(3), 1109–1142.

Quah, J., and B. Strulovici (2012): "Aggregating the Single Crossing Property," *Econometrica*, 80, 2333–2348.

Rayo, L. (2013): “Monopolistic Signal Provision,” *BE Journal of Theoretical Economics*, 13(1), 27–58.

Rayo, L., and I. Segal (2010): “Optimal Information Disclosure,” *Journal of Political Economy*, 118, 949–987.

Rogerson, W. P. (1985): “The First-Order Approach to Principal-Agent Problems,” *Econometrica*, pp. 1357–1367.

Santambrogio, F. (2015): *Optimal Transport for Applied Mathematicians*, vol. 55. Springer.

Seidmann, D., and E. Winter (1997): “Strategic Information Transmission with Verifiable Messages,” *Econometrica*, 65, 163–169.

Smolin, A., and T. Yamashita (2022): “Information Design in Concave Games,” *Working paper*.

Spence, M. (1973): “Job Market Signaling,” *Quarterly Journal of Economics*, 87, 355–374.

Tamura, W. (2018): “Bayesian Persuasion with Quadratic Preferences,” *Working paper*.

Villani, C. (2009): *Optimal Transport: Old and New*, vol. 338. Springer.

Winkler, G. (1988): “Extreme Points of Moment Sets,” *Mathematics of Operations Research*, 13(4), 581–587.

Yang, K. H., and A. Zentefis (2023): “Extreme Points of First-Order Stochastic Dominance Intervals: Theory and Applications,” *Working paper*.

Zhang, J., and J. Zhou (2016): “Information Disclosure in Contests: A Bayesian Persuasion Approach,” *Economic Journal*, 126(597), 2197–2217.

A. Characterization of Aggregate Quasi-Concavity

We present two alternative conditions that are equivalent to strict aggregate quasi-concavity of U. Condition (2) is analogous to the “signed-ratio monotonicity” conditions for weak aggregate quasi-concavity in Theorem 1 of Quah and Strulovici (2012) and Corollary 2 of Choi and Smith (2017). We give a shorter proof based on the optimality of pairwise signals (see Appendix F.1). Condition (3) is novel. It corresponds to strict concavity of U (i.e., $u_a(a, \theta) < 0$), up to a normalizing factor $g(a) > 0$.

Lemma 2. *Let Assumption 1 hold. The following statements are equivalent:*

(1) *Assumption 2 holds.*

(2) *For all θ, θ', and a, we have*

$$u(a,\theta) = 0 \implies u_a(a,\theta) < 0, \tag{10}$$

$$u(a,\theta) < 0 < u(a,\theta') \implies u(a,\theta')u_a(a,\theta) - u(a,\theta)u_a(a,\theta') < 0. \tag{11}$$

(3) *There exists a differentiable function $g(a) > 0$ such that $\tilde{u}(a,\theta) = u(a,\theta)/g(a)$ satisfies $\tilde{u}_a(a,\theta) < 0$ for all (a,θ).*

B. Complementary Slackness

This appendix develops a technical result in the spirit of complementary slackness, which forms the basis of all of our proofs. In short, we define a compact set $\Gamma \subset A \times \Theta$ with the properties that an implementable outcome π is optimal iff $\text{supp}(\pi) \subset \Gamma$, and equation (1) holds at any pair (a,θ) in a full-measure subset $\Gamma^\star \subset \Gamma$. In analogy with the optimal transport literature (e.g., Section 3 in Ambrosio, Brué, and Semola 2021), we refer to this set Γ as the *contact set*.

Let $C(\Theta)$ denote the set of continuous functions on Θ, and let $B(A)$ denote the set of bounded, measurable functions on A. Let $p \in C(\Theta)$ be the optimal price function (which we will see is unique), and let I be any sufficiently large compact interval (e.g., as defined in Lemma 24). Let

$$Q(a) := \{r \in I : \; p(\theta) \geq V(a,\theta) + ru(a,\theta) \text{ for all } \theta \in \Theta\}, \quad \text{for all } a \in A.$$

This is the set of possible values for $q(a) \in I$ that satisfy (D1) for all θ, given the optimal price function p. For any selection $q \in B(A)$ from Q, the pair (p,q) is a solution to (D).

By Lemma 1, together with (P1) and (P2), any optimal π and (p,q) satisfy

$$\int_{A\times\Theta} (p(\theta) - V(a,\theta) - q(a)u(a,\theta))\mathrm{d}\pi(a,\theta) = 0.$$

By (D1), the integrand is non-negative, and hence any optimal π is concentrated on the set Γ of points (a,θ) that satisfy (D1) with equality. We call any such set Γ *a contact set*. Note that Γ depends on the selection q from Q.

We will show that q given by

$$q(a) := \begin{cases} -\frac{v(a,\theta^\star(a))}{u_a(a,\theta^\star(a))}, & \theta^\star(a) \in \Theta \text{ and } p(\theta^\star(a)) = V(a,\theta^\star(a)), \\ \frac{\min Q(a) + \max Q(a)}{2}, & \text{otherwise}, \end{cases}$$

is a measurable selection from Q, and the associated contact set Γ given by

$$\Gamma := \{(a,\theta) \in A \times \Theta : p(\theta) = V(a,\theta) + q(a)u(a,\theta)\}$$

has the desired properties. We call this set Γ *the* contact set, to distinguish it from contact sets that result from different choices of q.

Recall that for each a, the a-section of Γ is defined as $\Gamma_a := \{\theta \in \Theta : (a,\theta) \in \Gamma\}$. Intuitively, Γ_a is the set of states that it is optimal to pool together to induce action a. Next, the projection of Γ on A is defined as $A_\Gamma := \{a \in A : (a,\theta) \in \Gamma \text{ for some } \theta \in \Theta\}$. Intuitively, A_Γ is the set of actions that it is ever optimal to induce. Finally, the set $\Gamma^\star \subset \Gamma$ is defined by letting its a-section be given by

$$\Gamma^\star_a := \begin{cases} \{\theta^\star(a)\}, & a \in A_\Gamma \text{ and } \theta^\star(a) \in \{\min \Gamma_a, \max \Gamma_a\}, \\ \Gamma_a, & \text{otherwise}, \end{cases} \qquad \text{for all } a \in A.$$

Intuitively, $\Gamma^\star$ is a subset of Γ that removes "redundant" states from each a-section.

Theorem 7. *Under Assumptions 1–4, the following hold:*

(1) *The set* Γ *is compact and satisfies* $\min \Gamma_a \le \theta^\star(a) \le \max \Gamma_a$ *for all* $a \in A_\Gamma$. *Moreover,* (p,q) *solves (D). Consequently, an implementable outcome* π *solves (P) iff* $\mathrm{supp}(\pi) \subset \Gamma$.

(2) *The set* $\Gamma^\star$ *is a Borel subset of* Γ *and* (1) *holds for all* $(a,\theta) \in \Gamma^\star$, *with the convention that* $q'(a) \cdot 0 = 0$ *even if* q *is not differentiable at* a. *Moreover, an implementable outcome* π *solves (P) iff there exists a conditional probability* π_a *of* π *given* a *such that* $\mathrm{supp}(\pi_a) \subset \Gamma^\star_a$ *and* $\int_\Theta u(a,\theta)\mathrm{d}\pi_a(\theta) = 0$ *for all* $a \in \mathrm{supp}(\alpha_\pi)$, *where* α_π *denotes the marginal distribution of* π *on* A.

Theorem 7 says that there is a compact contact set Γ such that an implementable outcome is optimal iff it is supported on Γ; and there is a measure-1 subset $\Gamma^\star \subset \Gamma$ such that the sender's FOC holds on $\Gamma^\star$. Theorem 7 is our key tool for characterizing optimal outcomes: by showing that points (a,θ) violate (1), we can exclude them from $\Gamma^\star$, and hence from the support of any optimal outcome.

Taking the expectation of (1) with respect to an optimal conditional probability π_a yields a useful formula for $q(a)$:

$$q(a) = -\frac{\mathbb{E}_{\pi_a}[v(a,\theta)]}{\mathbb{E}_{\pi_a}[u_a(a,\theta)]}, \quad \text{for all } a \in \mathrm{supp}(\alpha_\pi). \tag{12}$$

This says that $q(a)$ equals the product of the sender's expected marginal utility at a and the rate at which a increases as the obedience constraint is relaxed, where the latter term equals $-1/\mathbb{E}_{\pi_a}[u_a(a,\theta)]$ by the implicit function theorem applied to the obedience constraint. Note that we defined q so that (12) holds for a where $\pi_a = \delta_{\theta^\star(a)}$ (i.e., for actions induced at disclosed states); here we see that this equation also holds for a where π_a is non-degenerate (i.e., for actions induced at pooled states).

Remark 2. *Under Assumptions 1–4, there is a unique solution p to (D).*[56] *As shown by the examples in Appendix E, while the price function $p \in C(\Theta)$ is unique, there can be multiple functions $q \in B(A)$ such that (p,q) is a solution to (D).*

In the standard optimal transport problem the dual solutions are always unique, because the associated constraints on the two marginal distributions are non-homogenous. In contrast, in the persuasion problem the obedience constraint (P2) is homogeneous, so the multiplier $q \in B(A)$ for (P2) does not enter into the objective of the dual problem (D). Consequently, for any selection $q \in B(A)$ from Q, the pair (p,q) is feasible and thus optimal for (D). Our key technical contribution is to find a selection q from Q such that the associated contact set Γ is compact, and to construct a measure-1 subset $\Gamma^\star \subset \Gamma$ such that the FOC holds on $\Gamma^\star$. These challenges do not arise in standard optimal transport, where uniqueness of the dual solutions implies that there is a unique contact set, and the FOC turns out to hold on the entire contact set.

In particular, by selecting $q(a)$ from the interior of $Q(a)$ (when $p(\theta^\star(a)) > V(a,\theta^\star(a))$ and $Q(a)$ is multivalued), we ensure that A_Γ does not contain any actions a that are "redundant," in the sense that $\theta^\star(a) \notin [\min \Gamma_a, \max \Gamma_a]$—for such actions, $\int_\Theta u(a,\theta)\mathrm{d}\pi_a(\theta) \neq 0$ for all $\pi_a \in \Delta(\Gamma_a)$, so these actions are not induced by any optimal outcome. In turn, $\Gamma^\star$ is obtained from Γ by removing redundant states from each a-section—if $\theta^\star(a) \in \{\min \Gamma_a, \max \Gamma_a\}$, then $\pi_a(\theta^\star(a)) = 1$ for any $\pi_a \in \Delta(\Gamma_a)$ such that $\int_\Theta u(a,\theta)\mathrm{d}\pi_a(\theta) = 0$, so any states $\theta \neq \theta^\star(a)$ can be removed from Γ_a.[57] We provide examples illustrating these and other technical points in Appendix E.

Lemma 1 and Theorem 7 can be compared to results in the literature on *martingale optimal transport (MOT)*. The MOT problem is to find an optimal joint distribution of two variables (say, a and θ) with given marginals, subject to the martingale constraint $\mathbb{E}_{\pi_a}[\theta] = a$ for all a. This problem coincides with our linear receiver case, but with

[56] See Lemma 11.

[57] Thus, $\Gamma \setminus \Gamma^\star$ is the *polar* subset of Γ with respect to (P2), in the sense that a set $\Gamma^0 \subset \Gamma$ satisfies $\pi(\Gamma^0) = 0$ for all $\pi \in \Delta(\Gamma)$ satisfying (P2) iff $\Gamma^0 \subset \Gamma \setminus \Gamma^\star$.

an exogenously fixed distribution of the receiver's action. Motivated by problems in mathematical finance, Beiglböck, Henry-Labordere, and Penkner (2013) (see also Beiglböck, Nutz, and Touzi 2017) introduce MOT and prove that the primal and dual problems have the same value; however, they also show that their dual problem may not have a solution, unlike in our model with endogenous actions. Results in MOT also do not establish compactness of the contact set, which holds in our model as well as in standard optimal transport. Thus, MOT is related to our linear receiver case, but the endogenous action distribution apparently makes our model more tractable.

C. Uniqueness

This appendix presents a notable technical result: under a regularity condition, strict single-dippedness/-peakedness implies that there is a unique optimal outcome.

Theorem 8. *If $\Gamma^\star$ is strictly single-dipped (-peaked), ϕ has a density, and the set $\{a \in A_\Gamma : t_1(a) < t_2(a)\}$ is the union of finitely many intervals, then there is a unique optimal outcome.*

The regularity condition that the set $\{a \in A_\Gamma : t_1(a) < t_2(a)\}$ is the union of finitely many intervals rules out pathological cases, such as when this set is the complement of the Cantor set. This condition is satisfied in every example in the literature that we know of.

In martingale optimal transport, the optimal plan is unique under the martingale Spence-Mirrlees condition (e.g., Proposition 3.5 in Beiglböck, Henry-Labordère, and Touzi 2017), which as noted above coincides with our condition for the optimality of strict single-dippedness in the linear receiver case. The key implication of Theorem 8 is that the optimal marginal distribution of actions α_π is unique; there is no analog of this result in martingale optimal transport, where both marginals are fixed.

To outline the argument, consider the case where ϕ is discrete and t_2 is strictly increasing: when the recommended action is higher, the highest possible state under the induced posterior is also (strictly) higher. Suppose toward a contradiction that there are two distinct optimal outcomes, π and π'. Since every optimal signal is pairwise, we know that π and π' have the same conditional distribution: $\pi_a = \pi'_a = \rho_a \delta_{t_1(a)} + (1-\rho_a)\delta_{t_2(a)}$ for all $a \in A_S$, where ρ_a is pinned down by obedience. Thus, the marginal distributions α_π and $\alpha_{\pi'}$ must differ. So let $\hat{a} = \sup\{a \in A : \alpha_\pi([0,a]) \neq \alpha_{\pi'}([0,a])\}$, and consider the state $t_2(\hat{a})$. Since t_2 is strictly increasing, the state $t_2(\hat{a})$ can only induce actions $a \geq \hat{a}$: thus, $\pi([0,\hat{a}), t_2(\hat{a})) = \pi'([0,\hat{a}), t_2(\hat{a})) = 0$. Since the marginals α_π and $\alpha_{\pi'}$ coincide on $(\hat{a}, 1]$ (by the definition of $\hat{a}$), and the conditionals π_a and π'_a coincide everywhere, we also have $\pi((\hat{a}, 1], t_2(\hat{a})) = \pi'((\hat{a}, 1], t_2(\hat{a}))$. Thus, since $\pi(A, t_2(\hat{a})) = \pi'(A, t_2(\hat{a}))$ by (P1), we can conclude that $\pi(\hat{a}, t_2(\hat{a})) = \pi'(\hat{a}, t_2(\hat{a}))$, and hence

$$0 = \pi(\hat{a}, t_2(\hat{a})) - \pi'(\hat{a}, t_2(\hat{a})) = (1 - \rho_{\hat{a}})(\alpha_\pi(\hat{a}) - \alpha_{\pi'}(\hat{a})).$$

Since $\rho_{\hat{a}} < 1$ by convention, it follows that $\alpha_\pi(\hat{a}) = \alpha_{\pi'}(\hat{a})$, and hence $\alpha_\pi([0, \hat{a})) = \alpha_{\pi'}([0, \hat{a}))$. Finally, when ϕ is discrete, this implies that α_π and $\alpha_{\pi'}$ coincide on $(\hat{a} - \varepsilon, 1]$ for some $\varepsilon > 0$, which contradicts the definition of $\hat{a}$. When instead ϕ has a density and our regularity condition holds, a similar argument delivers the same conclusion. Moreover, when ϕ has a density, the possibility that t_2 may be only weakly increasing does not threaten uniqueness of the optimal joint distribution π, because the set of states corresponding to flat regions of t_2 has measure 0 (i.e., $\phi(\{\theta : \exists a < a' \text{ s.t. } \theta = \theta_2(a) = \theta_2(a')\}) = 0$).

D. Proofs

We prove the results in the following order. First, in the online appendix, we prove Lemmas 2 and 1. Second, in this appendix, we prove Theorems 7, 1–3, 8, and 4–6. Third, in the online appendix, we prove Propositions 1–3.

D.1. **Proof of Theorem 7.** We start by proving Theorem 7, because this result is used in most of the other proofs. However, the proof is fairly long and difficult, and it can be skipped on a first reading. We note that the proof remains valid if Assumption 4 is replaced with the weaker requirement that $u(a, \theta)$ satisfies strict single-crossing in θ: for all a and $\theta < \theta'$, we have $u(a, \theta) \geq 0 \implies u(a, \theta') > 0$.

Let
$$\mathring{q}(a) = \frac{\min Q(a) + \max Q(a)}{2}.$$
Define the set of *a-contact points of type 1* as

$$\Psi_1 = \{a \in A : \ \theta^\star(a) \in \Theta \text{ and } p(\theta^\star(a)) = V(a, \theta^\star(a))\},$$

and the set of *a-contact points of type 2* as

$$\Psi_2 = \{a \in A \setminus \Psi_1 : \exists \theta \in \Theta : p(\theta) = V(a, \theta) + \mathring{q}(a) u(a, \theta)\}.$$

Note that
$$q(a) = \begin{cases} \frac{v(a, \theta^\star(a))}{-u_a(a, \theta^\star(a))}, & a \in \Psi_1, \\ \mathring{q}(a), & \text{otherwise.} \end{cases}$$

Part (1) of the theorem follows from Lemmas 3–8, and part (2) of the theorem follows from Lemmas 9 and 10.

Lemma 3. Γ_a *is non-empty iff* $a \in \Psi_1 \cup \Psi_2$*. That is,* $\Psi_1 \cup \Psi_2 = A_\Gamma$*.*

Proof. Clearly, $\theta^\star(a) \in \Gamma_a$ if $a \in \Psi_1$. By the definition of Ψ_2, Γ_a is non-empty if $a \in \Psi_2$, and Γ_a is empty if $a \notin \Psi_1 \cup \Psi_2$. □

Lemma 4. *(p,q) solves (D).*

Proof. Note that $q \in B(A)$, as follows from the proof of Lemma 24 in the online appendix (measurability of q follows from continuity of p, v, u_a, and $\theta^\star$). Thus, by Lemma 23 in the online appendix, it suffices to show that $q(a_1) \in Q(a_1)$ for each $a_1 \in \Psi_1$: that is,

$$p(\theta) \geq V(a_1, \theta) + q(a_1)u(a_1, \theta) \quad \text{for all } a_1 \in \Psi_1 \text{ and } \theta \in \Theta.$$

Fix any $a_1 \in \Psi_1$ and $\theta \in \Theta$, and let $\theta_1 = \theta^\star(a_1)$. For any $\varepsilon \in (0,1)$, define $a_\varepsilon \in A$ as a unique solution to $(1-\varepsilon)u(a_\varepsilon, \theta_1) + \varepsilon u(a_\varepsilon, \theta) = 0$. By the implicit function theorem,

$$\lim_{\varepsilon \downarrow 0} \frac{a_\varepsilon - a_1}{\varepsilon} = \frac{u(a_1, \theta)}{-u_a(a_1, \theta_1)}.$$

By (D1), we have

$$V(a_1, \theta_1) \geq V(a_\varepsilon, \theta_1) + \mathring{q}(a_\varepsilon)u(a_\varepsilon, \theta_1) \quad \text{and} \quad p(\theta) \geq V(a_\varepsilon, \theta) + \mathring{q}(a_\varepsilon)u(a_\varepsilon, \theta).$$

Adding the first inequality multiplied by $1-\varepsilon$ and the second inequality multiplied by ε, and taking into account the definition of a_ε, we get

$$p(\theta) \geq V(a_1, \theta) + \frac{(1-\varepsilon)[V(a_\varepsilon, \theta_1) - V(a_1, \theta_1)] + \varepsilon[V(a_\varepsilon, \theta) - V(a_1, \theta)]}{\varepsilon}.$$

Taking the limit $\varepsilon \to 0$ gives

$$p(\theta) \geq V(a_1, \theta) + \frac{v(a_1, \theta_1)}{-u_a(a_1, \theta_1)} u(a_1, \theta) = V(a_1, \theta) + q(a_1)u(a_1, \theta). \qquad \square$$

Lemma 5. *For each* $a \in \Psi_1$*, we have* $\inf \Gamma_a \leq \theta^\star(a) \leq \sup \Gamma_a$*. For each* $a \in \Psi_2$*, we have* $\theta^\star(a) \notin \Gamma_a$*,* $\inf \Gamma_a < \theta^\star(a) < \sup \Gamma_a$*, and* $\mathring{q}(a) = \min Q(a) = \max Q(a)$*.*

Proof. We have $\theta^\star(a) \in \Gamma_a \subset [\inf \Gamma_a, \sup \Gamma_a]$ for each $a \in \Psi_1$, by the definition of Ψ_1. Fix $a \in \Psi_2$. By the definition of Ψ_2, we have $a \notin \Psi_1$, so

$$p(\theta^\star(a)) > V(a, \theta^\star(a)) = V(a, \theta^\star(a)) + \mathring{q}(a)u(a, \theta^\star(a)),$$

showing that $\theta^\star(a) \notin \Gamma_a$. By the definition of Ψ_2, Γ_a is non-empty, so it contains some $\theta \neq \theta^\star(a)$. Suppose for concreteness that $\theta > \theta^\star(a)$, so we write $\theta = \theta_+$ (the case $\theta < \theta^\star(a)$ is analogous and omitted). By the definition of $\mathring{q}(a)$ and $Q(a)$,

$$\mathring{q}(a) \leq \max Q(a) \leq \frac{p(\theta_+) - V(a, \theta_+)}{u(a, \theta_+)},$$

and, by the definition of Γ_a,

$$\mathring{q}(a) = \frac{p(\theta_+) - V(a, \theta_+)}{u(a, \theta_+)}.$$

Hence,

$$\mathring{q}(a) = \max Q(a) = \frac{p(\theta_+) - V(a, \theta_+)}{u(a, \theta_+)}.$$

Then,

$$\mathring{q}(a) = \min Q(a) = \sup_{\tilde{\theta} < \theta^\star(a)} \frac{V(a, \tilde{\theta}) - p(\tilde{\theta})}{-u(a, \tilde{\theta})},$$

where the first equality is by the definition of $\mathring{q}(a)$ and $\mathring{q}(a) = \max Q(a)$, and the second equality is by the definition of $Q(a)$. (Inspecting the definition gives $\min Q(a) = \sup_{\tilde{\theta}: u(a,\tilde{\theta})<0} (V(a, \tilde{\theta}) - p(\tilde{\theta}))/(-u(a, \tilde{\theta}))$, and $u(a, \tilde{\theta}) < 0$ iff $\tilde{\theta} < \theta^\star(a)$.)

Since p, V, and u are continuous and since $p(\theta^\star(a)) > V(a, \theta^\star(a))$, the supremum is attained at some $\theta_- < \theta^\star(a)$. Thus $\theta_- \in \Gamma_a$, by the definition of Γ_a. The lemma follows since $\inf \Gamma_a \leq \theta_- < \theta^\star(a) < \theta_+ \leq \sup \Gamma_a$. □

Lemma 6. *For each $a \in \Psi_1 \cup \Psi_2$ such that $\inf \Gamma_a < \theta^\star(a) < \sup \Gamma_a$, the function q has a derivative $q'(a)$, and (1) holds for all $\theta \in \Gamma_a$.*

Proof. Fix $a \in \Psi_1 \cup \Psi_2$ such that there exist $\theta_-, \theta_+ \in \Gamma_a$ with $\theta_- < \theta^\star(a) < \theta_+$. By (D1) and the definition of Γ, for every $\tilde{a} \in A$, we have

$$V(a, \theta_-) + q(a)u(a, \theta_-) \geq V(\tilde{a}, \theta_-) + q(\tilde{a})u(\tilde{a}, \theta_-).$$

Therefore, for every $\tilde{a} > a$, we have

$$\frac{q(\tilde{a}) - q(a)}{\tilde{a} - a} \geq \frac{1}{-u(\tilde{a}, \theta_-)} \left[\frac{V(\tilde{a}, \theta_-) - V(a, \theta_-)}{\tilde{a} - a} + q(a) \frac{u(\tilde{a}, \theta_-) - u(a, \theta_-)}{\tilde{a} - a} \right].$$

Since V and u have continuous partial derivatives in a, we have

$$\underline{q}'_+(a) := \liminf_{\tilde{a} \downarrow a} \frac{q(\tilde{a}) - q(a)}{\tilde{a} - a} \geq C_-,$$

where

$$C_- = -\frac{1}{u(a,\theta_-)}[v(a,\theta_-) + q(a)u_a(a,\theta_-)].$$

Applying a similar argument for $\theta = \theta_-$ and $\tilde{a} < a$, we get

$$\overline{q}'_-(a) := \limsup_{\tilde{a}\uparrow a} \frac{q(\tilde{a}) - q(a)}{\tilde{a} - a} \leq C_-.$$

Similarly, considering $\theta = \theta_+$ with $\tilde{a} > a$ and $\tilde{a} < a$, we get

$$\overline{q}'_+(a) := \limsup_{\tilde{a}\downarrow a} \frac{q(\tilde{a}) - q(a)}{\tilde{a} - a} \leq C_+ \quad \text{and} \quad \underline{q}'_-(a) := \liminf_{\tilde{a}\uparrow a} \frac{q(\tilde{a}) - q(a)}{\tilde{a} - a} \geq C_+,$$

where

$$C_+ = -\frac{1}{u(a,\theta_+)}[v(a,\theta_+) + q(a)u_a(a,\theta_+)].$$

In sum, we have

$$C_- \leq \underline{q}'_+(a) \leq \overline{q}'_+(a) \leq C_+ \quad \text{and} \quad C_+ \leq \underline{q}'_-(a) \leq \overline{q}'_-(a) \leq C_-.$$

We see that $C_- = C_+$ and all four Dini derivatives of q at a coincide, so q has a derivative $q'(a)$ at a that satisfies $q'(a) = C_- = C_+$.

Since $\theta_-, \theta_+ \in \Gamma_a$ are arbitrary, the lemma follows for $\theta \in \Gamma_a$ with $\theta \neq \theta^\star(a)$. For $\theta \in \Gamma_a$ with $\theta = \theta^\star(a)$, we have $a \in \Psi_1$, and the lemma follows by the definition of $q(a)$. □

Lemma 7. *The sets Γ and $\Psi_1 \cup \Psi_2$ are compact.*

Proof. To show that Γ is compact, we need to show that if $(a_n, \theta_n) \to (a, \theta)$ with $(a_n, \theta_n) \in \Gamma$, then $(a, \theta) \in \Gamma$. By Lemma 3, Γ_a is non-empty iff $a \in \Psi_1 \cup \Psi_2$. Thus, $a_n \in \Psi_1 \cup \Psi_2$. There are three cases to consider, up to taking a suitable subsequence.

(1) $a_n \in \Psi_1$ for all n. Since p, V, and $\theta^\star$ are continuous, the set Ψ_1 is closed. Thus, $a \in \Psi_1$. Since $(a_n, \theta_n) \in \Gamma$, we have $p(\theta_n) = V(a_n, \theta_n) + q(a_n)u(a_n, \theta_n)$. Since v, u_a, and $\theta^\star$ are continuous, q is continuous on Ψ_1. Since p, V, and u are also continuous, passing to the limit we have $p(\theta) = V(a, \theta) + q(a)u(a, \theta)$, so $(a, \theta) \in \Gamma$.

(2) $a_n \in \Psi_2$ for all n, and $a \notin \Psi_1$. Since $a_n \in \Psi_2$, we have $a_n \notin \Psi_1$, and hence, by Lemma 5, $\theta^\star(a_n) \notin \Gamma_{a_n}$. Taking another subsequence if necessary, we can assume that $\theta_n - \theta^\star(a_n)$ has the same sign for all n. Suppose for concreteness that $\theta_n > \theta^\star(a_n)$ (the case $\theta_n < \theta^\star(a_n)$ is analogous).

Since $a_n \in \Psi_2$, there exists $\tilde{\theta}_n \in \Gamma_{a_n}$ with $\tilde{\theta}_n < \theta^\star(a_n)$, by Lemma 5. Taking yet another subsequence, we can assume that

$$\tilde{\theta}_n \to \tilde{\theta} \leq \theta^\star(a) \quad \text{and} \quad \mathring{q}(a_n) \to r \in Q(a).$$

(Such a subsequence must exist because $A \times \Theta$ is compact, $\theta^\star$ is continuous, and Q is upper hemi-continuous.) Moreover, by continuity of p, V, and u, we have

$$p(\tilde{\theta}) = V(a, \tilde{\theta}) + ru(a, \tilde{\theta}) \quad \text{and} \quad p(\theta) = V(a, \theta) + ru(a, \theta).$$

Since $a \notin \Psi_1$, we have $\tilde{\theta}, \theta \neq \theta^\star(a)$. Thus, $\tilde{\theta} < \theta^\star(a) < \theta$. Next, $\theta > \theta^\star(a)$ implies that $r = \max Q(a)$; otherwise, $p(\theta) < V(a, \theta) + \max Q(a)u(a, \theta)$, contradicting the definition of $Q(a)$. Similarly, $\tilde{\theta} < \theta^\star(a)$ implies that $r = \min Q(a)$. Hence,

$$r = \min Q(a) = \max Q(a) = \mathring{q}(a).$$

We have $p(\theta) = V(a, \theta) + \mathring{q}(a)u(a, \theta)$. Since $a \notin \Psi_1$, this says that $(a, \theta) \in \Gamma$.

(3) $a_n \in \Psi_2$ for all n, and $a \in \Psi_1$. If $\theta = \theta^\star(a)$, then $(a, \theta) \in \Gamma$ because p, V, and u are continuous, and q is bounded. So suppose for concreteness that $\theta > \theta^\star(a)$ (the case $\theta < \theta^\star(a)$ is analogous). Taking another subsequence if necessary, we can assume that $\theta_n > \theta^\star(a_n)$ for all n. By Lemma 5, for each n there exist $\tilde{\theta}_n \in \Gamma_{a_n}$ with $\tilde{\theta}_n < \theta^\star(a_n)$. Taking a subsequence again, we can assume that

$$\tilde{\theta}_n \to \tilde{\theta} \leq \theta^\star(a) \quad \text{and} \quad \mathring{q}(a_n) \to r \in Q(a).$$

Passing to the limit, we get

$$p(\tilde{\theta}) = V(a, \tilde{\theta}) + ru(a, \tilde{\theta}) \quad \text{and} \quad p(\theta) = V(a, \theta) + ru(a, \theta).$$

If $\tilde{\theta} < \theta^\star(a)$, then as in the previous case $r = \min Q(a) = \max Q(a)$. Since $q(a) \in Q(a)$ by Lemma 4, this yields $r = q(a)$, and hence $(a, \theta) \in \Gamma$.

Finally, if $\tilde{\theta} = \theta^\star(a)$, then by Lemma 6 and $a_n \in \Psi_2$ we have

$$v(a_n, \theta_n) + \mathring{q}(a_n)u_a(a_n, \theta_n) + \mathring{q}'(a_n)u(a_n, \theta_n) = 0,$$
$$v(a_n, \tilde{\theta}_n) + \mathring{q}(a_n)u_a(a_n, \tilde{\theta}_n) + \mathring{q}'(a_n)u(a_n, \tilde{\theta}_n) = 0.$$

Thus,

$$\mathring{q}(a_n) = \frac{v(a_n, \tilde{\theta}_n)u(a_n, \theta_n) - v(a_n, \theta_n)u(a_n, \tilde{\theta}_n)}{-u_a(a_n, \tilde{\theta}_n)u(a_n, \theta_n) + u_a(a_n, \theta_n)u(a_n, \tilde{\theta}_n)}.$$

As $\mathring{q}(a_n) \to r$ and $a \in \Psi_1$, passing to the limit we have

$$r = \frac{v(a, \theta^\star(a))}{-u_a(a, \theta^\star(a))} = q(a).$$

This shows that $(a, \theta) \in \Gamma$.

We have shown that Γ is compact. By Lemma 3, $\Psi_1 \cup \Psi_2 = A_\Gamma$, and thus is compact as the projection of a compact set. □

Lemma 8. *An implementable outcome π is optimal iff* $\operatorname{supp}(\pi) \subset \Gamma$.

Proof. For any implementable outcome π, we have, by (P1), (D1), and (P2),

$$\begin{aligned}\int_\Theta p(\theta)\mathrm{d}\phi(\theta) &= \int_{A\times\Theta} p(\theta)\mathrm{d}\pi(a,\theta) \\ &\geq \int_{A\times\Theta} (V(a,\theta) + q(a)u(a,\theta))\mathrm{d}\pi(a,\theta) \\ &= \int_{A\times\Theta} V(a,\theta)\mathrm{d}\pi(a,\theta).\end{aligned}$$

By Lemma 1, π is optimal iff the inequality holds with equality, or equivalently $\pi(\Gamma) = 1$. In turn, since Γ is compact, $\pi(\Gamma) = 1$ iff $\operatorname{supp}(\pi) \subset \Gamma$, because $\operatorname{supp}(\pi)$ is defined as the smallest compact set of measure one. □

Lemma 9. *The set $\Gamma^\star$ is Borel, and* (1) *holds for all $(a, \theta) \in \Gamma^\star$.*

Proof. Since Γ is compact, $\min \Gamma_a$ and $\max \Gamma_a$ are measurable functions from A_Γ to Θ that satisfy $\min \Gamma_a, \max \Gamma_a \in \Gamma_a$ for all $a \in A_\Gamma$. Since $\theta^\star(a)$ is a continuous function that satisfies $\min \Gamma_a \leq \theta^\star(a) \leq \max \Gamma_a$ for all $a \in A_\Gamma$, it follows that $\Gamma^\star$ is a Borel subset of Γ. Finally, if $\Gamma^\star_a = \{\theta^\star(a)\}$ then (1) holds at $(a, \theta^\star(a))$ by the definition of $q(a)$; otherwise, $\min \Gamma_a < \theta^\star(a) < \max \Gamma_a$, so (1) holds at (a, θ) for all $\theta \in \Gamma_a = \Gamma^\star_a$, by Lemma 6. □

Lemma 10. *An implementable outcome π satisfies* $\operatorname{supp}(\pi) \subset \Gamma$ *iff there exists a conditional probability π_a such that* $\operatorname{supp}(\pi_a) \subset \Gamma^\star_a$ *and* $\int u(a, \theta)\mathrm{d}\pi_a(\theta) = 0$ *for all* $a \in \operatorname{supp}(\alpha_\pi)$.

Proof. If an outcome π admits such a conditional probability then $\pi(\Gamma^\star) = \pi(\Gamma) = 1$, so $\operatorname{supp}(\pi) \subset \Gamma$. Now fix an implementable outcome π such that $\operatorname{supp}(\pi) \subset \Gamma$. Recall that α_π is the a-marginal distribution. Let π_a be any version of the conditional probability. By (P2) and $\text{supp}(\pi) \subset \Gamma$, there exists a Borel set $S_\pi \subset \text{supp}(\alpha_\pi)$ with $\alpha_\pi(S_\pi) = 1$ such that

$$\text{supp}(\pi_a) \subset \Gamma_a \quad \text{and} \quad \int_\Theta u(a,\theta)\mathrm{d}\pi_a(\theta) = 0 \quad \text{for all } a \in S_\pi.$$

Hence, for each $a \in S_\pi$,

$$\text{either} \quad \min \Gamma_a < \theta^\star(a) < \max \Gamma_a \quad \text{or} \quad \min \Gamma_a = \theta^\star(a) = \max \Gamma_a.$$

By definition, $\Gamma^\star_a$ coincides with Γ_a for such a, so

$$\text{supp}(\pi_a) \subset \Gamma_a = \Gamma^\star_a \quad \text{for all } a \in S_\pi.$$

Finally, for all $a \in A_\Gamma \setminus S_\pi$, we can redefine π_a as follows:

$$\pi_a = \rho_a \delta_{\min \Gamma^\star_a} + (1-\rho_a)\delta_{\max \Gamma^\star_a},$$
$$\text{where} \quad \rho_a = \frac{u(a, \max \Gamma^\star_a)}{u(a, \max \Gamma^\star_a) - u(a, \min \Gamma^\star_a)} \mathbf{1}\{\min \Gamma^\star_a < \max \Gamma^\star_a\}.$$

With this definition, π_a automatically satisfies the conditions of the lemma for all $a \in A_\Gamma \setminus S_\pi$. Lastly, since $\alpha_\pi(S_\pi) = 1$, the redefined π_a coincides with the original π_a for α_π-almost all a, and thus is a valid version of the conditional probability. □

Lemma 11. *There exists a unique $p \in C(\Theta)$ that solves (D).*

Proof. Recall that to define Γ we took an arbitrary solution p to (D) and then selected $q(a) \in Q(a)$ so that the associated contact set is compact. By the definition of Γ, we have

$$p(\theta) = V(a,\theta) + q(a)u(a,\theta), \quad \text{for all } (a,\theta) \in \Gamma.$$

Fix any solution π to (P). By Theorem 7, $\Sigma := \text{supp}(\pi) \subset \Gamma$. Let Σ_a denote the a-section of Σ. Define the set $\Sigma^\star \subset \Sigma$ by letting its a-section be given by

$$\Sigma^\star_a = \begin{cases} \{\theta^\star(a)\}, & \theta^\star(a) \in \{\min \Sigma_a, \max \Sigma_a\}, \\ \Sigma_a, & \text{otherwise}, \end{cases} \quad \text{for all } a \in A.$$

Since $\Sigma \subset \Gamma$, we get $\Sigma^\star \subset \Gamma^\star$. By Lemma 10, $\pi(\Sigma^\star) = 1$. Let the projection of $\Sigma^\star$ on Θ be defined as $\Theta_{\Sigma^\star} = \{\theta \in \Theta : (a,\theta) \in \Sigma^\star \text{ for some } a \in A\}$. Then, $\phi(\Theta_{\Sigma^\star}) = 1$ and the closure of $\Theta_{\Sigma^\star}$ is Θ.

Next take any $\theta \in \Theta_{\Sigma^\star}$. If $(a^\star(\delta_\theta), \theta) \in \Sigma^\star$, then $p(\theta) = V(a^\star(\delta_\theta), \theta)$. Otherwise, by the definition of $\Sigma^\star$, there exist $a \in A$ and $\theta' \in \Theta$ such that $(a,\theta), (a,\theta') \in \Sigma^\star$ and either $\theta < \theta^\star(a) < \theta'$ or $\theta' < \theta^\star(a) < \theta$. Suppose that $\theta < \theta^\star(a) < \theta'$ (the other case is analogous and omitted). By Theorem 7, we have

$$v(a,\theta)+q(a)u_a(a,\theta)+q'(a)u(a,\theta)=0,$$
$$v(a,\theta')+q(a)u_a(a,\theta')+q'(a)u(a,\theta')=0.$$

Adding the first equation multiplied by $u(a,\theta')$ and the second multipliled by $-u(a,\theta)$, we obtain

$$q(a)=-\frac{v(a,\theta)u(a,\theta')-v(a,\theta')u(a,\theta)}{u_a(a,\theta)u(a,\theta')-u_a(a,\theta')u(a,\theta)},$$

which is well-defined because the denominator is strictly negative by Assumption 2. Consequently, $p(\theta)=V(a,\theta)+q(a)u(a,\theta)$. In sum, for each $\theta\in\Theta_{\Sigma^\star}$, an arbitrary solution $p(\theta)$ to (D) is determined by $\Sigma^\star$, which is constructed from a fixed solution π to (P). Moreover, since Θ is the closure of $\Theta_{\Sigma^\star}$, there is a unique continuous extension of p from $\Theta_{\Sigma^\star}$ to Θ. This shows that there is a unique $p\in C(\Theta)$ that solves (D). □

D.2. **Proof of Theorem 1.** We first prove the second part where the twist condition holds. The proof of this part remains valid if Assumption 4 is replaced with strict single crossing of $u(a,\theta)$ in θ. Suppose by contradiction that there exist (a,θ_1), (a,θ_2), and (a,θ_3) in $\Gamma^\star$ with $\theta_1<\theta_2<\theta_3$. Then, by the definition of $\Gamma^\star$, we have $\min\Gamma^\star_a<\theta^\star(a)<\max\Gamma^\star_a$. Thus, by redefining $\theta_1=\min\Gamma^\star_a$ and $\theta_3=\max\Gamma^\star_a$ if necessary, we can assume that $\theta_1<\theta^\star(a)<\theta_3$, so (2) holds. But this implies that the rows of the matrix S are linearly independent, which contradicts the fact that (1) holds at (a,θ_1), (a,θ_2), and (a,θ_3). Thus, $|\Gamma^\star_a|\leq 2$ for all $a\in A$.

We now turn to the first part. The proof of this part does not require Assumption 4, and it remains valid when Θ is an arbitrary compact metric space. For any $\mu\in\Delta(\Theta)$, denote the set of distributions of posteriors with average posterior equal to μ by

$$\Delta_2(\mu)=\left\{\tau\in\Delta(\Delta(\Theta)):\int_{\Delta(\Theta)}\eta\mathrm{d}\tau(\eta)=\mu\right\}.$$

Let $\Delta_2^{Bin}(\mu)\subset\Delta_2(\mu)$ denote the set of such distributions where in addition the posterior is always supported on at most two states:

$$\Delta_2^{Bin}(\mu)=\left\{\tau\in\Delta_2(\mu):\mathrm{supp}(\tau)\subset\Delta_1^{Bin}\right\},$$

where

$$\Delta_1^{Bin}=\{\eta\in\Delta(\Theta):|\mathrm{supp}(\eta)|\leq 2\}.$$

We wish to show that for each $\tau\in\Delta_2(\phi)$, there exists $\hat\tau\in\Delta_2^{Bin}(\phi)$ such that $\pi_{\hat\tau}=\pi_\tau$.

We set the stage by defining some key objects and establishing their properties. Define $\Delta_1 = \Delta(\Theta)$ and $\Delta_2 = \Delta(\Delta(\Theta))$. Since Θ is compact, the sets Δ_1 and Δ_2 are also compact (in the weak* topology), by Prokhorov's Theorem (Theorem 15.11 in Aliprantis and Border 2006). Moreover, $\Delta_2(\mu)$ is compact, since it is a closed subset of the compact set Δ_2.

Define the correspondence $P : \Delta_1 \rightrightarrows \Delta_1$ as

$$P(\mu) = \left\{ \eta \in \Delta_1 : \int u(a^\star(\mu), \theta) \, \mathrm{d}\eta(\theta) = 0 \right\}.$$

For each $\mu \in \Delta_1$, $P(\mu)$ is a *moment set*—a set of probability measures $\eta \in \Delta_1$ satisfying a given moment condition (e.g., Winkler 1988). By Assumption 2, we have, for all $\mu, \eta \in \Delta_1$,

$$\eta \in P(\mu) \iff a^\star(\mu) = a^\star(\eta). \tag{13}$$

Clearly, $P(\mu)$ is nonempty (as $\mu \in P(\mu)$) and convex. Since u is continuous in θ, $P(\mu)$ is a closed subset of Δ_1, and hence is compact. Moreover, the correspondence P has a closed graph. Indeed, consider two sequences $\mu_n \to \mu \in \Delta_1$ and $\eta_n \to \eta \in \Delta_1$ with $\mu_n \in \Delta_1$ and $\eta_n \in P(\mu_n)$, so that

$$\int u(a^\star(\mu_n), \theta) \, \mathrm{d}\eta_n(\theta) = 0.$$

Note that $a^\star(\mu)$ is a continuous function of μ, by Berge's theorem (Theorem 17.31 in Aliprantis and Border 2006). Since u is also continuous, by Corollary 15.7 in Aliprantis and Border (2006) we have

$$\int u(a^*(\mu), \theta) \, \mathrm{d}\eta(\theta) = 0,$$

proving that $\eta \in P(\mu)$, so P has a closed graph.

Define the correspondence $E : \Delta_1 \rightrightarrows \Delta_1$ as

$$E(\mu) = P(\mu) \cap \Delta_1^{Bin} = \{ \eta \in P(\mu) : |\operatorname{supp} \eta| \leq 2 \}.$$

Notice that for each $\mu \in \Delta_1$, the support of μ is well defined, by Theorem 12.14 in Aliprantis and Border (2006). Moreover, from the proof of Theorem 15.8 in Aliprantis and Border (2006), it follows that Δ_1^{Bin} is a closed subset of Δ_1, so both Δ_1^{Bin} and $E(\mu)$ are compact.

Define the correspondence $\Lambda : \Delta_1 \rightrightarrows \Delta_2$ as

$$\Lambda(\mu) = \left\{ \lambda \in \Delta(E(\mu)) : \mu = \int_{E(\mu)} \eta \mathrm{d}\lambda(\eta) \right\}.$$

Lemma 13 shows that the correspondence Λ admits a measurable selection. In turn, Lemma 13 relies on the following lemma, which follows immediately from the Choquet Theorem (Theorem 3.1 in Winkler 1988) and Richter-Rogosinsky's Theorem (Theorem 2.1 in Winkler 1988).

Lemma 12. *Let Assumptions 1 and 2 hold. For any $a \in A$ and $\mu \in \Delta(\Theta)$ such that $\int u(a,\theta)\mathrm{d}\mu = 0$, there exists $\lambda_\mu \in \Delta(\Delta(\Theta))$ such that $\int \eta \mathrm{d}\lambda_\mu = \mu$ and for each $\eta \in \operatorname{supp}(\lambda_\mu)$ we have $\int u(a,\theta)\mathrm{d}\eta = 0$ and $|\operatorname{supp}(\eta)| \leq 2$.*

Lemma 13. *There exists a measurable function $\mu \mapsto \lambda_\mu \in \Lambda(\mu)$.*

Proof. The correspondence Λ is nonempty-valued, by Lemma 12. Next, fix $\mu \in \Delta_1$, and consider a sequence $\lambda_n \to \lambda \in \Delta_2$ with $\lambda_n \in \Lambda(\mu)$. By the Portmanteau Theorem (Theorem 15.3 in Aliprantis and Border 2006), we have

$$\int_{E(\mu)} \eta \mathrm{d}\lambda_n(\eta) \to \int_{E(\mu)} \eta \mathrm{d}\lambda(\eta) \quad \text{and} \quad \limsup_n \lambda_n(E(\mu)) \leq \lambda(E(\mu)),$$

where the last inequality holds because $E(\mu)$ is closed. Thus,

$$\int_{E(\mu)} \eta \mathrm{d}\lambda(\eta) = \mu \quad \text{and} \quad 1 = \limsup_n \lambda_n(E(\mu)) \leq \lambda(E(\mu)) \leq 1,$$

proving that $\lambda \in \Lambda(\mu)$. Thus, Λ is closed-valued.

Next, consider two sequences $\mu_n \to \mu \in \Delta_1$ and $\lambda_n \to \lambda \in \Delta_2$ with $\mu_n \in \Delta_1$ and $\lambda_n \in \Lambda(\mu_n)$, so that

$$\mu_n = \int \eta \mathrm{d}\lambda_n(\eta), \quad \lambda_n\left(\Delta_1^{Bin}\right) = 1, \quad \text{and} \quad \lambda_n(P(\mu_n)) = 1.$$

The Portmanteau Theorem implies that $\mu = \int \eta \mathrm{d}\lambda(\eta)$ and $\lambda\left(\Delta_1^{Bin}\right) = 1$, since Δ_1^{Bin} is closed. Define $\overline{P}(\mu_n)$ as the closure of $\cup_{k=n}^{\infty} P(\mu_k)$. By construction, $P(\mu_k) \subset \overline{P}(\mu_k) \subset \overline{P}(\mu_n)$ for $k \geq n$, so the Portmanteau Theorem implies that $\lambda(\overline{P}(\mu_n)) = 1$. Moreover, $\overline{P}(\mu_n) \downarrow \overline{P} \subset P(\mu)$, because P has a closed graph. Hence, $\lambda(P(\mu)) = 1$, by the continuity of probability measures (Theorem 10.8 in Aliprantis and Border 2006). That is, $\lambda \in \Lambda(\mu)$, showing that the correspondence Λ has a closed graph.

Therefore, Λ is measurable, by Theorem 18.20 in Aliprantis and Border (2006), as well as nonempty- and closed-valued. Hence, there exists a measurable function $\mu \mapsto \lambda_\mu \in \Lambda(\mu)$, by Theorem 18.13 in Aliprantis and Border (2006). □

Finally, taking a measurable selection, for each $\tau \in \Delta_2(\phi)$, define $\hat{\tau} \in \Delta_2$ as

$$\hat{\tau}\left(\widetilde{\Delta}_1\right) = \int_{\Delta_1} \lambda_\mu \left(\widetilde{\Delta}_1\right) \mathrm{d}\tau(\mu) \tag{14}$$

for every measurable set $\widetilde{\Delta}_1 \subset \Delta_1$. By construction, $\hat{\tau} \in \Delta_2^{Bin}(\phi)$, since

$$\hat{\tau}(\Delta_1^{Bin}) = \int_{\Delta_1} \lambda_\mu(\Delta_1^{Bin})\mathrm{d}\tau(\mu) = 1$$

and

$$\phi = \int_{\Delta_1} \mu \mathrm{d}\tau(\mu) = \int_{\Delta_1} \left(\int_{E(\mu)} \eta \mathrm{d}\lambda_\mu(\eta) \right) \mathrm{d}\tau(\mu) = \int_{\Delta_1} \eta \mathrm{d}\hat{\tau}(\eta),$$

where the first equality holds by $\tau \in \Delta_2(\phi)$, the second by $\lambda_\mu \in \Lambda$, and the third by (14). Similarly, for each measurable $\widetilde{A} \subset A$ and $\widetilde{\Theta} \subset \Theta$, we have

$$\begin{aligned}
\pi_\tau(\widetilde{A}, \widetilde{\Theta}) &= \int_{\Delta_1} 1\{a^*(\mu) \in \widetilde{A}\}\mu(\widetilde{\Theta})\mathrm{d}\tau(\mu) \\
&= \int_{\Delta_1} 1\{a^*(\mu) \in \widetilde{A}\} \left(\int_{E(\mu)} \eta(\widetilde{\Theta}) d\lambda_\mu(\eta) \right) \mathrm{d}\tau(\mu) \\
&= \int_{\Delta_1} \left(\int_{E(\mu)} 1\{a^*(\eta) \in \widetilde{A}\}\eta(\widetilde{\Theta})\mathrm{d}\lambda_\mu(\eta) \right) \mathrm{d}\tau(\mu) \\
&= \int_{\Delta_1} 1\{a^*(\eta) \in \widetilde{A}\}\eta(\widetilde{\Theta})\mathrm{d}\hat{\tau}(\eta) \\
&= \pi_{\hat{\tau}}(\widetilde{A}, \widetilde{\Theta}),
\end{aligned}$$

where the second equality holds by $\lambda_\mu \in \Lambda$, the third by (13) and $E(\mu) \subset P(\mu)$, and the fourth by (14).

D.3. **Proof of Corollary 1.** The proof of Corollary 1 remains valid if Assumption 4 is replaced with strict single-crossing of $u(a, \theta)$ in θ. Let a be such that $\int u(a, \theta)\mathrm{d}\phi = 0$. Since $|\Theta| \geq 3$, Assumption 4 and $\int u(a, \theta)\mathrm{d}\phi = 0$ imply that there exist $\theta_1 < \theta_2 < \theta_3$ in Θ such that $\theta_1 < \theta^\star(a) < \theta_3$.

Suppose that no disclosure is optimal. Then, by part (2) of Theorem 7, it follows that $\Gamma_a^\star = \Gamma_a = \Theta$ and (1) holds for all $\theta \in \Theta$, so there exist constants $q(a), q'(a) \in \mathbb{R}$

such that

$$v(a,\theta) = -q(a)u_a(a,\theta) - q'(a)u(a,\theta) \quad \text{for all } \theta \in \Theta.$$

That is, $v(a,\cdot)$ lies in a linear space L spanned by $u_a(a,\cdot)$ and $u(a,\cdot)$, whose dimension is at most 2. But the space of functions $v(a,\cdot)$ satisfying Assumption 1 is the linear space $C(\Theta)$, whose dimension is at least 3, since $|\Theta| \geq 3$. Hence, the space L is a proper subspace of $C(\Theta)$, so generically $v(a,\cdot)$ does not belong to L, and thus generically no disclosure is suboptimal.

D.4. **Proof of Theorem 2.** The proof of Theorem 2 remains valid if the condition $u_\theta(a,\theta) > 0$ in Assumption 4 is replaced with strict single-crossing of $u(a,\theta)$ in θ. We will prove that Γ is single-dipped, which implies that every optimal outcome is single-dipped. We start with an appropriate version of the theorem of alternative.

Lemma 14. *Exactly one of the following two alternatives holds.*

(1) *There exists $x > 0$ such that $xR \leq 0$.*

(2) *There exists $y \geq 0$ such that $Ry \geq 0$ and $Ry \neq 0$.*

Proof. Clearly, (1) and (2) cannot both hold, because premultiplying $Ry \geq 0$ with $Ry \neq 0$ by $x > 0$ yields $xRy > 0$, whereas postmultiplying $xR \leq 0$ by $y \geq 0$ yields $xRy \leq 0$.

Now suppose that (1) does not hold. Then there does not exist $x \geq 0$ such that

$$x \begin{pmatrix} R & -I \end{pmatrix} \leq \begin{pmatrix} 0 & -e \end{pmatrix}$$

where I is an identity matrix and e is a row vector of ones. Thus, by the theorem of alternative (e.g., Theorem 2.10 in Gale 1989), there exists $y \geq 0$ and $z \geq 0$ such that

$$\begin{pmatrix} R \\ -I \end{pmatrix} \begin{pmatrix} y & z \end{pmatrix} \geq 0 \quad \text{and} \quad -ez < 0,$$

which in turn shows that (2) holds. □

We prove the theorem by contraposition. Suppose that Γ is not single-dipped, so it contains a strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3). Without loss, we can assume that $\theta_1 \leq \theta^\star(a_1) \leq \theta_3$. This is because $\min\Gamma_{a_1} \leq \theta^\star(a_1) \leq \max\Gamma_{a_1}$ by Theorem 7, and thus the triple $(a_1, \min\Gamma_{a_1})$, (a_2,θ_2), $(a_1, \max\Gamma_{a_1})$ is strictly single-peaked and lies in Γ.

By (D1) and Theorem 7, we have

$$\begin{aligned}
V(a_1,\theta_1)+q(a_1)u(a_1,\theta_1) &\geq V(a_2,\theta_1)+q(a_2)u(a_2,\theta_1),\\
V(a_2,\theta_2)+q(a_2)u(a_2,\theta_2) &\geq V(a_1,\theta_2)+q(a_1)u(a_1,\theta_2),\\
V(a_1,\theta_3)+q(a_1)u(a_1,\theta_3) &\geq V(a_2,\theta_3)+q(a_2)u(a_2,\theta_3).
\end{aligned}$$

By (12), for an optimal π_a, and for $i \in \{1,2\}$, we have

$$q(a_i)=\frac{\mathbb{E}_{\pi_{a_i}}[v(a_i,\theta)]}{-\mathbb{E}_{\pi_{a_i}}[u_a(a_i,\theta)]}>0,$$

where the inequality follows from Assumptions 2 and 4. Thus, the vector $x = (1, q(a_1), q(a_2))$ is strictly positive and satisfies $xR \leq 0$. By Lemma 14, there does not exist a vector $y \geq 0$ such that $Ry \geq 0$ and $Ry \neq 0$, as desired.

D.5. **Proof of Theorem 3.** The proof uses the following five lemmas, whose proofs are deferred to Appendix F. For the second part of the theorem, we will prove that Γ is single-dipped (-peaked) and $\Gamma^\star$ is strictly single-dipped (-peaked), which implies that every outcome is strictly single-dipped (-peaked).

Lemma 15. *If $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ and $v_\theta(a,\theta)/u_\theta(a,\theta)$ are increasing (decreasing) in θ for all a, with at least one of them strictly increasing (decreasing), then $|S| > (<)0$ for all a and $\theta_1 < \theta_2 < \theta_3$ such that $\theta_1 < \theta^\star(a) < \theta_3$.*

Lemma 16. *If $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ and $v_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ are increasing (decreasing) in θ for all a and $a_2 \geq (\leq) a_1$, with at least one of them strictly increasing (decreasing), then $|R| > (<)0$ for all $\theta_1 < \theta_2 < \theta_3$ and all $a_2 > (<) a_1$ such that $\theta_1 \leq \theta^\star(a_1) \leq \theta_3$.*

Lemma 17. *If $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ is increasing in θ for all a, then for all $\theta_1 < \theta_2 < \theta_3$ and all $a_2 > a_1$ such that $\theta_1 \leq \theta^\star(a_1) \leq \theta_3$, we have*

$$\begin{aligned}
u(a_2,\theta_3)u(a_1,\theta_1) &> u(a_2,\theta_1)u(a_1,\theta_3),\\
u(a_2,\theta_2)u(a_1,\theta_1) &> u(a_2,\theta_1)u(a_1,\theta_2),\\
u(a_2,\theta_3)u(a_1,\theta_2) &> u(a_2,\theta_2)u(a_1,\theta_3).
\end{aligned}$$

Lemma 18. *If $v_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ is decreasing in θ for all $a_2 \leq a_1$, then for all $\theta_1 < \theta_2 < \theta_3$ and all $a_2 < a_1$ such that $\theta_1 \leq \theta^\star(a_1) \leq \theta_3$, we have*

$$\frac{u(a_1,\theta_1)}{V(a_1,\theta_1)-V(a_2,\theta_1)}<\frac{u(a_1,\theta_2)}{V(a_1,\theta_2)-V(a_2,\theta_2)}<\frac{u(a_1,\theta_3)}{V(a_1,\theta_3)-V(a_2,\theta_3)}.$$

Lemma 19. *Suppose that v^n is a sequence of continuous functions converging uniformly to v, and suppose that the corresponding contact sets Γ^n are single-dipped (-peaked). Then there exists a single-dipped (-peaked) optimal outcome.*

Now, the set Γ is single-dipped (-peaked) by Theorem 2 with

$$y = \begin{pmatrix} u(a_2,\theta_3)u(a_1,\theta_2) - u(a_2,\theta_2)u(a_1,\theta_3) \\ u(a_2,\theta_3)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_3) \\ u(a_2,\theta_2)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_2) \end{pmatrix} \quad \left(y = \begin{pmatrix} \frac{u(a_2,\theta_1)}{V(a_2,\theta_1)-V(a_1,\theta_1)} \\ \frac{u(a_2,\theta_2)}{V(a_2,\theta_2)-V(a_1,\theta_2)} \\ \frac{u(a_2,\theta_3)}{V(a_2,\theta_3)-V(a_1,\theta_3)} \end{pmatrix} \right),$$

as follows from Lemma 16 and Lemma 17 (Lemma 18). Moreover, $|\Gamma_a^\star| \leq 2$ for all a by Theorem 1 and Lemma 15, showing that $\Gamma^\star$ is strictly single-dipped (-peaked). Finally, consider

$$v^n(a,\theta) = v(a,\theta) + \int_0^\theta \frac{\tilde{v}(\theta)}{n} u_\theta(a,\tilde{\theta})\mathrm{d}\tilde{\theta},$$

where $\tilde{v}(\theta)$ is a continuous, strictly positive, and strictly increasing (decreasing) function on $\overline{\Theta}$. Then $v^n(a,\theta) > 0$ because $v(a,\theta) > 0$ and $u_\theta(a,\theta) > 0$ for all (a,θ), by Assumption 4. Moreover, for all $a_2 \geq (\leq) a_1$,

$$\frac{v_\theta^n(a_2,\theta)}{u_\theta(a_1,\theta)} = \frac{v_\theta(a_2,\theta)}{u_\theta(a_1,\theta)} + \frac{\tilde{v}(\theta)}{n}\frac{u_\theta(a_2,\theta)}{u_\theta(a_1,\theta)}$$

is strictly increasing (decreasing) in θ, because $\tilde{v}(\theta)$ is strictly positive and strictly increasing (decreasing) in θ; $v_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ is increasing (decreasing) in θ; and $u_\theta(a_2,\theta)/u_\theta(a_1,\theta)$ is increasing in θ, since $u_{a\theta}(a,\theta)/u_\theta(a,\theta)$ is increasing (decreasing) in θ. Thus, by Lemma 19, there exists an optimal single-dipped (-peaked) outcome.

D.6. **Proof of Theorem 8.** We first show that the contact set can be described as in Remark 1. Since $\Gamma^\star$ is strictly single-dipped, we have $|\Gamma_a^\star| \leq 2$ for all $a \in A_\Gamma$, so $\Gamma_a^\star = \{t_1(a), t_2(a)\}$ with $t_1(a) = \min \Gamma_a^\star \leq \theta^\star(a) \leq \max \Gamma_a^\star = t_2(a)$ for all $a \in A_\Gamma$. Since Γ is compact, and $\Gamma^\star$ is constructed from Γ using a continuous function $\theta^\star(a)$, the functions t_1 and t_2 are measurable. Since $\Gamma^\star$ is single-dipped, for all $a < a'$ in A_Γ, we have $t_2(a) \leq t_2(a')$, as otherwise $(a, t_1(a))$, $(a', t_2(a'))$, $(a, t_2(a))$ would be a strictly single-peaked triple in $\Gamma^\star$; and $t_1(a') \notin (t_1(a), t_2(a))$, as otherwise $(a, t_1(a))$, $(a', t_1(a'))$, $(a, t_2(a))$ would be a strictly single-peaked triple in $\Gamma^\star$.

Suppose now that the set $\{a \in A_\Gamma : t_1(a) < t_2(a)\}$ is the union of finitely many intervals. We claim that for each $a \in A_\Gamma$ there exists $\varepsilon > 0$ such that, for all $\tilde{a}_1, \tilde{a}_2 \in [a-\varepsilon, a] \cap A_\Gamma$ with $t_1(\tilde{a}_1) \neq t_2(\tilde{a}_1)$ and $t_1(\tilde{a}_2) \neq t_2(\tilde{a}_2)$, we have $t_1(\tilde{a}_1) < t_2(\tilde{a}_2)$. This claim is obvious if there does not exist a sequence $a_n \in A_\Gamma$ such that $a_n \uparrow a$, so suppose that such a sequence a_n exists. By monotonicity of t_2, the sequence $t_2(a_n)$ converges to some $t_2(a_-) \leq t_2(a)$. In fact, we must have $t_2(a_-) = t_2(a)$, meaning that t_2 is left-continuous at a. First, if $t_1(a) < \theta^\star(a) < t_2(a)$, then $t_2(a_-) = t_2(a)$, as otherwise $\Gamma_a^\star$ would contain at least three distinct states $t_1(a)$, $t_2(a_-)$, and $t_2(a)$, by compactness of Γ, contradicting that $|\Gamma_a^\star| \leq 2$. Second, if $t_1(a) = \theta^\star(a) = t_2(a)$, then $t_2(a_-) = t_2(a)$, as otherwise there would exist $a_n \in A_\Gamma$ such that $t_2(a_n) < \theta^\star(a_n)$, contradicting that $t_1(a_n) \leq \theta^\star(a_n) \leq t_2(a_n)$. We will now show that there exists $\varepsilon > 0$ with the required property. If $t_1(a) < \theta^\star(a) < t_2(a)$, then, by left-continuity of t_2, there exists $\varepsilon > 0$ such that, for all $\tilde{a}_2 \in [a-\varepsilon, a] \cap A_\Gamma$, we have $\theta^\star(a) < t_2(\tilde{a}_2)$, and thus, for all $\tilde{a}_1 \in [a-\varepsilon, a] \cap A_\Gamma$, we have $t_1(\tilde{a}_1) \leq \theta^\star(\tilde{a}_1) \leq \theta^\star(a) < t_2(\tilde{a}_2)$. If $t_1(a) = \theta^\star(a) = t_2(a)$, then, by left-continuity of t_2 and the regularity condition, there exists $\varepsilon > 0$ such that t_2 is continuous on $[a-\varepsilon, a] \cap A_\Gamma$ and either (i) $t_1(\tilde{a}) = \theta^\star(\tilde{a}) = t_2(\tilde{a})$ for all $\tilde{a} \in [a-\varepsilon, a] \cap A_\Gamma$, or (ii) $t_1(\tilde{a}) < \theta^\star(\tilde{a}) < t_2(\tilde{a})$ for all $\tilde{a} \in [a-\varepsilon, a) \subset A_\Gamma$, in which case $t_1(\tilde{a}_1) \leq t_1(a-\varepsilon) < t_2(a-\varepsilon) \leq t_2(\tilde{a}_2)$ for all $\tilde{a}_1, \tilde{a}_2 \in [a-\varepsilon, a)$. In particular, the inequality $t_1(\tilde{a}_1) \leq t_1(a-\varepsilon)$ holds because $t_1(\tilde{a}_1) \notin (t_1(a-\varepsilon), t_2(a-\varepsilon))$, as shown in the first paragraph, and $t_1(\tilde{a}_1) \notin [t_2(a-\varepsilon), \theta^\star(a))$, as otherwise $(a^\star(\delta_{t_1(\tilde{a}_1)}), t_1(a^\star(\delta_{t_1(\tilde{a}_1)})),$ $(\tilde{a}_1, t_1(\tilde{a}_1))$, $(a^\star(\delta_{t_1(\tilde{a}_1)}), t_2(a^\star(\delta_{t_1(\tilde{a}_1)}))$ would be a strictly single-peaked triple in $\Gamma^\star$. Thus, in both cases (i) and (ii), there exists $\varepsilon > 0$ with the required property.

Suppose now that ϕ has a density. Suppose for contradiction that there exist two distinct optimal outcomes π and π'. Recall that, because $|\Gamma_a^\star| \leq 2$ for all a, we have $\pi_a = \pi'_a = \rho_a \delta_{t_1(a)} + (1-\rho_a)\delta_{t_2(a)}$ for all $a \in A_\Gamma$ where $1 - \rho_a > 0$ is given by

$$1 - \rho_a = \begin{cases} \frac{-u(a,t_1(a))}{u(a,t_2(a)) - u(a,t_1(a))}, & t_1(a) < t_2(a), \\ 1, & t_1(a) = t_2(a). \end{cases}$$

Thus, $\alpha_\pi \neq \alpha_{\pi'}$. Define $\hat{a} = \sup\{a \in A : \alpha_\pi([0,a]) \neq \alpha_{\pi'}([0,a])\} \in A_\Gamma$, where the inclusion follows from $\alpha_\pi \neq \alpha_{\pi'}$ and $\alpha_\pi(A_\Gamma) = \alpha_{\pi'}(A_\Gamma) = 1$. As shown above, there exists $\varepsilon > 0$ such that, for all $\tilde{a}_1, \tilde{a}_2 \in [\hat{a}-\varepsilon, \hat{a}] \cap A_\Gamma$ with $t_1(\tilde{a}_1) \neq t_2(\tilde{a}_1)$ and $t_1(\tilde{a}_2) \neq t_2(\tilde{a}_2)$, we have $t_1(\tilde{a}_1) < t_2(\tilde{a}_2)$. We will now show that $\alpha_\pi([0,\tilde{a}]) = \alpha_{\pi'}([0,\tilde{a}])$ for all $\tilde{a} \in [\hat{a}-\varepsilon, \hat{a}]$ contradicting the definition of $\hat{a}$.

By (P1), the marginals of π and π' on Θ are both equal to ϕ. Since t_2 is increasing in a, states $\theta > t_2(\tilde{a})$ can only induce actions $a > \tilde{a}$. Thus, since $\alpha_{\pi'}([0,a]) = \alpha_\pi([0,a])$ for all $a \geq \hat{a}$, and since $t_1(\tilde{a}_1) < t_2(\tilde{a}_2)$ for all $\tilde{a}_1, \tilde{a}_2 \in [\hat{a}-\varepsilon, \hat{a}] \cap A_\Gamma$ with $t_1(\tilde{a}_1) \neq t_2(\tilde{a}_1)$ and $t_1(\tilde{a}_2) \neq t_2(\tilde{a}_2)$, it follows that, for all $\tilde{a} \in [\hat{a}-\varepsilon, \hat{a}] \cap A_\Gamma$, we have

$$\phi((t_2(\tilde{a}), 1]) - \phi([t_2(\tilde{a}), 1]) \leq \int_{[\tilde{a},\hat{a}]} (1-\rho_a) \mathrm{d}\alpha_{\pi'}(a) - \int_{[\tilde{a},\hat{a}]} (1-\rho_a) \mathrm{d}\alpha_\pi(a)$$

$$\leq \phi([t_2(\tilde{a}), 1]) - \phi((t_2(\tilde{a}), 1]).$$

Moreover, since ϕ has a density, we have $\phi((t_2(\tilde{a}), 1]) = \phi([t_2(\tilde{a}), 1])$, and hence

$$\int_{[\tilde{a},\hat{a}]} (1-\rho_a) \mathrm{d}\alpha_{\pi'}(a) = \int_{[\tilde{a},\hat{a}]} (1-\rho_a) \mathrm{d}\alpha_{\pi}(a).$$

Then, since $1-\rho_a > 0$ for all $a \in A_\Gamma$, and since $\mathrm{supp}(\alpha_{\pi'}) \subset A_\Gamma$ and $\mathrm{supp}(\alpha_\pi) \subset A_\Gamma$, it follows that $\alpha_\pi([\tilde{a}, \hat{a}]) = \alpha_\pi([\tilde{a}, \hat{a}])$ for all $\tilde{a} \in [\hat{a}-\varepsilon, \hat{a}]$. Thus, since $\alpha_{\pi'}([0, a]) = \alpha_\pi([0, a])$ for all $a \geq \hat{a}$, it follows that $\alpha_\pi([0, \tilde{a}]) = \alpha_{\pi'}([0, \tilde{a}])$ for all $\tilde{a} \in [\hat{a}-\varepsilon, \hat{a}]$.

D.7. **Proof of Theorem 4.** The support of the full disclosure outcome is $\cup_{\theta\in\Theta}(a^\star(\delta_\theta), \theta)$. Thus, by Lemma 1 and Theorem 7, full disclosure is optimal iff there exists $q \in B(A)$ such that

$$V(a^\star(\delta_\theta), \theta) \geq V(a, \theta) + q(a)u(a, \theta), \quad \text{for all } (a, \theta) \in A \times \Theta,$$
$$\iff \frac{V(a, \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1)}{-u(a, \theta_1)} \leq q(a) \leq \frac{V(a^\star(\delta_{\theta_2}), \theta_2) - V(a, \theta_2)}{u(a, \theta_2)},$$

for all $a \in A$ and $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta^\star(a) < \theta_2$. As shown in the proof of Lemma 24, the left-hand side and right-hand side functions are bounded on $A \times \Theta$, so full disclosure is optimal iff, for all $a \in A$ and $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta^\star(a) < \theta_2$, we have

$$\frac{V(a, \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1)}{-u(a, \theta_1)} \leq \frac{V(a^\star(\delta_{\theta_2}), \theta_2) - V(a, \theta_2)}{u(a, \theta_2)},$$
$$\iff u(a, \theta_2)V(a, \theta_1) - u(a, \theta_1)V(a, \theta_2) \leq u(a, \theta_2)V(a^\star(\delta_{\theta_1}), \theta_1) - u(a, \theta_1)V(a^\star(\delta_{\theta_2}), \theta_2),$$
$$\iff \rho V(a^\star(\mu), \theta_1) + (1-\rho)V(a^\star(\mu), \theta_2) \leq \rho V(a^\star(\delta_{\theta_1})), \theta_1) + (1-\rho)V(a^\star(\delta_{\theta_2}), \theta_2),$$

where $\rho = u(a, \theta_2)/(u(a, \theta_2) - u(a, \theta_1))$, $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$, and $a^\star(\mu) = a$, by the definition of $a^\star(\mu)$. To complete the proof that full disclosure is optimal iff (3) holds for all μ, note that for all a and $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta^\star(a) < \theta_2$, we have $\rho = u(a, \theta_2)/(u(a, \theta_2) - u(a, \theta_1)) \in (0, 1)$; and conversely, for each $\theta_1 < \theta_2$ and $\rho \in (0, 1)$, there exists a unique $a \in (a^\star(\delta_{\theta_1}), a^\star(\delta_{\theta_2}))$ such that $\rho = u(a, \theta_2)/(u(a, \theta_2) - u(a, \theta_1))$.

Finally, assume that (3) holds with strict inequality for all μ. Suppose by contradiction that full disclosure is not uniquely optimal. Then, by Theorem 7, there exist $a \in A_\Gamma$ and $\theta_1, \theta_2 \in \Gamma_a^\star$ such that $\theta_1 < \theta^\star(a) < \theta_2$, so

$$V(a, \theta_1) + q(a)u(a, \theta_1) \geq V(a^\star(\delta_{\theta_1}), \theta_1),$$
$$V(a, \theta_2) + q(a)u(a, \theta_2) \geq V(a^\star(\delta_{\theta_2}), \theta_1).$$

Denote $\rho = u(a,\theta_2)/(u(a,\theta_2) - u(a,\theta_1) \in (0,1)$ and $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$. Notice that $a = a^\star(\mu)$. Adding the first inequality multiplied by ρ and the second inequality multiplied by $1-\rho$ gives

$$\rho V(a^\star(\mu),\theta_1) + (1-\rho)V(a^\star(\mu),\theta_2) \geq \rho V(a^\star(\delta_{\theta_1}),\theta_1) + (1-\rho)V(a^\star(\delta_{\theta_2}),\theta_2),$$

contradicting that (3) holds with strict inequality.

D.8. **Proof of Corollary 2'.** Condition (3) holds because

$$\begin{aligned}
&\rho V(p\theta_1 + (1-\rho)\theta_2,\theta_1) + (1-\rho)V(\rho\theta_1 + (1-\rho)\theta_2,\theta_2)\\
&\leq \rho(\rho V(\theta_1,\theta_1) + (1-\rho)V(\theta_2,\theta_1)) + (1-\rho)(\rho V(\theta_1,\theta_2) + (1-\rho)V(\theta_2,\theta_2))\\
&\leq \rho V(\theta_1,\theta_1) + (1-\rho)V(\theta_2,\theta_2),
\end{aligned}$$

where the first inequality holds because $V(a,\theta)$ is convex in a, and the second holds because $V(\theta_1,\theta_2) + V(\theta_2,\theta_1) \leq V(\theta_1,\theta_1) + V(\theta_2,\theta_2)$.

D.9. **Proof of Theorem 5.** By $\Theta = [0,1]$ and Assumptions 1–4, $\theta^\star(a)$ is a strictly increasing, continuous function from A onto $\overline{\Theta} = \Theta$. Since the range of $\theta^\star$ is Θ and full disclosure is optimal, Theorem 7 implies that $\theta^\star(a) \in \Gamma_a^\star$ for all a. Thus, since the contact set is pairwise (i.e., $|\Gamma_a^\star| \leq 2$) and $\min\Gamma_a^\star < \theta^\star(a) < \max\Gamma_a^\star$ whenever $\Gamma_a^\star$ is multivalued (by the definition of $\Gamma^\star$), it follows that $\Gamma_a^\star = \{\theta^\star(a)\}$ for all a, as otherwise $\min\Gamma_a^\star$, $\theta^\star(a)$, and $\max\Gamma_a^\star$ would be three distinct elements in $\Gamma_a^\star$. Hence, $\Gamma^\star = \cup_{\theta\in\Theta}(a^\star(\delta_\theta),\theta)$, so full disclosure is optimal.

D.10. **Proof of Theorem 6.** We give the proof for the single-dipped case. Since for all $\theta_1 < \theta_2$ there exists $p \in (0,1)$ such that (5) holds, it follows that there do not exist $\theta_1 < \theta_2$ such that $(a^\star(\delta_{\theta_1}),\theta_1)$ and $(a^\star(\delta_{\theta_2}),\theta_2)$ are in Γ. Suppose by contradiction that such θ_1 and θ_2 exist. For any $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$ with $\rho \in (0,1)$, we have

$$\begin{aligned}
p(\theta_1) = V(a^\star(\theta_1),\theta_1) \geq V(a^\star(\mu),\theta_1) + q(a^\star(\mu))u(a^\star(\mu),\theta_1),\\
p(\theta_2) = V(a^\star(\theta_2),\theta_2) \geq V(a^\star(\mu),\theta_2) + q(a^\star(\mu))u(a^\star(\mu),\theta_2),
\end{aligned}$$

by (D1) and the definition of Γ. Adding the first inequality multiplied by ρ and the second inequality multiplied by $1-\rho$, we obtain that (5) fails for all $\rho \in (0,1)$, yielding a contradiction.

Since $\Theta = [0,1]$, Γ^* is strictly single-dipped, and for all $\theta_1 < \theta_2$ there exists $p \in (0,1)$ such that (5) holds, it follows that $t_1(a_2) \leq t_1(a_1)$ for all $a_1 < a_2$ in A_Γ, and thus $\Gamma^\star$ is single-dipped negative assortative. As shown in the first paragraph of the proof of Theorem 8, $\Gamma^\star(a) = \{t_1(a), t_2(a)\}$ for all $a \in A_\Gamma$ where $t_2(a)$ is increasing in a. Suppose by contradiction that there exist $a_1 < a_2$ in A_Γ such that $t_1(a_2) > t_1(a_1)$. Then $t_1(a_2) \geq t_2(a_1)$, as otherwise $(a_1, t_1(a_1))$, $(a_2, t_1(a_2))$, $(a_1, t_2(a_1))$ is a strictly single-peaked triple in $\Gamma^\star$. Define

$$\underline{a}_i = \inf\{a \in A_\Gamma : t_1(a_i) \leq t_1(a) \leq t_2(a) \leq t_2(a_i)\} \leq a_i, \quad \text{for } i = 1, 2.$$

Since $A_{\Gamma^\star} = A_\Gamma$ and A_Γ is compact, we have $\underline{a}_1, \underline{a}_2 \in A_{\Gamma^\star}$. We claim that $\Gamma^\star_{\underline{a}_i} = \{\theta^\star(\underline{a}_i)\}$ for $i = 1, 2$. Suppose by contradiction that $\Gamma^\star_{\underline{a}_i} \neq \{\theta^\star(\underline{a}_i)\}$, so $\Gamma^\star_{\underline{a}_i} = \{t_1(\underline{a}_i), t_2(\underline{a}_i)\}$ with $t_1(\underline{a}_i) < \theta^\star(\underline{a}_i) < t_2(\underline{a}_i)$. Let $\Theta_{\Gamma^\star}$ be the projection of $\Gamma^\star$ on Θ. Since $\pi(\Gamma^\star) = 1$ for an optimal π, we have $\phi(\Theta_{\Gamma^\star}) = 1$ by (P1), and the closure of $\Theta_{\Gamma^\star}$ is $\Theta = [0, 1]$. Thus, there exists $(a, \theta) \in \Gamma^\star$ with $t_1(\underline{a}_i) < \theta < t_2(\underline{a}_i)$. Since $\Gamma^\star$ is strictly single-dipped, it follows that $a < \underline{a}$ (otherwise $(\underline{a}, t_1(\underline{a}))$, (a, θ), $(\underline{a}, t_2(\underline{a}))$ is a single-peaked triple in $\Gamma^\star$) and $t_1(\underline{a}_i) \leq t_1(a) \leq t_2(a) \leq t_2(\underline{a}_i)$ (otherwise either $(a, t_1(a))$, $(\underline{a}_i, t_1(\underline{a}_i)$, (a, θ) or (a, θ), $(\underline{a}_i, t_2(\underline{a}_i)$, $(a, t_2(a))$ is a strictly single-peaked triple in $\Gamma^\star$), contradicting the definition of $\underline{a}$. Hence, $(\underline{a}_1, \theta^\star(\underline{a}_1))$ and $(\underline{a}_2, \theta^\star(\underline{a}_2))$ are in Γ, so by the second step of the proof we must have $\underline{a}_1 = \underline{a}_2$. But, by construction, $t_1(a_1) \leq \theta^\star(\underline{a}_1) \leq t_2(a_1) \leq t_1(a_2) \leq \theta^\star(\underline{a}_2) \leq t_2(a_2)$, and $\underline{a}_1 = \underline{a}_2$ implies that these inequalities all hold with equality, contradicting $t_1(a_2) > t_1(a_1)$.

Now suppose that ϕ has a density f and $\Gamma^\star$ is single-dipped negative assortative. Finally, we show that the functions t_1 and t_2 are continuous and satisfy the differential equations (6)–(7) and the boundary condition (8). Since the closure of the projection $\Theta_{\Gamma^\star}$ of $\Gamma^\star$ on Θ is Θ, it follows that the the closure of the image of the functions t_1 and t_2 must also be equal to Θ. Since t_1 is decreasing and t_2 is increasing on the compact domain A_Γ, and since $t_1(a) \leq \theta^\star(a) \leq t_2(a)$ for all $a \in A_\Gamma$, it follows that t_1 and t_2 are continuous functions such that $t_1(\underline{a}) = \theta^\star(a) = t_2(\underline{a})$, $t_1(a) < \theta^\star(a) < t_2(a)$ for all $a > \underline{a}$, $t_1(\overline{a}) = 0$, $t_2(\overline{a}) = 1$, and $(t_1(\underline{b}_i), t_2(\underline{b}_i)) = (t_1(\overline{b}_i), t_2(\overline{b}_i))$ for all i, where $\{(\underline{b}_i, \overline{b}_i)\}_i$ is an at most countable set of disjoint open intervals comprising the set $[\underline{a}, \overline{a}] \setminus A_\Gamma$. Since ϕ has a density, the measure of the endpoints of these intervals is zero, and hence the set of optimal outcomes is unaffected if we redefine A_Γ as $[\underline{a}, \overline{a}]$ and extend the domain of t_1 and t_2 to $[\underline{a}, \overline{a}]$ by setting $t_1(a) = t_1(\underline{b}_i) = t_1(\overline{b}_i)$ and $t_2(a) = t_2(\underline{b}_i) = t_2(\overline{b}_i)$ for all $a \in (\underline{b}_i, \overline{b}_i)$. In sum, without loss of generality, we can assume that t_1 and t_2 are continuous monotone functions defined on $[\underline{a}, \overline{a}]$ that satisfy (8) and $t_1(a) < \theta^\star(a) < t_2(a)$ for all $a \in (\underline{a}, \overline{a}]$.

Since ϕ has a density and $\Gamma^\star_a = \{t_1(a), t_2(a)\}$ for all $a \in [\underline{a}, \overline{a}]$, where t_1 is continuously decreasing and t_2 is continuously increasing, we can rewrite (P2) for $\tilde{A} = [a, a']$, with

$\underline{a} \leq a < a' \leq \overline{a}$, as

$$\int_a^{a'} u(\tilde{a}, t_1(\tilde{a}))(-\mathrm{d}\phi([0, t_1(\tilde{a})])) + \int_a^{a'} u(\tilde{a}, t_2(\tilde{a}))\mathrm{d}\phi([0, t_2(\tilde{a})]) = 0.$$

Taking the limit $a' \downarrow a$, we obtain (6) for all $a \in [\underline{a}, \overline{a}]$.

Since $\Gamma_a^\star = \{t_1(a), t_2(a)\}$ for all $a \in [\underline{a}, \overline{a}]$, Theorem 7 gives the FOC, for all $a \in (\underline{a}, \overline{a}]$,

$$\begin{aligned} v(a, t_1(a)) + q(a)u_a(a, t_1(a)) + q'(a)u(a, t_1(a)) = 0, \\ v(a, t_2(a)) + q(a)u_a(a, t_2(a)) + q'(a)u(a, t_2(a)) = 0. \end{aligned}$$

Solving for $q(a)$ and $q'(a)$, we get, for all $a \in (\underline{a}, \overline{a}]$,

$$\begin{aligned} q(a) &= \frac{v(a, t_1(a))u(a, t_2(a)) - v(a, t_2(a))u(a, t_1(a))}{u(a, t_1(a))u_a(a, t_2(a)) - u(a, t_2(a))u_a(a, t_1(a))}, \\ q'(a) &= \frac{v(a, t_1(a))u_a(a, t_2(a)) - v(a, t_2(a))u_a(a, t_1(a))}{u_a(a, t_1(a))u(a, t_2(a)) - u_a(a, t_2(a))u(a, t_1(a))}, \end{aligned}$$

where the denominators in the expressions for $q(a)$ and $q'(a)$ are not equal to 0, by Assumption 2. Recalling that q' is the derivative of q, we obtain (7) for all $a \in (\underline{a}, \overline{a}]$.

D.11. **Proof of Corollary 3.** We give the proof for the single-dipped case. Noting that $\rho u(a^\star(\mu), \theta_1) + (1-\rho)u(a^\star(\mu), \theta_2) = 0$ and denoting $a = a^\star(\mu)$, we infer that (5) fails if there exist $\theta_1 < \theta_2$ such that for all $a \in (a^\star(\delta_{\theta_1}), a^\star(\delta_{\theta_2}))$, we have

$$u(a, \theta_2)(V(a, \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1)) - u(a, \theta_1)(V(a, \theta_2) - V(a^\star(\delta_{\theta_2}), \theta_2)) \leq 0.$$

By Taylor's theorem and some algebra, we get

$$\begin{aligned} &u(a, \theta_2)(V(a, \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1)) - u(a, \theta_1)(V(a, \theta_2) - V(a^\star(\delta_{\theta_2}), \theta_2)) \\ &= \frac{1}{2}u_a(a, \theta^\star(a)) \Bigg(v_a(a, \theta^\star(a)) - \frac{v(a, \theta^\star(a))u_{aa}(a, \theta^\star(a))}{u_a(a, \theta^\star(a))} \\ &\quad -2\frac{v_\theta(a, \theta^\star(a))u_a(a, \theta^\star(a)) - v(a, \theta^\star(a))u_{a\theta}(a, \theta^\star(a))}{u_\theta(a, \theta^\star(a))} \Bigg) \\ &\quad \cdot (a - a^\star(\delta_{\theta_1}))(a^\star(\delta_{\theta_2}) - a)(a^\star(\delta_{\theta_2}) - a^\star(\delta_{\theta_1})) \\ &\quad + o((a - a^\star(\delta_{\theta_1}))(a^\star(\delta_{\theta_2}) - a)(a^\star(\delta_{\theta_2}) - a^\star(\delta_{\theta_1}))). \end{aligned}$$

Hence, if (9) fails at some a, then there exist $\theta_2 > \theta_1$ with $a^\star(\delta_{\theta_2}) - a > 0$ and $a - a^\star(\delta_{\theta_1}) > 0$ small enough such that (5) fails for all $\rho \in (0, 1)$.

Note that $\mathrm{d}\theta^\star(a)/\mathrm{d}a = -u_a(a, \theta^\star(a))/u_\theta(a, \theta^\star(a))$, by the implicit function theorem applied to $u(a, \theta^\star(a)) = 0$. Thus, denoting the partial derivatives of v and u_a in a by

v_a and u_{aa}, we get that the derivative of $q(a) = -v(a, \theta^\star(a))/u_a(a, \theta^\star(a))$ is given by

$$q'(a) = -\frac{v_a(a, \theta^\star(a))}{u_a(a, \theta^\star(a))} + \frac{v_\theta(a, \theta^\star(a))}{u_\theta(a, \theta^\star(a))} + \frac{v(a, \theta^\star(a)) u_{aa}(a, \theta^\star(a))}{(u_a(a, \theta^\star(a)))^2} - \frac{v(a, \theta^\star(a)) u_{a\theta}(a, \theta^\star(a))}{u_a(a, \theta^\star(a)) u_\theta(a, \theta^\star(a))}.$$

Conversely, suppose that (9), together with all other assumptions of the corollary, holds. Then, for $a > a^\star(\delta_\theta)$, we have

$$\begin{aligned}
& V(a,\theta) - \frac{v(a, \theta^\star(a))}{u_a(a, \theta^\star(a))} u(a,\theta) - V(a^\star(\delta_\theta), \theta) \\
=& (V(\tilde{a}, \theta) + q(\tilde{a}) u(\tilde{a}, \theta))|_{a^\star(\delta_\theta)}^{a} \\
=& \int_{a^\star(\delta_\theta)}^{a} [v(\tilde{a}, \theta) + q(\tilde{a}) u_a(\tilde{a}, \theta) + q'(\tilde{a}) u(\tilde{a}, \theta)] \mathrm{d}\tilde{a} \\
\geq& \int_{a^\star(\delta_\theta)}^{a} \left[v(\tilde{a}, \theta) - \frac{v(\tilde{a}, \theta^\star(\tilde{a}))}{u_a(\tilde{a}, \theta^\star(\tilde{a}))} u_a(\tilde{a}, \theta) \right] \mathrm{d}\tilde{a} \\
& + \int_{a^\star(\delta_\theta)}^{a} \left[\frac{v(\tilde{a}, \theta^\star(\tilde{a})) u_{a\theta}(\tilde{a}, \theta^\star(\tilde{a}))}{u_a(\tilde{a}, \theta^\star(\tilde{a})) u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} - \frac{v_\theta(\tilde{a}, \theta^\star(\tilde{a}))}{u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} \right] u(\tilde{a}, \theta) \mathrm{d}\tilde{a} \\
=& \int_{a^\star(\delta_\theta)}^{a} \int_{\theta}^{\theta^\star(\tilde{a})} \left[\frac{v(\tilde{a}, \theta^\star(\tilde{a}))}{u_a(\tilde{a}, \theta^\star(\tilde{a}))} u_{a\theta}(\tilde{a}, \tilde{\theta}) - v_\theta(\tilde{a}, \tilde{\theta}) \right] \mathrm{d}\tilde{\theta} \mathrm{d}\tilde{a} \\
& + \int_{a^\star(\delta_\theta)}^{a} \int_{\theta}^{\theta^\star(\tilde{a})} \left[\frac{v_\theta(\tilde{a}, \theta^\star(\tilde{a}))}{u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} - \frac{v(\tilde{a}, \theta^\star(\tilde{a})) u_{a\theta}(\tilde{a}, \theta^\star(\tilde{a}))}{u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} \right] u_\theta(\tilde{a}, \tilde{\theta}) \mathrm{d}\tilde{\theta} \mathrm{d}\tilde{a} \\
=& \int_{a^\star(\delta_\theta)}^{a} \int_{\theta}^{\theta^\star(\tilde{a})} \left[\frac{v_\theta(\tilde{a}, \theta^\star(\tilde{a}))}{u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} - \frac{v_\theta(\tilde{a}, \tilde{\theta})}{u_\theta(\tilde{a}, \tilde{\theta})} \right] u_\theta(\tilde{a}, \tilde{\theta}) \mathrm{d}\tilde{\theta} \mathrm{d}\tilde{a} \\
& + \int_{a^\star(\delta_\theta)}^{a} \int_{\theta}^{\theta^\star(\tilde{a})} \frac{v(\tilde{a}, \theta^\star(\tilde{a}))}{-u_a(\tilde{a}, \theta^\star(\tilde{a}))} \left[\frac{u_{a\theta}(\tilde{a}, \theta^\star(\tilde{a}))}{u_\theta(\tilde{a}, \theta^\star(\tilde{a}))} - \frac{u_{a\theta}(\tilde{a}, \tilde{\theta})}{u_\theta(\tilde{a}, \tilde{\theta})} \right] u_\theta(\tilde{a}, \tilde{\theta}) \mathrm{d}\tilde{\theta} \mathrm{d}\tilde{a} > 0,
\end{aligned}$$

where the first and last equalities are by rearrangement, the second and third equalities are by the fundamental theorem of calculus, the first inequality is by (9) and substitution of $q(\tilde{a})$ and $q'(\tilde{a})$, and the last inequality is by our assumptions imposed in the corollary.

By Taylor's theorem, we have, for $\theta_1 < \theta_2$ and $a \in (a^\star(\delta_{\theta_1}), a^\star(\delta_{\theta_2}))$,

$$\begin{aligned}
u(a, \theta_2)&(V(a, \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1)) - u(a, \theta_1)(V(a, \theta_2) - V(a^\star(\delta_{\theta_2}), \theta_2)) \\
&= \left[V(a^\star(\theta_2), \theta_1) - \frac{v(a^\star(\theta_2), \theta_2)}{u_a(a^\star(\theta_2), \theta_2)} u(a^\star(\theta_2), \theta_1) - V(a^\star(\delta_{\theta_1}), \theta_1) \right] \\
&\qquad \cdot (-u_a(a^\star(\delta_{\theta_2}), \theta_2))(a^\star(\delta_{\theta_2}) - a) + o(a^\star(\delta_{\theta_2}) - a).
\end{aligned}$$

Hence (5) holds for sufficiently small $\rho > 0$.

Online Appendix

E. Additional Examples

Example 4 ($\Gamma^\star$ might not be compact; with the "wrong" selection from Q, Γ might not be compact either.)**.** Consider the linear case with $V(a) = 0$ if $a < 1/2$ and $V(a) = (a - 1/2)^2$ otherwise. Let ϕ be uniform on $\Theta = \{0, 1/2, 1\}$. Note that $p(\theta) = V(\theta, \theta)$ solves (D). Moreover, $Q(a) = 0$ if $a < 1/2$ and $Q(a) = [a - 1/2, a]$ otherwise. Our selection from Q is given by $q(a) = 0$ if $a < 1/2$ and $q(a) = 2a - 1$ otherwise. Note that this selection is from the interior of $Q(a)$ for all $a \in (1/2, 1)$.

With our selection, the contact set $\Gamma = ([0, 1/2] \times \{0, 1/2\}) \cup \{1, 1\}$ is compact, but $\Gamma^\star = \Gamma \setminus (\{0, 1/2\} \cup \{1/2, 0\})$ is not compact. Note also that there exists an optimal outcome with $\mathrm{supp}(\pi) = \Gamma$ (e.g., the outcome that induces action 1 with certainty if $\theta = 1$, and induces action $a \in [0, 1/2]$ with densities $4 - 8a$ and $8a$ if $\theta = 0$ and $\theta = 1/2$, respectively.) However, for any such an outcome there exists a conditional probability π_a such that $\mathrm{supp}(\pi_a) = \Gamma^\star_a$ for all $a \in A_\Gamma = [0, 1/2] \cup \{1\}$ (i.e., $\pi_1 = \delta_1$ and $\pi_a = (1 - 2a)\delta_0 + 2a\delta_{1/2}$ for all $a \in [0, 1/2]$).

In contrast, consider an alternative selection from Q given by $\tilde{q}(a) = 0$ if $a < 1/2$ and $\tilde{q}(a) = a$ otherwise. The associated contact set $\tilde{\Gamma} = \Gamma \cup ([1/2, 1) \times \{1\}) \setminus \{(1/2, 0)\}$ is not compact because $(1/2, 0) \notin \tilde{\Gamma}$, and $A_{\tilde{\Gamma}}$ contains redundant actions $a \in (1/2, 1)$ that are not induced by any optimal outcome.

Example 5 (The FOC (1) might not hold on all of Γ.)**.** Consider the linear receiver case. Let ϕ be uniform on $\Theta = \{0, 1/3, 1\}$, and $V(a, \theta) = -a^2$ if $\theta = 0$ and $V(a, \theta) = -a/3 + a^2 - 3a^3/4$ if $\theta \in \{1/3, 1\}$. Since $V(a, \theta) \leq 0$ for all (a, θ) with equality on $\Gamma = \{0, 0\} \cup (\{0, 2/3\} \times \{1/3, 1\})$ and strict inequality elsewhere, the unique optimal outcome reveals state 0 (which induces action 0) and pools states 1/3 and 1 (which induces action 2/3). The contact set is Γ, so $\Gamma_0 = \Theta$. But (1) cannot hold on Γ, because the following system of equations does not have a solution $(q(0), q'(0))$,

$$\begin{cases} 0 - q(0) + q'(0)0 = 0, \\ -\frac{1}{3} - q(0) + q'(0)\frac{1}{3} = 0, \\ -\frac{1}{3} - q(0) + q'(0)1 = 0. \end{cases}$$

Intuitively, $\theta^\star(a) \in (\min \Gamma_a, \max \Gamma_a)$ is an interior case, so the FOC is valid on $\Gamma^\star_a = \Gamma_a$; while $\theta^\star(a) \in \{\min \Gamma_a, \max \Gamma_a\}$ is a boundary case, so the FOC may be invalid on Γ_a, but it is still valid on $\Gamma_a^\star = \{\theta^\star(a)\}$ given our selection $q(a) = v(a,\theta^\star(a))/(-u_a(a,\theta^\star(a))$.

Example 6 (Without Assumption 4, Γ might not be compact and the FOC might fail.)**.** Let ϕ be uniform on $\Theta = \{0, 1/3, 2/3, 1\}$; $u(a,0) = -a$, $u(a,1/3) = u(a,2/3) = 1/2 - a$, and $u(a,1) = 1 - a$; and $V(a,0) = V(a,1/3) = 0$, $V(a,2/3) = a - 1/2$, and $V(a,1) = a - 1$. Note that $p = 0$ solves (D). Moreover, $Q(a) = 0$ if $a < 1/2$, $Q(a) = 1$ if $a > 1/2$, and $Q(1/2) = [0,1]$.

For any selection $\tilde{q}$ from Q, the associated contact set $\tilde{\Gamma}$ satisfies $\tilde{\Gamma}_a = \{0, 1/3\}$ if $a < 1/2$, $\tilde{\Gamma}_a = \{2/3, 1\}$ if $a > 1/2$, and $\tilde{\Gamma}_{1/2} = \{1/3, 2/3\} \cup (\{\tilde{q}(1/2)\} \cap \Theta)$. The set $\tilde{\Gamma}$ is not compact because $(1/2, 0) \notin \tilde{\Gamma}$ if $\tilde{q}(1/2) \neq 0$ and $(1/2, 1) \notin \tilde{\Gamma}$ if $\tilde{q}(1/2) \neq 1$. Moreover, there does not exist a full measure set where the FOC holds: since the full-disclosure outcome $\pi = (\delta_{(0,0)} + \delta_{(1/2,1/3)} + \delta_{(1/2,2/3)} + \delta_{(1,1)})/4$ is supported on $\tilde{\Gamma}$, it is optimal, but the FOC does not hold at $(1/2, 1/3)$ if $\tilde{q}(a) \neq 0$ and at $(1/2, 2/3)$ if $\tilde{q}(a) \neq 1$.

F. Additional Proofs

F.1. **Proof of Lemma 2.** (1) $\implies$ (2). It is easy to see that Assumption 2 for $\mu = \delta_\theta$ such that $u(a,\theta) = 0$ yields (10). Similarly, Assumption 2 for $\mu = \rho\delta_\theta + (1-\rho)\delta_\theta$ such that $u(a,\theta) < 0 < u(a,\theta')$ and $\rho u(a,\theta) + (1-\rho)u(a,\theta') = 0$ yields (11).

(2) $\implies$ (1). By Lemma 12, for any $a \in A$ and $\mu \in \Delta(\Theta)$ such that $\int u(a,\theta)\mathrm{d}\mu = 0$, there exists $\lambda_\mu \in \Delta(\Delta(\Theta))$ such that $\int \eta \mathrm{d}\lambda_\mu = \mu$, and for each $\eta \in \mathrm{supp}(\lambda_\mu)$ there exist $\theta, \theta' \in \Theta$ and $\rho \in [0,1]$ such that $\eta = \rho\delta_\theta + (1-\rho)\delta_{\theta'}$ and

$$\rho u(a,\theta) + (1-\rho)u(a,\theta') = 0. \tag{15}$$

It suffices to show that

$$\rho u_a(a,\theta) + (1-\rho)u_a(a,\theta') < 0. \tag{16}$$

There are two cases to consider. First, if $\rho u(a,\theta) = 0$, then (16) follows from (10) and (15). Second, if $\rho u(a,\theta) \neq 0$, then (16) follows from (11) and (15).

(3) $\implies$ (1). Notice that

$$\int u(a,\theta)\mathrm{d}\mu = 0 \iff \int \tilde{u}(a,\theta)\mathrm{d}\mu = 0.$$

Hence, if $\tilde{u}_a(a,\theta) < 0$ for all (a,θ) and $\int u(a,\theta)\mathrm{d}\mu = 0$, then

$$\int u_a(a,\theta)\mathrm{d}\mu = g(a)\int \tilde{u}_a(a,\theta)\mathrm{d}\mu + g'(a)\int \tilde{u}(a,\theta)\mathrm{d}\mu = g(a)\int \tilde{u}_a(a,\theta)\mathrm{d}\mu < 0,$$

yielding Assumption 2.

(1) $\implies$ (3). We rely on the following lemma.

Lemma 20. *If Assumptions 1 and 2 hold, then there exists a continuous function $\gamma(a)$ such that*

$$u_a(a,\theta) + \gamma(a)u(a,\theta) < 0, \quad \textit{for all } (a,\theta) \in A \times \Theta. \tag{17}$$

Given this lemma, the required g is given by

$$g(a) = e^{-\int_0^a \gamma(\tilde{a})\mathrm{d}\tilde{a}},$$

as follows from

$$\tilde{u}_a(a,\theta) = \frac{\partial}{\partial a}\left(\frac{u(a,\theta)}{e^{-\int_0^a \gamma(\tilde{a})\mathrm{d}\tilde{a}}}\right) = \frac{u_a(a,\theta) + \gamma(a)u(a,\theta)}{e^{-\int_0^a \gamma(\tilde{a})\mathrm{d}\tilde{a}}} < 0.$$

Proof of Lemma 20. Fix $a \in [0,1]$. Let $M_+([0,1])$ be the set of positive Borel measures on $[0,1]$. Define the set $C \subset \mathbb{R}^3$ as follows

$$C = \left\{\left(\int u(a,\theta)\mathrm{d}\mu, \int u_a(a,\theta)\mathrm{d}\mu - z, \int \mathrm{d}\mu\right) \;\middle|\; \mu \in M_+([0,1]),\ z \geq 0\right\}.$$

Clearly, C is a convex cone.

Moreover, C is closed, because $u(a,\theta)$ and $u_a(a,\theta)$ are continuous in θ. To see this, let sequences $\mu_n \in M_+([0,1])$ and $z_n \in \mathbb{R}^n_+$ be such that

$$\int u(a,\theta)\mathrm{d}\mu_n \to c_1,\ \int u_a(a,\theta)\mathrm{d}\mu_n - z_n \to c_2,\ \int \mathrm{d}\mu_n \to c_3$$

for some $(c_1,c_2,c_3) \in \mathbb{R}^3$. It follows from $\int \mathrm{d}\mu_n \to c_3$ that all μ_n belong to a compact subset of positive measures whose total variation is bounded by $\sup_n \int \mathrm{d}\mu_n$, and hence, up to extraction of a subsequence, $\mu_n \to \mu \in M_+([0,1])$, with $\int \mathrm{d}\mu = c_3$. Since $u(a,\theta)$ and $u_a(a,\theta)$ are continuous in θ, we get $\int u(a,\theta)\mathrm{d}\mu_n \to \int u(a,\theta)\mathrm{d}\mu = c_1$ and $\int u_a(a,\theta)\mathrm{d}\mu_n \to \int u_a(a,\theta)\mathrm{d}\mu$. Hence, $z_n \to \int u_a(a,\theta)\mathrm{d}\mu - c_2 = z \geq 0$. In sum,

$$\int u(a,\theta)\mathrm{d}\mu = c_1,\ \int u_a(a,\theta)\mathrm{d}\mu - z = c_2,\ \int \mathrm{d}\mu = c_3,$$

showing that C is closed.

Next, notice that Assumption 2 implies that $(0,0,1) \notin C$. Thus, by the separation theorem (e.g., Corollary 5.84 in Aliprantis and Border 2006), there exists $y \in \mathbb{R}^3$ such that, for all $\mu \in M_+([0,1])$ and $z \geq 0$,

$$0y_1 + 0y_2 + 1y_3 < 0 \leq \left(\int u(a,\theta)\mathrm{d}\mu\right) y_1 + \left(\int u_a(a,\theta)\mathrm{d}\mu - z\right) y_2 + \left(\int \mathrm{d}\mu\right) y_3,$$

or equivalently

$$\begin{aligned} u(a,\theta)y_1 + u_a(a,\theta)y_2 + y_3 &\geq 0, \quad \text{for all } \theta \in [0,1], \\ -y_2 &\geq 0, \\ y_3 &< 0. \end{aligned} \tag{18}$$

We now show that there exists a scalar $\gamma(a) \in \mathbb{R}$ satisfying

$$u_a(a,\theta) + \gamma(a)u(a,\theta) < 0, \quad \text{for all } \theta \in [0,1]. \tag{19}$$

There are two cases. First, if $y_2 < 0$ then $\gamma(a) = y_1/y_2 \in \mathbb{R}$ satisfies (19). Second, if $y_2 = 0$ then (18) implies that

$$u(a,\theta)y_1 \geq -y_3 > 0, \quad \text{for all } \theta \in [0,1].$$

Thus, we have either (i) $u(a,\theta) > 0$ for all $\theta \in [0,1]$, so, taking into account continuity of $u(a,\theta)$ and $u_a(a,\theta)$ in θ,

$$\gamma(a) = \min_{\theta\in[0,1]} \left\{-\frac{u_a(a,\theta)}{u(a,\theta)}\right\} - 1 \in \mathbb{R}$$

satisfies (19); or (ii) $u(a,\theta) < 0$ for all $\theta \in [0,1]$, so

$$\gamma(a) = \max_{\theta\in[0,1]} \left\{-\frac{u_a(a,\theta)}{u(a,\theta)}\right\} + 1 \in \mathbb{R}$$

satisfies (19).

It remains to show that if for all $a \in [0,1]$ there exists $\gamma(a) \in \mathbb{R}$ satisfying (19), then there exists a continuous function $\tilde{\gamma} : [0,1] \to \mathbb{R}$ satisfying (19). Define a correspondence $\varphi : [0,1] \rightrightarrows \mathbb{R}$,

$$\varphi(a) = \{r \in \mathbb{R} : u_a(a,\theta) + ru(a,\theta) < 0, \quad \text{for all } \theta \in [0,1]\}.$$

Note that φ is nonempty valued by assumption, and is clearly convex valued. In addition, φ has open lower sections, because for each $r \in \mathbb{R}$ the set

$$\{a \in [0,1] : u_a(a,\theta) + ru(a,\theta) < 0, \quad \text{for all } \theta \in [0,1]\}$$

is open, since u_a and u are continuous on the compact set $[0,1] \times [0,1]$. Thus, by Browder's Selection Theorem (Theorem 17.63 in Aliprantis and Border 2006), φ admits a continuous selection $\tilde{\gamma}$, which by construction satisfies (19). □

F.2. **Proof of Lemma 1.**

We first prove dual attainment, and then prove primal attainment and no-duality-gap, which are more standard. The proof of Lemma 1 remains valid wihout Assumption 4 and when Θ is an arbitrary compact metric space.

Dual attainment. For any nonempty, compact interval $I \subset \mathbb{R}$, let $B(A, I)$ denote the set of bounded, measurable functions q such that $q(A) \subset I$. Define

$$F_I = \left\{ p \in \mathbb{R}^\Theta : \exists q \in B(A, I), p(\theta) = \sup_{a \in A} \{V(a, \theta) + q(a) u(a, \theta)\} \text{ for all } \theta \in \Theta \right\},$$

and consider the problem

$$\inf_{p \in F_I} \int_\Theta p(\theta) \mathrm{d}\phi(\theta). \tag{D'}$$

(D') is a reformulated version of (D) that involves only the function p. Denote the value of (D) by $\mathbf{D}$, and denote the value of (D') (which depends on the interval I) by $\mathbf{D'_I}$. We first show that there exists a solution $p \in F_I$ to (D') and that p, together with any measurable selection q from Q, is feasible for (D) (Lemma 23). Finally, we show that for a sufficiently large interval $I = [-C, C]$, $\mathbf{D} = \mathbf{D'_I}$ (Lemma 24), so (p, q) solve (D).

For the moment, let $I = [-C, C]$ for an arbitrary choice of $C > 0$. The existence of a solution to (D') relies on the following lemma.

Lemma 21. *The family of functions F_I is uniformly bounded and equicontinuous. Thus, there exists a convergent sequence $p_n \to p$ such that $p_n \in F_I$ for all n, $p \in C(\Theta)$, and $\int p d\phi = \mathbf{D'_I}$.*

Proof. For each $p \in F_I$, there exists $q \in B(A, I)$ such that $p(\theta) = \sup_{a \in A}\{V(a, \theta) + q(a) u(a, \theta)\}$ for all $\theta \in \Theta$, and thus

$$\sup_{\theta \in \Theta} |p(\theta)| \leq \sup_{(a,\theta) \in A \times \Theta} |V(a, \theta) + q(a) u(a, \theta)| \leq \sup_{(a,\theta,r) \in A \times \Theta \times I} |V(a, \theta) + r u(a, \theta)|,$$

This upper bound is finite by compactness of $A \times \Theta \times I$ and continuity of V and u, so the family of functions F_I is uniformly bounded.

Next, since V and u are continuous on the compact set $A \times \Theta$, they are uniformly continuous on $A \times \Theta$. This implies that there exists an increasing, continuous function

$w:\mathbb{R}_+ \to \mathbb{R}_+$ (known as the modulus of continuity) such that $w(0)=0$ and, for every $\theta,\theta' \in \Theta$ and $a \in A$, we have

$$\begin{aligned}
&\left|V(a,\theta)+q(a)u(a,\theta)-V(a,\theta')-q(a)u(a,\theta')\right| \\
\leq &\left|V(a,\theta)-V(a,\theta')\right|+|q(a)|\left|u(a,\theta)-u(a,\theta')\right| \\
\leq &\left|V(a,\theta)-V(a,\theta')\right|+C\left|u(a,\theta)-u(a,\theta')\right| \\
\leq &w\left(d\left(\theta,\theta'\right)\right),
\end{aligned}$$

where $d(\theta,\theta')$ denotes the distance between $\theta,\theta' \in \Theta$. We claim that for all $p \in F_I$ and $\theta,\theta' \in \Theta$, we have $|p(\theta)-p(\theta')| \leq w(d(\theta,\theta'))$. Indeed, for each $a \in A$,

$$\begin{aligned}
p(\theta) = &\sup_{\tilde{a}\in A}\left\{V(\tilde{a},\theta)+q(\tilde{a})u(\tilde{a},\theta)\right\} \\
\geq &V(a,\theta)+q(a)u(a,\theta) \\
\geq &V(a,\theta')+q(a)u(a,\theta')-w\left(d\left(\theta,\theta'\right)\right).
\end{aligned}$$

Taking the supremum over $a \in A$ gives $p(\theta) \geq p(\theta') - w(d(\theta,\theta'))$, and switching the roles of θ and θ' gives $|p(\theta)-p(\theta')| \leq w(d(\theta,\theta'))$. Consequently, the family of functions F_I is equicontinuous.

Now consider a minimizing sequence $p_n \in F_I$ such that $\int p_n(\theta)\mathrm{d}\phi(\theta) \to \mathbf{D_I'}$. Since Θ is compact, and F_I is uniformly bounded and equicontinuous, Arzelà-Ascoli's theorem implies that there exists a subsequence p_{n_k} uniformly converging to some function $p \in C(\Theta)$, and thus $\int p_{n_k}(\theta)\mathrm{d}\phi(\theta) \to \int p(\theta)\mathrm{d}\phi(\theta) = \mathbf{D_I'}$. □

Now fix $p \in C(\Theta)$ as in Lemma 21. To show that $p \in F_I$, recall the correspondence

$$Q(a)=\left\{r \in I : p(\theta) \geq V(a,\theta)+ru(a,\theta) \text{ for all } \theta \in \Theta\right\}, \quad \text{for all } a \in A. \tag{20}$$

We first derive some properties of this correspondence, which will also be used in the subsequent analysis.[58]

Lemma 22. *The correspondence Q is nonempty, convex and compact valued, and upper hemicontinuous, and hence admits a measurable selection q.*

Proof. By Lemma 21, there exists a sequence $p_n \in F_I$, such that $p_n \to p$ uniformly. For every $n \in \mathbb{N}$, define

$$Q_n(a)=\left\{r \in I : p_n(\theta) \geq V(a,\theta)+ru(a,\theta) \text{ for all } \theta \in \Theta\right\} \quad \text{for all } a \in A.$$

[58]In the current proof, the correspondence Q is defined in reference to the price function p defined in Lemma 21. In Appendix B, Q is defined in reference to an optimal price function. We will see that p is indeed optimal, so the definitions are equivalent.

For every $a \in A$ and $n \in \mathbb{N}$, we have $Q_n(a) \neq \emptyset$ since $p_n \in F_I$. Fix $a \in A$, and for every $n \in \mathbb{N}$ fix $r_n \in Q_n(a) \subset I$. Since I is compact, there exists a convergent subsequence $r_{n_k} \to r$ with $r \in I$. For all $k \in \mathbb{N}$, we have $p_{n_k}(\theta) \geq V(a, \theta) + r_{n_k} u(a, \theta)$ for all $\theta \in \Theta$, which implies that $p(\theta) \geq V(a, \theta) + ru(a, \theta)$ for all $\theta \in \Theta$. This shows that $r \in Q(a)$. Since a was arbitrary, it follows that Q is nonempty valued.

Next, for all $a \in A$, $Q(a)$ is closed because V and u are continuous, and $Q(a)$ is convex because it is defined by a linear inequality. Now consider a sequence (a_n, r_n) in the graph of Q such that $(a_n, r_n) \to (a, r)$. For every $n \in \mathbb{N}$, we have $p(\theta) \geq V(a_n, \theta) + r_n u(a_n, \theta)$ for all $\theta \in \Theta$. By continuity of V and u, this implies that $p(\theta) \geq V(a, \theta) + ru(a, \theta)$ for all $\theta \in \Theta$. This shows that (a, r) is in the graph of Q. By the closed-graph theorem, the correspondence Q is upper hemicontinuous. Finally, by Theorem 18.20 in Aliprantis and Border (2006), Q admits a measurable selection q. □

We next show that (p, q) is feasible for (D), for any measurable selection q from Q. Consider the problem

$$\inf\left\{\int p(\theta)\mathrm{d}\phi(\theta) : p \in C(\Theta), \exists q \in B(A, I) \text{ such that } (p, q) \text{ satisfy (D1)}\right\}. \tag{D”}$$

Denote the the value of (D”) by $\mathbf{D''_I}$

Lemma 23. *For every measurable selection q from Q, we have*

$$p(\theta) = \sup_{a \in A}\{V(a, \theta) + q(a)\, u(a, \theta)\}, \quad \textit{for all } \theta \in \Theta,$$

and hence $p \in F_I$, and (p, q) satisfy (D1). Therefore, $\mathbf{D'_I} = \mathbf{D''_I}$.

Proof. Fix a measurable selection q from Q, and let $\hat{p}(\theta) := \sup_{a \in A}\{V(a, \theta) + q(a)\, u(a, \theta)\}$ for all $\theta \in \Theta$. Note that $\hat{p} \in F_I \subset C(\Theta)$, and that $p(\theta) \geq \hat{p}(\theta)$ for all $\theta \in \Theta$ by construction of Q. Conversely, if $\hat{p}(\theta) < p(\theta)$ for some $\theta \in \Theta$, then $\int_\Theta \hat{p}(\theta)\, \mathrm{d}\phi(\theta) < \int_\Theta p(\theta)\, \mathrm{d}\phi(\theta)$ (by continuity of p and $\hat{p}$, together with full support of ϕ), which contradicts the definition of p. Hence, $p = \hat{p}$, establishing the first part of the lemma. Next, since p is continuous and (p, q) satisfy (D1) for every selection q from Q, p is feasible for (D”). Moreover, for any $\tilde{p} \in C(\Theta)$ that satisfies (D1) for some $q \in B(A, I)$, the function $\hat{p}$ defined above satisfies $\hat{p}(\theta) \leq \tilde{p}(\theta)$ for all $\theta \in \Theta$, so we have $\int p \mathrm{d}\phi \leq \int \hat{p} \mathrm{d}\phi \leq \int \tilde{p} \mathrm{d}\phi$. Hence, p solves (D”), and $\mathbf{D''_I} = \mathbf{D'_I}$. □

The following lemma implies that for a sufficiently large interval $I = [-C, C]$, we have $\mathbf{D} = \mathbf{D}''_{\mathbf{I}}$, so that the pair (p, q) constructed in Lemma 23 solve (D). This proves dual attainment.

Lemma 24. *There exists $C > 0$ such that $\mathbf{D} = \mathbf{D}''_{\mathbf{I}}$, where $I = [-C, C]$.*

Proof. It is enough to find $C > 0$ such that the additional constraint $q(a) \in [-C, C]$ for all $a \in A$ is non-binding in (D).

Define

$$\widetilde{q}(a, \theta) = \begin{cases} \frac{v(a,\theta)}{-u_a(a,\theta)}, & u(a,\theta) = 0, \\ \frac{V(a^\star(\delta_\theta),\theta) - V(a,\theta)}{u(a,\theta)}, & u(a,\theta) \neq 0. \end{cases}$$

Recall that Assumption 2 requires that $u_a(a, \theta) < 0$ when $u(a, \theta) = 0$; so $\tilde{q}(a, \theta)$ is well-defined. Since $a^\star(\delta_\theta)$ is a unique maximizer of a continuous function $U(a, \theta)$, it is continuous in θ by Berge's theorem.

We now prove that $\widetilde{q}$ is continuous at each $(a, \theta) \in A \times \Theta$. First, $\widetilde{q}$ is continuous at each (a, θ) such that $u(a, \theta) \neq 0$, because V, u, and $a^\star$ are continuous. Next, consider (a, θ) such that $u(a, \theta) = 0$, or equivalently $a = a^\star(\delta_\theta)$. For each $(a', \theta') \in A \times \Theta$, there exists $\hat{a}$ between $a^\star(\delta_{\theta'})$ and a' such that

$$[V(a^\star(\delta_{\theta'}), \theta') - V(a', \theta')]u_a(\hat{a}, \theta') = -v(\hat{a}, \theta')u(a', \theta'),$$

by the mean value theorem applied to the function

$$[V(a^\star(\delta_{\theta'}), \theta') - V(\tilde{a}, \theta')]u(a', \theta') - [V(a^\star(\delta_{\theta'}), \theta') - V(a', \theta')]u(\tilde{a}, \theta'),$$

where the argument $\tilde{a}$ is between $a^\star(\delta_{\theta'})$ and a'. Thus,

$$\widetilde{q}(a', \theta') - \widetilde{q}(a, \theta) = \frac{v(\hat{a}, \theta')}{-u_a(\hat{a}, \theta')} - \frac{v(a, \theta)}{-u_a(a, \theta)}.$$

If $(a', \theta') \to (a, \theta)$ then $(\hat{a}, \theta') \to (a, \theta)$, because $a^\star(\delta_\theta)$ is continuous in θ. Hence, $\widetilde{q}(a', \theta') \to \widetilde{q}(a, \theta)$, because v and u_a are continuous. This shows that $\widetilde{q}$ is continuous on $A \times \Theta$.

Next, define $\underline{C} = \min_{(a,\theta) \in A \times \Theta} \widetilde{q}(a, \theta) - 1$ and $\overline{C} = \max_{(a,\theta) \in A \times \Theta} \widetilde{q}(a, \theta) + 1$, where $\underline{C}$ and $\overline{C}$ are finite because $\widetilde{q}$ is continuous on the compact set $A \times \Theta$. To see why the constraint $q(a) \leq \overline{C}$ is non-binding, notice that decreasing $q(a)$ weakly tightens (D1) for θ such that $u(a, \theta) < 0$, and weakly relaxes (D1) for θ such that $u(a, \theta) \geq 0$. If $q(a) > \overline{C}$ and $u(a, \theta) < 0$, then $V(a, \theta) + q(a)u(a, \theta) < V(a^\star(\delta_\theta), \theta)$, so decreasing $q(a)$ to $\overline{C}$ does not strictly tighten (D1), because $p(\theta) \geq V(a^\star(\delta_\theta), \theta)$ by feasibility.

Thus, since the dual objective function does not depend on $q(a)$, adding the constraint $q(a) \le \overline{C}$ does not affect the value of (D). Similarly, increasing $q(a)$ to $\underline{C}$ does not strictly tighten (D1) for θ such that $u(a,\theta) > 0$, and weakly relaxes (D1) for θ such that $u(a,\theta) \le 0$; so we can add the non-binding constraint $q(a) \ge \underline{C}$.

In sum, adding the constraint $q(A) \subset I = [-C, C]$ where $C = \max\{|\underline{C}|, |\overline{C}|\}$ does not alter the value of (D), so $\mathbf{D} = \mathbf{D}''_{\mathbf{I}}$. □

Primal attainment. The set of feasible solutions to (P) is clearly nonempty, as $\pi(a,\theta) = \phi(\theta)\delta_{a^*(\phi)}(a)$ (i.e., no disclosure) is feasible. Since the set $A \times \Theta$ is compact, the set of probability measures $\Delta(A \times \Theta)$ is also compact (in the weak* topology), by Prokhorov's theorem. The constraint map in (P1) is continuous because it is a projection, and the constraint map in (P2) is continuous because $u(a,\theta)$ is continuous in (a,θ); so the set of feasible solutions is a closed subset of the compact set $\Delta(A\times\Theta)$, and is thus itself compact. Since $V(a,\theta)$ is continuous, the objective function is continuous, and thus attains its maximum on the compact set of feasible solutions.

No duality gap. Consider a tightened dual problem in which $(p,q) \in C(\Theta) \times C(A)$, and let F_D be the set of feasible solutions of the original dual problem: $(p,q) \in C(\Theta) \times B(A)$ satisfying (D1). Let F_P be the set of feasible solutions of the primal problem: $\pi \in \Delta(A \times \Theta)$ satisfying (P1) and (P2). Weak duality follows easily:

$$\begin{aligned}
\inf_{(p,q)\in F_D,\ q\in C(A)} \int_\Theta p(\theta)\mathrm{d}\phi(\theta) &\ge \inf_{(p,q)\in F_D} \int_\Theta p(\theta)\mathrm{d}\phi(\theta) \\
&= \inf_{(p,q)\in F_D,\ \pi\in F_P} \int_{A\times\Theta} p(\theta)\mathrm{d}\pi(a,\theta) \\
&\ge \sup_{(p,q)\in F_D,\ \pi\in F_P} \int_{A\times\Theta} (V(a,\theta) + q(a)u(a,\theta))\mathrm{d}\pi(a,\theta) \\
&= \sup_{\pi\in F_P} \int_{A\times\Theta} V(a,\theta)\mathrm{d}\pi(a,\theta),
\end{aligned} \tag{21}$$

where the first inequality holds because the original dual problem is more relaxed than the tightened dual problem, the first equality holds by (P1), the second inequality holds by (D1), and the second equality holds by (P2).

By the Riesz representation theorem, the space $M_r(A \times \Theta)$ of regular, signed Borel measures on the compact set $A \times \Theta$ with the total variation norm is the topological dual of the space $C(A \times \Theta)$ of continuous functions on $A \times \Theta$ with the supremum norm. Moreover, the set of (positive) measures in $M_r(A\times\Theta)$, $M_r(\Theta)$, and $M_r(A)$ are all weak* closed, so the positive cones in the primal variable space and the primal constraint space are closed.

The tightened dual problem has a finite value, since it is bounded below by the value of the primal problem and is bounded above by $\overline{V} := \max_{a,\theta} V(a,\theta)$, as $(p,q) = (\overline{V},0)$ is feasible. Moreover, since $u, V \in C(A \times \Theta)$, there is an interior feasible solution $(p,q) = (1+\overline{V},0)$ of the tightened dual problem, as the function $p(\theta) - q(a)u(a,\theta) - V(a,\theta) = 1+\overline{V} - V(a,\theta)$ lies in the interior of the positive cone of $C(A\times\Omega)$. Together with the closedness properties established in the previous paragraph, this implies that the (generalized) Slater condition is satisfied for the tightened dual problem, so there is no duality gap by Corollary 3.14 in Anderson and Nash (1987): that is,

$$\inf_{(p,q)\in F_D,\ q\in C(A)} \int_\Theta p(\theta)\mathrm{d}\phi(\theta) = \sup_{\pi\in F_P} \int_{A\times\Theta} V(a,\theta)\mathrm{d}\pi(a,\theta),$$

It follows that all inequalities in (21) hold with equality. Finally, as the original dual and primal problems admit solutions, we have

$$\min_{(p,q)\in F_D} \int_\Theta p(\theta)\mathrm{d}\phi(\theta) = \max_{\pi\in F_P} \int_{A\times\Theta} V(a,\theta)\mathrm{d}\pi(a,\theta).$$

F.3. **Proof of Lemma 15.** We consider the case where $u_{a\theta}/u_\theta$ and v_θ/u_θ are increasing in θ; the case where $u_{a\theta}/u_\theta$ and v_θ/u_θ are decreasing in θ is analogous and thus omitted.

Fix $\theta_1 < \theta_2 < \theta_3$ and a such that $u(a,\theta_1) < 0 < u(a,\theta_3)$. The inequality $|S| > 0$ follows from the following displayed equations:

$$u(a,\theta_3) - u(a,\theta_1) = \int_{\theta_1}^{\theta_3} u_\theta(a,\theta)\mathrm{d}\theta > 0,$$

where the inequality holds by Assumption 4;

$$\begin{vmatrix} u(a,\theta_1) & u(a,\theta_3) \\ u_a(a,\theta_1) & u_a(a,\theta_3) \end{vmatrix} = -u(a,\theta_3)u_a(a,\theta_1) + u(a,\theta_1)u_a(a,\theta_3) > 0,$$

where the inequality holds by part (2) of Lemma 2;

$$\begin{vmatrix} v(a,\theta_1) & v(a,\theta_3) \\ u(a,\theta_1) & u(a,\theta_3) \end{vmatrix} = u(a,\theta_3)v(a,\theta_1) - u(a,\theta_1)v(a,\theta_3) > 0,$$

where the inequality holds by Assumption 4;

$$-\begin{vmatrix} v(a,\theta_2) - v(a,\theta_1) & v(a,\theta_3) - v(a,\theta_2) \\ u(a,\theta_2) - u(a,\theta_1) & u(a,\theta_3) - u(a,\theta_2) \end{vmatrix}$$

$$= (v(a,\theta_3) - v(a,\theta_2))(u(a,\theta_2) - u(a,\theta_1)) - (v(a,\theta_2) - v(a,\theta_1))(u(a,\theta_3) - u(a,\theta_2))$$
$$= \int_{\theta_2}^{\theta_3}\int_{\theta_1}^{\theta_2}(v_\theta(a,\tilde{\theta})u_\theta(a,\theta) - v_\theta(a,\theta)u_\theta(a,\tilde{\theta}))\mathrm{d}\theta\mathrm{d}\tilde{\theta} \geq (>)0,$$

where the inequality holds by Assumption 4 and (strict) monotonicity of v_θ/u_θ in θ;

$$\begin{vmatrix} u(a,\theta_2) - u(a,\theta_1) & u(a,\theta_3) - u(a,\theta_2) \\ u_a(a,\theta_2) - u_a(a,\theta_1) & u_a(a,\theta_3) - u_a(a,\theta_2) \end{vmatrix}$$
$$= (u(a,\theta_2) - u(a,\theta_1))(u_a(a,\theta_3) - u_a(a,\theta_2)) - (u(a,\theta_3) - u(a,\theta_2))(u_a(a,\theta_2) - u_a(a,\theta_1))$$
$$= \int_{\theta_2}^{\theta_3}\int_{\theta_1}^{\theta_2}(u_\theta(a,\theta)u_{a\theta}(a,\tilde{\theta}) - u_\theta(a,\tilde{\theta})u_{a\theta}(a,\theta))\mathrm{d}\theta\mathrm{d}\tilde{\theta} \geq (>)0,$$

where the inequality holds by Assumption 4 and (strict) monotonicity of $u_{a\theta}/u_\theta$ in θ;

$$\frac{\begin{vmatrix} v(a,\theta_1) & v(a,\theta_2) & v(a,\theta_3) \\ u(a,\theta_1) & u(a,\theta_2) & u(a,\theta_3) \\ u_a(a,\theta_1) & u_a(a,\theta_2) & u_a(a,\theta_3) \end{vmatrix}}{\begin{vmatrix} u(a,\theta_1) & u(a,\theta_3) \\ u_a(a,\theta_1) & u_a(a,\theta_3) \end{vmatrix}}(u(a,\theta_3) - u(a,\theta_1))$$
$$= -\begin{vmatrix} v(a,\theta_2) - v(a,\theta_1) & v(a,\theta_3) - v(a,\theta_2) \\ u(a,\theta_2) - u(a,\theta_1) & u(a,\theta_3) - u(a,\theta_2) \end{vmatrix}$$
$$+\frac{\begin{vmatrix} v(a,\theta_1) & v(a,\theta_3) \\ u(a,\theta_1) & u(a,\theta_3) \end{vmatrix}}{\begin{vmatrix} u(a,\theta_1) & u(a,\theta_3) \\ u_a(a,\theta_1) & u_a(a,\theta_3) \end{vmatrix}}\begin{vmatrix} u(a,\theta_2) - u(a,\theta_1) & u(a,\theta_3) - u(a,\theta_2) \\ u_a(a,\theta_2) - u_a(a,\theta_1) & u_a(a,\theta_3) - u_a(a,\theta_2) \end{vmatrix},$$

where the equality holds by rearrangement.

F.4. **Proof of Lemma 16.** We consider the case where $u_{a\theta}/u_\theta$ and v_θ/u_θ are increasing in θ; the case where $u_{a\theta}/u_\theta$ and v_θ/u_θ are decreasing in θ is analogous and thus omitted.

Fix $\theta_1 < \theta_2 < \theta_3$ and $a_2 > a_1$ such that $u(a_1,\theta_1) < 0 < u(a_1,\theta_2)$. The inequality $|R| > 0$ follows from the following displayed equations:

$$u(a_1,\theta_3) - u(a_1,\theta_1) = \int_{\theta_1}^{\theta_3} u_\theta(a_1,\theta)\mathrm{d}\theta > 0,$$

where the inequality holds by Assumption 4;

$$\begin{aligned}
&\begin{vmatrix} u(a_1,\theta_1) & u(a_1,\theta_3) \\ u(a_2,\theta_1) & u(a_2,\theta_3) \end{vmatrix} \\
&= -u(a_1,\theta_3)u(a_2,\theta_1) + u(a_1,\theta_1)u(a_2,\theta_3) \\
&= -g(a_1)\tilde{u}(a_1,\theta_3)g(a_2)\tilde{u}(a_2,\theta_1) + g(a_1)\tilde{u}(a_1,\theta_1)g(a_2)\tilde{u}(a_2,\theta_3) \\
&= g(a_1)g(a_2)[-\tilde{u}(a_1,\theta_3)(\tilde{u}(a_2,\theta_1) - \tilde{u}(a_1,\theta_1)) + \tilde{u}(a_1,\theta_1)(\tilde{u}(a_2,\theta_3) - \tilde{u}(a_1,\theta_3))] \\
&= g(a_1)g(a_2)\int_{a_1}^{a_2} [-\tilde{u}(a_1,\theta_3)\tilde{u}_a(a,\theta_1) + \tilde{u}(a_1,\theta_1)\tilde{u}_a(a,\theta_3)]\mathrm{d}a > 0,
\end{aligned}$$

where the inequality and the second equality hold by parts (2) and (3) of Lemma 2;

$$\begin{aligned}
&\begin{vmatrix} V(a_2,\theta_1) - V(a_1,\theta_1) & V(a_2,\theta_3) - V(a_1,\theta_3) \\ u(a_1,\theta_1) & u(a_1,\theta_3) \end{vmatrix} \\
&= u(a_1,\theta_3)\int_{a_1}^{a_2} v(a,\theta_1)\mathrm{d}a - u(a_1,\theta_1)\int_{a_1}^{a_2} v(a,\theta_3)\mathrm{d}a > 0,
\end{aligned}$$

where the inequality holds by Assumption 4;

$$\begin{aligned}
&-\begin{vmatrix} V(a_2,\theta_2) - V(a_1,\theta_2) - V(a_2,\theta_1) + V(a_1,\theta_1) & V(a_2,\theta_3) - V(a_1,\theta_3) - V(a_2,\theta_2) + V(a_1,\theta_2) \\ u(a_1,\theta_2) - u(a_1,\theta_1) & u(a_1,\theta_3) - u(a_1,\theta_2) \end{vmatrix} \\
&= (V(a_2,\theta_3) - V(a_1,\theta_3) - V(a_2,\theta_2) + V(a_1,\theta_2))(u(a_1,\theta_2) - u(a_1,\theta_1)) \\
&\quad -(V(a_2,\theta_2) - V(a_1,\theta_2) - V(a_2,\theta_1) + V(a_1,\theta_1))(u(a_1,\theta_3) - u(a_1,\theta_2)) \\
&= \int_{a_1}^{a_2}\int_{\theta_2}^{\theta_3}\int_{\theta_1}^{\theta_2} (v_\theta(a,\tilde{\theta})u_\theta(a_1,\theta) - v_\theta(a,\theta)u_\theta(a_1,\tilde{\theta}))\mathrm{d}\theta\mathrm{d}\tilde{\theta}\mathrm{d}a \geq (>)0,
\end{aligned}$$

where the inequality holds by Assumption 4 and (strict) monotonicity of v_θ/u_θ in θ;

$$\begin{aligned}
&\begin{vmatrix} u(a_1,\theta_2) - u(a_1,\theta_1) & u(a_1,\theta_3) - u(a_1,\theta_2) \\ u(a_2,\theta_2) - u(a_2,\theta_1) & u(a_2,\theta_3) - u(a_2,\theta_2) \end{vmatrix} \\
&= (u(a_1,\theta_2) - u(a_1,\theta_1))(u(a_2,\theta_3) - u(a_2,\theta_2)) - (u(a_1,\theta_3) - u(a_1,\theta_2))(u(a_2,\theta_2) - u(a_2,\theta_1)) \\
&= \int_{\theta_2}^{\theta_3}\int_{\theta_1}^{\theta_2} (u_\theta(a_1,\theta)u_\theta(a_2,\tilde{\theta}) - u_\theta(a_1,\tilde{\theta})u_\theta(a_2,\theta))\mathrm{d}\theta\mathrm{d}\tilde{\theta} \geq (>)0,
\end{aligned}$$

where the inequality holds by Assumption 4 and (strict) monotonicity of $u_{a\theta}/u_\theta$ in θ, which imply that, for $a_2 > a_1$ and $\tilde{\theta} > \theta$, we have

$$\ln\frac{u_\theta(a_1,\theta)u_\theta(a_2,\tilde{\theta})}{u_\theta(a_1,\tilde{\theta})u_\theta(a_2,\theta)} = \int_{a_1}^{a_2}\frac{\partial}{\partial a}[\ln u_\theta(a,\tilde{\theta}) - \ln u_\theta(a,\theta)]\mathrm{d}a = \int_{a_1}^{a_2}\left[\frac{u_{a\theta}(a,\tilde{\theta})}{u_\theta(a,\tilde{\theta})} - \frac{u_{a\theta}(a,\theta)}{u_\theta(a,\theta)}\right]\mathrm{d}a \geq (>)0;$$

$$\frac{\begin{vmatrix} V(a_2,\theta_1)-V(a_1,\theta_1) & -(V(a_2,\theta_2)-V(a_1,\theta_2)) & V(a_2,\theta_3)-V(a_1,\theta_3) \\ -u(a_1,\theta_1) & u(a_1,\theta_2) & -u(a_1,\theta_3) \\ u(a_2,\theta_1) & -u(a_2,\theta_2) & u(a_2,\theta_3) \end{vmatrix}}{\begin{vmatrix} u(a_1,\theta_1) & u(a_1,\theta_3) \\ u(a_2,\theta_1) & u(a_2,\theta_3) \end{vmatrix}}(u(a_1,\theta_3)-u(a_1,\theta_1))$$

$$= -\begin{vmatrix} V(a_2,\theta_2)-V(a_1,\theta_2)-V(a_2,\theta_1)+V(a_1,\theta_1) & V(a_2,\theta_3)-V(a_1,\theta_3)-V(a_2,\theta_2)+V(a_1,\theta_2) \\ u(a_1,\theta_2)-u(a_1,\theta_1) & u(a_1,\theta_3)-u(a_1,\theta_2) \end{vmatrix}$$

$$+\frac{\begin{vmatrix} V(a_2,\theta_1)-V(a_1,\theta_1) & V(a_2,\theta_3)-V(a_1,\theta_3) \\ u(a_1,\theta_1) & u(a_1,\theta_3) \end{vmatrix}}{\begin{vmatrix} u(a_1,\theta_1) & u(a_1,\theta_3) \\ u(a_2,\theta_1) & u(a_2,\theta_3) \end{vmatrix}}\begin{vmatrix} u(a_1,\theta_2)-u(a_1,\theta_1) & u(a_1,\theta_3)-u(a_1,\theta_2) \\ u(a_2,\theta_2)-u(a_2,\theta_1) & u(a_2,\theta_3)-u(a_2,\theta_2) \end{vmatrix},$$

where the equality holds by rearrangement.

F.5. **Proof of Lemma 17.** Fix $\theta_1 < \theta_2 < \theta_3$ and $a_2 > a_1$ such that $u(a_1,\theta_1) < 0 < u(a_1,\theta_3)$. The first claimed inequality follows as in the proof of Lemma 16, by Assumption 2 and $u(a_1,\theta_1) < 0 < u(a_1,\theta_3)$. We thus focus on the second and third inequalities.

As in the proof of Lemma 16, Assumption 4 and monotonicity of $u_{a\theta}/u_\theta$ in θ yield

$$u(a_1,\theta_3) > u(a_1,\theta_2) > u(a_1,\theta_1),$$
$$\frac{u(a_2,\theta_3)-u(a_2,\theta_2)}{u(a_1,\theta_3)-u(a_1,\theta_2)} \geq \frac{u(a_2,\theta_2)-u(a_2,\theta_1)}{u(a_1,\theta_2)-u(a_1,\theta_1)}.$$

There are three cases to consider.

(1) $u(a_1,\theta_2)=0$. In this case, $u(a_2,\theta_2)<0$, by Assumption 2. Thus,

$$u(a_2,\theta_2)u(a_1,\theta_1) > 0 = u(a_2,\theta_1)u(a_1,\theta_2),$$
$$u(a_2,\theta_3)u(a_1,\theta_2) = 0 > u(a_2,\theta_2)u(a_1,\theta_3).$$

(2) $u(a_1,\theta_2)>0$. In this case, as follows from the proof of Lemma 16,

$$u(a_2,\theta_2)u(a_1,\theta_1) > u(a_2,\theta_1)u(a_1,\theta_2),$$

by Assumption 2 and $u(a_1,\theta_1) < 0 < u(a_1,\theta_2)$. Thus,

$$\frac{u(a_2,\theta_3)-u(a_2,\theta_2)}{u(a_1,\theta_3)-u(a_1,\theta_2)} \geq \frac{u(a_2,\theta_2)-u(a_2,\theta_1)}{u(a_1,\theta_2)-u(a_1,\theta_1)} > \frac{u(a_2,\theta_2)}{u(a_1,\theta_2)}$$

$$\implies u(a_2,\theta_3)u(a_1,\theta_2) > u(a_2,\theta_2)u(a_1,\theta_3).$$

(3) $u(a_1,\theta_2) < 0$. In this case, as follows from the proof of Lemma 16,

$$u(a_2,\theta_3)u(a_1,\theta_2) > u(a_2,\theta_2)u(a_1,\theta_3),$$

by Assumption 2 and $u(a_1,\theta_2) < 0 < u(a_1,\theta_3)$. Thus,

$$\frac{u(a_2,\theta_2)}{u(a_1,\theta_2)} > \frac{u(a_2,\theta_3)-u(a_2,\theta_2)}{u(a_1,\theta_3)-u(a_1,\theta_2)} \geq \frac{u(a_2,\theta_2)-u(a_2,\theta_1)}{u(a_1,\theta_2)-u(a_1,\theta_1)}$$
$$\implies u(a_2,\theta_2)u(a_1,\theta_1) > u(a_2,\theta_1)u(a_1,\theta_2).$$

F.6. **Proof of Lemma 18.** Fix $\theta_1 < \theta_2 < \theta_3$ and $a_2 < a_1$ such that $\theta_1 \leq \theta^\star(a_1) \leq \theta_3$. As in the proof of Lemma 16, Assumption 4 and monotonicity of v_θ/u_θ in θ yield

$$V(a_1,\theta_j) - V(a_2,\theta_j) > 0 \quad \text{for } j = 1,2,3, \tag{22}$$

$$u(a_1,\theta_3) > u(a_1,\theta_2) > u(a_1,\theta_1), \tag{23}$$

$$\begin{aligned}&\frac{V(a_1,\theta_3)-V(a_2,\theta_3)-V(a_1,\theta_2)+V(a_2,\theta_2)}{u(a_1,\theta_3)-u(a_1,\theta_2)}\\ \leq\ &\frac{V(a_1,\theta_2)-V(a_2,\theta_2)-V(a_1,\theta_1)+V(a_2,\theta_1)}{u(a_1,\theta_2)-u(a_1,\theta_1)}.\end{aligned} \tag{24}$$

There are two cases to consider.

(1) $u(a_1,\theta_2) \geq 0$. In this case, we have

$$\frac{u(a_1,\theta_1)}{V(a_1,\theta_1)-V(a_2,\theta_1)} < \frac{u(a_1,\theta_2)}{V(a_1,\theta_2)-V(a_2,\theta_2)},$$

by (22) and $u(a_1,\theta_1) < 0 \leq u(a_1,\theta_2)$, and

$$\frac{u(a_1,\theta_2)}{V(a_1,\theta_2)-V(a_2,\theta_2)} < \frac{u(a_1,\theta_3)}{V(a_1,\theta_3)-V(a_2,\theta_3)},$$

by

$$\begin{aligned}u(a_1,\theta_2)(V(a_1,\theta_3)-V(a_2,\theta_3)) \leq\ & u(a_1,\theta_2)\frac{u(a_1,\theta_3)-u(a_1,\theta_1)}{u(a_1,\theta_2)-u(a_1,\theta_1)}(V(a_1,\theta_2)-V(a_2,\theta_2))\\ <\ & u(a_1,\theta_3)(V(a_1,\theta_2)-V(a_2,\theta_2)),\end{aligned}$$

where the first inequality holds by (24), $V(a_1,\theta_1) > V(a_2,\theta_1)$, $u(a_1,\theta_3) > u(a_1,\theta_2)$, and $u(a_1,\theta_2) \geq 0$, and the second inequality holds by $V(a_1,\theta_2) > V(a_2,\theta_2)$, $u(a_1,\theta_3) > u(a_1,\theta_2)$, and $u(a_1,\theta_1) < 0$.

(2) $u(a_1, \theta_2) \leq 0$. In this case, we have

$$\frac{u(a_1, \theta_2)}{V(a_1, \theta_2) - V(a_2, \theta_2)} < \frac{u(a_1, \theta_3)}{V(a_1, \theta_3) - V(a_2, \theta_3)},$$

by (22) and $u(a_1, \theta_2) \leq 0 < u(a_1, \theta_3)$, and

$$\frac{u(a_1, \theta_1)}{V(a_1, \theta_1) - V(a_2, \theta_1)} < \frac{u(a_1, \theta_2)}{V(a_1, \theta_2) - V(a_2, \theta_2)},$$

by

$$\begin{aligned}-u(a_1, \theta_2)(V(a_1, \theta_1) - V(a_2, \theta_1)) &\leq - u(a_1, \theta_2)\frac{u(a_1, \theta_3) - u(a_1, \theta_1)}{u(a_1, \theta_3) - u(a_1, \theta_2)}(V(a_1, \theta_2) - V(a_2, \theta_2)) \\ &< - u(a_1, \theta_1)(V(a_1, \theta_2) - V(a_2, \theta_2)),\end{aligned}$$

where the first inequality holds by (24), $V(a_1, \theta_3) > V(a_2, \theta_3)$, $u(a_1, \theta_3) > u(a_1, \theta_2)$, and $u(a_1, \theta_2) \leq 0$, and the second inequality holds by $V(a_1, \theta_2) > V(a_2, \theta_2)$, $u(a_1, \theta_2) > u(a_1, \theta_1)$, and $u(a_1, \theta_3) > 0$.

F.7. **Proof for Example 2.** First, notice that the outcome π is implementable. (P1) holds because, for all $a \in [\underline{a}, a]$

$$\pi_a = \frac{\mathrm{d}\phi([0, t_1(a)])}{\mathrm{d}\phi([0, t_1(a)] + \mathrm{d}\phi([a, 1])}\delta_{t_1(a)} + \frac{\mathrm{d}\phi([a, 1])}{\mathrm{d}\phi([0, t_1(a)] + \mathrm{d}\phi([a, 1])}\delta_{t_1(a)},$$
$$\alpha_\pi([a, 1]) = \phi([0, t_1(a)]) + \phi([a, 1]),$$

as follows from $\kappa\phi([0, t_1(a)]) = (1 - \kappa)\phi([a, 1])$, which implies that $\kappa \mathrm{d}\phi([0, t_1(a)]) = (1 - \kappa)\mathrm{d}\phi([a, 1])$ and that t_1 is a continuous, strictly decreasing function. (P2) holds because, for all $a \in [\underline{a}, 1]$,

$$\mathbb{E}_{\pi_a}[u(a, \theta)] = \mathbb{E}_{\pi_a}[\mathbf{1}\{\theta \geq a\} - \kappa] = \pi_a([a, 1]) - \kappa = 0.$$

Consider now any other implementable outcome $\tilde{\pi}$. By (P2), there exists $\tilde{\pi}_a$ with $\tilde{\pi}_a([a, 1]) \geq \kappa$, as otherwise $\mathbb{E}_{\tilde{\pi}_a}[u(a, \theta)] < 0$. Thus, by (P1), $\alpha_{\tilde{\pi}}([a, 1]) \leq \phi([a, 1])/\kappa$, as follows from

$$\phi([a, 1]) = \int_A \tilde{\pi}_{\tilde{a}}([a, 1])\mathrm{d}\alpha_{\tilde{\pi}}(\tilde{a}) \geq \int_a^1 \tilde{\pi}_{\tilde{a}}([a, 1])\mathrm{d}\alpha_{\tilde{\pi}}(\tilde{a}) \geq \kappa\alpha_{\tilde{\pi}}([a, 1]).$$

Since $\alpha_\pi([a, 1]) = \phi([a, 1])/\kappa$, it follows that α_π first-order stochastically dominates $\alpha_{\tilde{\pi}}$, and thus, for an increasing V,

$$\int_{A\times\Theta} V(a)\mathrm{d}\pi(a, \theta) = \int_A V(a)\mathrm{d}\alpha_\pi(a) \geq \int_A V(a)\mathrm{d}\alpha_{\tilde{\pi}}(a) = \int_{A\times\Theta} V(a)\mathrm{d}\pi(a, \theta),$$

showing that π is optimal.

F.8. **Proof of Lemma 19.** We give the proof for the single-dipped case. The proof remains valid if Assumption 4 is replaced with strict single-crossing of $u(a,\theta)$ in θ. Let π^n be any optimal outcome, so that $\text{supp}(\pi^n) \subset \Gamma^n$. Since the set of compact subsets of a compact set is compact (in the Hausdorff topology), taking a subsequence if necessary, Γ^n converges to some compact set $\overline{\Gamma} \subset A \times \Theta$. Since the set of implementable outcomes is compact (in the weak* topology), taking a subsequence if necessary, π^n converges weakly to some implementable outcome π. Finally, since $\Gamma^n \to \overline{\Gamma}$, $\pi^n \to \pi$, and $\text{supp}(\pi^n) \subset \Gamma^n$, it follows that $\text{supp}(\pi) \subset \overline{\Gamma}$, by Box 1.13 in Santambrogio (2015).

We claim that π is optimal under v. Since v^n converges uniformly to v, for each $\delta > 0$ there exists $n_\delta \in \mathbb{N}$ such that, for all $n \geq n_\delta$, we have $|v_n(a,\theta) - v(a,\theta)| \leq \delta$ for all (a,θ). Since π^n is optimal under v^n, for each implementable outcome $\tilde{\pi}$ we have

$$\begin{aligned}
\int_{A\times\Theta}\int_0^a v(\tilde{a},\theta)\mathrm{d}\pi^n(a,\theta) &\geq \int_{A\times\Theta}\int_0^a v^n(\tilde{a},\theta)\mathrm{d}\pi^n(a,\theta) - \delta \\
&\geq \int_{A\times\Theta}\int_0^a v^n(\tilde{a},\theta)\mathrm{d}\tilde{\pi}(a,\theta) - \delta \\
&\geq \int_{A\times\Theta}\int_0^a v(\tilde{a},\theta)\mathrm{d}\pi(a,\theta) - 2\delta.
\end{aligned}$$

Passing to the limit as $\delta \to 0$ and $n \to \infty$ establishes the optimality of π under v.

Let Φ be a subset of A such that $a \in \Phi$ iff there exists a strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3) in $\overline{\Gamma}$ with $a = a_1$. Define

$$\Gamma^\dagger = \overline{\Gamma} \setminus \cup_{a\in\Phi}\left(\{a\} \times \Theta \setminus \{\theta^\star(a)\}\right).$$

We show that $\Gamma^\dagger$ is a Borel single-dipped set satisfying $\pi(\Gamma^\dagger) = 1$, and hence π is single-dipped.

First, we show that $\Gamma^\dagger$ is single-dipped. For each strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3) in $\overline{\Gamma}$, we have $a_1 \in \Phi$, and thus $(a_1,\theta) \in \Gamma^\dagger$ only if $\theta = \theta^\star(a_1)$. But then (a_1,θ_1) and (a_1,θ_3) cannot both be in $\Gamma^\dagger$, as $\theta_1 \neq \theta_3$.

Second, we show that for each strictly single-peaked triple (a_1,θ_1), (a_2,θ_2), (a_1,θ_3) in $\overline{\Gamma}$, we have $\theta^\star(a_1) = \theta_2$. Fix such a triple. Since $\Gamma^n \to \overline{\Gamma}$ and $\theta^\star(a)$ is uniformly continuous on A, for each $\delta > 0$ there exist $n \in \mathbb{N}$ and a triple (a_1^n,θ_1^n), (a_2^n,θ_2^n), (a_3^n,θ_3^n) in Γ^n such that $\theta_1^n < \theta_2^n < \theta_3^n$, $a_1^n < a_2^n$, $a_3^n < a_2^n$, $|\theta_i^n - \theta_i| \leq \delta$, and $|\theta^\star(a_i^n) - \theta^\star(a_i)| \leq \delta$ for all $i \in \{1,2,3\}$ (where $a_3 = a_1$). Hence,

$$\theta^\star(a_1) - \delta \leq \theta^\star(a_1^n) \leq \theta_2^n \leq \theta_2 + \delta \implies \theta^\star(a_1) \leq \theta_2 + 2\delta.$$

To understand the middle inequality, suppose by contradiction that $\theta^\star(a_1^n) > \theta_2^n$. Recall that, by Theorem 7, each contact set Γ^n satisfies $\min \Gamma^n_{a_1^n} \leq \theta^\star(a_1^n) \leq \max \Gamma^n_{a_1^n}$. Hence, there exists $\hat{\theta}_1^n \in \Gamma^n_{a_1^n}$ with $\hat{\theta}_1^n \geq \theta^\star(a_1^n) > \theta_2^n$ (for example, $\hat{\theta}_1^n = \max \Gamma^n_{a_1^n}$). But then Γ^n cannot be single-dipped, as it contains the strictly single-peaked triple $(a_1^n, \theta_1^n), (a_2^n, \theta_2^n), (a_1^n, \hat{\theta}_1^n)$. By an analogous argument, we get

$$\theta_2 - \delta \leq \theta_2^n \leq \theta^\star(a_3^n) \leq \theta^\star(a_1) + \delta \implies \theta^\star(a_1) \geq \theta_2 - 2\delta.$$

Since $\delta > 0$ is arbitrary, we get $\theta^\star(a_1) = \theta_2$.

Third, we show that for any two strictly single-peaked triples (a_1, θ_1), (a_2, θ_2), (a_1, θ_3) and $(\tilde{a}_1, \tilde{\theta}_1)$, $(\tilde{a}_2, \tilde{\theta}_2)$, $(\tilde{a}_1, \tilde{\theta}_3)$ in $\overline{\Gamma}$, we have $\tilde{\theta}_2 \notin (\theta_2, \theta_3)$. Suppose by contradiction that $\tilde{\theta}_2 \in (\theta_2, \theta_3)$. By the previous paragraph, $\theta^\star(a_1) = \theta_2$ and $\theta^\star(\tilde{a}_1) = \tilde{\theta}_2$. Moreover, for each $\delta > 0$, there exist $n \in \mathbb{N}$ and two triples (a_1^n, θ_1^n), (a_2^n, θ_2^n), (a_3^n, θ_3^n) and $(\tilde{a}_1^n, \tilde{\theta}_1^n)$, $(\tilde{a}_2^n, \tilde{\theta}_2^n)$, $(\tilde{a}_3^n, \tilde{\theta}_3^n)$ in Γ^n such that $|\theta_i^n - \theta_i| \leq \delta$, $|\theta^\star(a_i^n) - \theta^\star(a_i)| \leq \delta$, $|\tilde{\theta}_i^n - \tilde{\theta}_i| \leq \delta$, and $|\theta^\star(\tilde{a}_i^n) - \theta^\star(\tilde{a}_i)| \leq \delta$ for all $i \in \{1, 2, 3\}$ (where $a_3 = a_1$ and $\tilde{a}_3 = \tilde{a}_1$). Next, since $\min \Gamma^n_{a_3^n} \leq \theta^\star(a_3^n) \leq \max \Gamma^n_{a_3^n}$, there exists $\hat{\theta}_3^n \in \Gamma^n_{a_3^n}$ such that $\hat{\theta}_3^n \leq \theta^\star(a_3^n)$ (for example, $\hat{\theta}_3^n = \min \Gamma^n_{a_3^n}$). Since Γ^n is single-dipped, to reach a contradiction it suffices to show that the triple $(a_3^n, \hat{\theta}_3^n)$, $(\tilde{a}_2^n, \tilde{\theta}_2^n)$, (a_3^n, θ_3^n) (which is in Γ^n by construction) is strictly single-peaked for small enough $\delta > 0$. To see this, notice that we have

$$\begin{aligned}
\theta^\star(a_3^n) &\leq \theta^\star(a_1) + \delta = \theta_2 + \delta, \\
\theta^\star(\tilde{a}_2^n) &\geq \theta^\star(\tilde{a}_2) - \delta > \theta^\star(\tilde{a}_1) - \delta = \tilde{\theta}_2 - \delta, \\
\hat{\theta}_3^n &\leq \theta^\star(a_3^n) \leq \theta^\star(a_1) + \delta = \theta_2 + \delta, \\
\tilde{\theta}_2^n &\in [\tilde{\theta}_2 - \delta, \tilde{\theta}_2 + \delta], \\
\theta_3^n &\geq \theta_3 - \delta.
\end{aligned}$$

Thus, if $\delta \in (0, \min\{(\tilde{\theta}_2 - \theta_2)/2, (\theta_3 - \tilde{\theta}_2)/2\})$, then $a_3^n < \tilde{a}_2^n$ and $\hat{\theta}_3^n < \tilde{\theta}_2^n < \theta_3^n$, so the triple $(a_3^n, \hat{\theta}_3^n)$, $(\tilde{a}_2^n, \tilde{\theta}_2^n)$, (a_3^n, θ_3^n) is strictly single-peaked.

Fourth, we show that the set Φ is countable, and thus $\Gamma^\dagger$ is Borel. If $a_1 \in \Phi$, then there exists a strictly single-peaked triple (a_1, θ_1), (a_2, θ_2), (a_1, θ_3) in $\overline{\Gamma}$ with $\theta^\star(a_1) = \theta_2$. Let us associate with each such a_1 some rational number $r(a_1) \in (\theta_2, \theta_3)$. Since for any other strictly single-peaked triple $(\tilde{a}_1, \tilde{\theta}_1)$, $(\tilde{a}_2, \tilde{\theta}_2)$, $(\tilde{a}_1, \tilde{\theta}_3)$ in $\overline{\Gamma}$, we have $\tilde{\theta}_2 \notin (\theta_2, \theta_3)$ and, by symmetry, $\theta_2 \notin (\tilde{\theta}_2, \tilde{\theta}_3)$, we see that $(\theta_2, \theta_3) \cap (\tilde{\theta}_2, \tilde{\theta}_3) = \emptyset$ if $\theta_2 \neq \tilde{\theta}_2$. Consequently, $r(a_1) \neq r(\tilde{a}_1)$ if $a_1, \tilde{a}_1 \in \Phi$ and $a_1 \neq \tilde{a}_1$. Thus, r is a one-to-one mapping of Φ into a subset of the set of rational numbers, so Φ is countable.

Finally, we show that $\pi(\Gamma^\dagger) = 1$. Since Φ is countable and probability measures are countably additive, it suffices to show that $\pi(\{a_1\}\times\Theta\setminus\{\theta^\star(a_1)\}) = 0$ for each $a_1 \in \Phi$. In turn, this follows if for each $\varepsilon_\theta > 0$, we have

$$\pi((a_1-\varepsilon, a_1+\varepsilon)\times\Theta\setminus[\theta^\star(a_1)-\varepsilon_\theta, \theta^\star(a_1)+\varepsilon_\theta]) \to 0 \text{ as } \varepsilon \to 0.$$

Fix $a_1 \in \Phi$ and a strictly single-peaked triple (a_1, θ_1), $(a_2, \theta^\star(a_1))$, (a_1, θ_3) in $\overline{\Gamma}$. Let $\varepsilon \in (0, a_2 - a_1)$. Since $\Gamma^n \to \overline{\Gamma}$, $(a_2, \theta^\star(a_1)) \in \overline{\Gamma}$, and $a_2 > a_1$ (and thus $\theta^\star(a_2) > \theta^\star(a_1)$), there exists $n \in \mathbb{N}$ and $(a_2^n, \theta_2^n) \in \Gamma^n$ with $\theta^\star(a_1-\varepsilon) < \theta_2^n < \theta^\star(a_1+\varepsilon) < \theta^\star(a_2^n)$.

Since Γ^n is a compact single-dipped set with $\min\Gamma_a^n \le \theta^\star(a) \le \max\Gamma_a^n$ for all $a \in A_{\Gamma^n}$, the triple $(a, \min\Gamma_a^n)$, (a_2^n, θ_2^n), $(a, \max\Gamma_a^n)$ cannot be strictly single-peaked. Hence, we have the following implications:
(i) if $a \in [a_1-\varepsilon, a^\star(\delta_{\theta_2^n}))$, then $\Gamma_a^n \cap (\theta_2^n, 1] = \emptyset$;
(ii) if $a \in (a^\star(\delta_{\theta_2^n}), \theta_2+\varepsilon]$, then $\Gamma_a^n \cap [0, \theta_2^n) = \emptyset$;
(iii) if $a = a^\star(\delta_{\theta_2^n})$, then $\Gamma_a^n \cap (\theta_2^n, 1] = \emptyset$ or $\Gamma_a^n \cap [0, \theta_2^n) = \emptyset$.

Let $\chi_0^n = \pi^n((a_1-\varepsilon, a_1+\varepsilon)\times[0, \theta^\star(a_1)-\varepsilon_\theta))$ and $\chi_1^n = \pi^n((a_1-\varepsilon, a_1+\varepsilon)\times(\theta^\star(a_1)+\varepsilon_\theta, 1])$. By (P2) and condition (iii), we have $\pi^n(a^\star(\delta_{\theta_2^n})\times\Theta\setminus\{\theta_2^n\}) = 0$, and hence

$$\int_{\left\{a^\star(\delta_{\theta_2^n})\right\}\times[0,\theta_2^n]} u(a,\theta)\mathrm{d}\pi^n(a,\theta) = \int_{\left\{a^\star(\delta_{\theta_2^n})\right\}\times[\theta_2^n,1]} u(a,\theta)\mathrm{d}\pi^n(a,\theta) = 0.$$

Together with conditions (i) and (ii) (and again using (P2)), we have

$$\begin{aligned}
0 &= \int_{(a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[0,\theta_2^n]} u(a,\theta)\mathrm{d}\pi^n(a,\theta) \\
&\le \max_{(a,\theta)\in[a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[0,\theta^\star(a_1)-\varepsilon_\theta]} u(a,\theta)\chi_0^n \\
&\quad + \max_{(a,\theta)\in[a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[\theta^\star(a_1)-\varepsilon_\theta,\theta_2^n]} u(a,\theta)(1-\chi_0^n) \\
\implies \chi_0^n &\le \frac{\max_{(a,\theta)\in[a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[\theta^\star(a_1)-\varepsilon_\theta,\theta_2^n]} u(a,\theta)}{\max_{(a,\theta)\in[a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[\theta^\star(a_1)-\varepsilon_\theta,\theta_2^n]} u(a,\theta) - \max_{(a,\theta)\in[a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]\times[0,\theta^\star(a_1)-\varepsilon_\theta]} u(a,\theta)},
\end{aligned}$$

and

$$\begin{aligned}
0 &= \int_{[a^\star(\delta_{\theta_2^n}), a_1+\varepsilon)\times[\theta_2^n,1]} u(a,\theta)\mathrm{d}\pi^n(a,\theta) \\
&\ge \min_{(a,\theta)\in[a^\star(\delta_{\theta_2^n}), a_1+\varepsilon]\times[\theta^\star(a_1)+\varepsilon_\theta,1]} u(a,\theta)\chi_1^n \\
&\quad + \min_{(a,\theta)\in[a^\star(\delta_{\theta_2^n}), a_1+\varepsilon]\times[\theta_2^n,\theta^\star(a_1)+\varepsilon_\theta]} u(a,\theta)(1-\chi_1^n)
\end{aligned}$$

$$\implies \chi_1^n \le \frac{-\min_{(a,\theta)\in[a^\star(\delta_{\theta_2^n}),a_1+\varepsilon]\times[\theta_2^n,\theta^\star(a_1)+\varepsilon_\theta]} u(a,\theta)}{\min_{(a,\theta)\in[a^\star(\delta_{\theta_2^n}),a_1+\varepsilon]\times[\theta^\star(a_1)+\varepsilon_\theta,1]} u(a,\theta) - \min_{(a,\theta)\in[a^\star(\delta_{\theta_2^n}),a_1+\varepsilon]\times[\theta_2^n,\theta^\star(a_1)+\varepsilon_\theta]} u(a,\theta)},$$

where the inequalities hold because $u(a_1-\varepsilon,\theta_2^n) > 0$ and $u(a_1+\varepsilon,\theta_2^n) < 0$, by Assumption 4 and $\theta^\star(a_1-\varepsilon) < \theta_2^n < \theta^\star(a_1+\varepsilon)$.

By Assumptions 1 and 2, $u_a(a,\theta) < 0$ for all (a,θ) in a neighborhood of $(a,\theta^*(a))$. Hence, for sufficiently small ε, $u(a,\theta)$ is maximized over $a \in [a_1-\varepsilon, a^\star(\delta_{\theta_2^n})]$ at $a = a_1-\varepsilon$, and $u(a,\theta)$ is minimized over $a \in [a^\star(\delta_{\theta_2^n}), a_1+\varepsilon]$ at $a = a_1+\varepsilon$. Therefore, By the Portmanteau Theorem (Theorem 15.3 in Aliprantis and Border 2006), for sufficiently small $\varepsilon > 0$, we get

$$\pi((a_1-\varepsilon,a_1+\varepsilon)\times[0,\theta^\star(a_1)-\varepsilon_\theta)) \le \liminf_n \chi_0^n$$

$$\le \frac{\max_{\theta\in[\theta^\star(a_1)-\varepsilon_\theta,\theta^\star(a_1)]} u(a_1-\varepsilon,\theta)}{\max_{\theta\in[\theta^\star(a_1)-\varepsilon_\theta,\theta^\star(a_1)]} u(a_1-\varepsilon,\theta) - \max_{\theta\in[0,\theta^\star(a_1)-\varepsilon_\theta]} u(a_1-\varepsilon,\theta)},$$

$$\pi((a_1-\varepsilon,a_1+\varepsilon)\times(\theta^\star(a_1)+\varepsilon_\theta,1]) \le \liminf_n \chi_1^n$$

$$\le \frac{-\min_{\theta\in[\theta^\star(a_1),\theta^\star(a_1)+\varepsilon_\theta]} u(a_1+\varepsilon,\theta)}{\min_{\theta\in[\theta^\star(a_1)+\varepsilon_\theta,1]} u(a_1+\varepsilon,\theta) - \min_{\theta\in[\theta^\star(a_1),\theta^\star(a_1)+\varepsilon_\theta]} u(a_1+\varepsilon,\theta)}.$$

Next, taking into account Assumption 4, we get

$$\lim_{\varepsilon\to 0}\max_{\theta\in[\theta^\star(a_1)-\varepsilon_\theta,\theta^\star(a_1)]} u(a_1-\varepsilon,\theta) = 0, \qquad \lim_{\varepsilon\to 0}\max_{\theta\in[0,\theta^\star(a_1)-\varepsilon_\theta]} u(a_1-\varepsilon,\theta) < 0,$$

$$\lim_{\varepsilon\to 0}\min_{\theta\in[\theta^\star(a_1),\theta^\star(a_1)+\varepsilon_\theta]} u(a_1+\varepsilon,\theta) = 0, \qquad \lim_{\varepsilon\to 0}\min_{\theta\in[\theta^\star(a_1)+\varepsilon_\theta,1]} u(a_1+\varepsilon,\theta) > 0.$$

Consequently, $\pi((a_1-\varepsilon,a_1+\varepsilon)\times\Theta\setminus[\theta^\star(a_1)-\varepsilon_\theta,\theta^\star(a_1)+\varepsilon_\theta]) \to 0$ as $\varepsilon \to 0$.

F.9. **Proof for Example 3.** The optimal outcome π is unique, because there is a unique implementable outcome π with $\pi(\Gamma) = 1$. To illustrate how the argument works more generally, we suppose that ϕ has a density on $\Theta = [\underline{\theta},\overline{\theta}]$, and that there exists a bifurcation point a_0 in the interior of $A_\Gamma = [\underline{a},\overline{a}]$ such that $\Gamma_a = \{t_1(a), t_2(a)\}$ with $t_1(a) = \theta^\star(a) = t_2(a)$ for $a \in [\underline{a}, a_0]$, and $t_1(a) < \theta^\star(a) < t_2(a)$ for $a \in [a_0,\overline{a}]$ where $t_1 : (a_0,\overline{a}] \to [\underline{\theta},\theta^\star(a_0))$ is continuous, strictly decreasing, and bijective and $t_2 : (a_0,\overline{a}] \to (\theta^\star(m),\overline{\theta}]$ is continuous, strictly increasing, and bijective.

Define the continuous, strictly decreasing, and bijective inverse $t_1^{-1} : [\underline{\theta},\theta^\star(a_0)) \to (a_0,\overline{a}]$ by

$$t_1^{-1}(\theta) = \{a \in (a_0,\overline{a}] : t_1(a) = \theta\}.$$

Define the distribution functions $F(\theta) = \phi([-1, \theta])$ and $H(a) = \alpha_\pi([-1, a])$ representing measures ϕ and α_π. Define the θ-section of Γ by $\Gamma^\theta = \{a \in A : (a, \theta) \in \Gamma\}$. Recall that, for $a \in (a_0, \overline{a}]$, $\pi_a = \rho_a \delta_{t_1(a)} + (1 - \rho_a)\delta_{t_2(a)}$ with $\rho_a = u(a, t_2(a))/(u(a, t_2(a) - u(a, t_1(a)) \in (0, 1)$, by (P2).

For all $a \in (a_0, \underline{a}]$, we have $\Gamma^{t_2(a)} = \{a\}$ and $\Gamma_a = \{t_1(a), t_2(a)\}$, and thus, by (P1) and (P2),

$$\mathrm{d}F(t_2(a)) = (1 - \rho_a)\mathrm{d}H(a).$$

For all $a \in [\underline{a}, a_0)$, we have $\Gamma_a = \{\theta^\star(a)\}$ and $\Gamma^{\theta^\star(a)} = \{a, t_1^{-1}(\theta^\star(a))\}$, with $t_1^{-1}(\theta^\star(a)) \in (a_0, \overline{a}]$ and thus, by (P1) and (P2),

$$\mathrm{d}F(\theta^\star(a)) = \mathrm{d}H(a) - \rho_{t_1^{-1}(\theta^\star(a))}\mathrm{d}H(t_1^{-1}(\theta^\star(a))),$$

where the last term has a minus sign because $t_1^{-1}(\theta^\star(a))$ is decreasing in a. So,

$$\mathrm{d}H(a) = \begin{cases} \frac{1}{1-\rho_a}\mathrm{d}F(t_2(a)), & a \in (a_0, \overline{a}], \\ \mathrm{d}F(\theta^\star(a)) + \frac{\rho_{t_1^{-1}(\theta^\star(a))}}{1-\rho_{t_1^{-1}(\theta^\star(a))}}\mathrm{d}F(t_2(t_1^{-1}(\theta^\star(a)))), & a \in [\underline{a}, a_0). \end{cases}$$

Substituting $\theta^\star(a) = a$ for $a \in [\underline{a}, a_0) = [-1, 0)$, and $\rho_a = 1/2$, $t_1(a) = -a$, and $t_2(a) = 3a$ for $a \in (a_0, \overline{a}] = (0, 1]$, we obtain that α_π has the stated density h.

Finally, to see that the contact set is the stated set Γ, we invoke the following lemma from Kolotilin and Wolitzky (2020).

Lemma 25. *Functions*

$$p(\theta) = \begin{cases} T(2\theta), & \theta \in [-1, 0), \\ 3T(\frac{2}{3}\theta), & \theta \in [0, 3], \end{cases} \quad \text{and} \quad q(a) = \begin{cases} \frac{2T'(2a)}{T'(0)}, & a \in [-1, 0), \\ 2, & a \in [0, 3]. \end{cases}$$

satisfy (D1) with equality for all $(a, \theta) \in \Gamma$ and strict inequality for all $(a, \theta) \notin \Gamma$.

F.10. **Proof of Proposition 1.** Recall that most results remain valid if the condition $u_\theta(a, \theta) > 0$ in Assumption 4 is replaced with strict single-crossing of $u(a, \theta)$ in θ. Clearly, $a^\star(\mu) = \mathbb{E}_\mu[\theta]/(1 + \mathbb{E}_\mu[\theta^2])$. To ensure that Assumption 3 holds, we normalize $A = [\min_{\theta \in [\underline{\theta}, \overline{\theta}]} a^\star(\delta_\theta), \max_{\theta \in [\underline{\theta}, \overline{\theta}]} a^\star(\delta_\theta)]$. Assumptions 1 and 2 obviously hold. Moreover, since $a^\star(\delta_\theta)$ is strictly increasing on $[0, 1]$ and strictly decreasing on $[1, +\infty)$, it follows that $u(a^\star(\delta_\theta), \theta') > 0$ if $\theta < \theta' \leq 1$ and if $1 \leq \theta' < \theta$. Thus, if $\overline{\theta} \leq 1$, then $u(a, \theta)$ satisfies strict single-crossing in θ, whereas, if $\underline{\theta} \geq 1$, $u(a, \theta)$ also satisfies strict single-crossing in θ once the state is redefined as $-\theta$. So Theorems 1, 2, 6 and Theorem 4 apply.

Lemma 26 replicates Lemma 1 and Proposition 3 in Zhang and Zhou (2016).

Lemma 26. *If* $\theta_1 < \theta_2$ *and* $\theta_1\theta_2 > (<)1$*, then* $\rho V(a^\star(\delta_{\theta_1}),\theta_1)+(1-\rho)V(a^\star(\delta_{\theta_2}),\theta_2) > (<)\rho V(a^\star(\mu),\theta_1) + (1-\rho)V(a^\star(\mu),\theta_2)$ *for all* $\rho \in (0,1)$.

$a^\star(\mu)\,(\rho/\theta_1 + (1-\rho)/\theta_2)$.

Proof. For $\mu = \rho\delta_{\theta_1} + (1-\rho)\delta_{\theta_2}$, $a^\star(\mu) = (\rho\theta_1 + (1-\rho)\theta_2)/(1+\rho\theta_1^2 + (1-\rho)\theta_2^2)$. Thus, if $\theta_1 < \theta_2$ and $\theta_1\theta_2 > (<)1$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\rho}a^\star(\mu) = \frac{(\theta_2-\theta_1)(\theta_1\theta_2-1)}{(1+\rho\theta_1^2+(1-\rho)\theta_2^2)^2} > (<)0,$$
$$\frac{\mathrm{d}^2}{\mathrm{d}\rho^2}a^\star(\mu) = \frac{(\theta_2-\theta_1)(\theta_1\theta_2-1)(\theta_2^2-\theta_1^2)}{(1+\rho\theta_1^2+(1-\rho)\theta_2^2)^3} > (<)0.$$

Define $\varphi(\rho) = a^\star(\mu)\,(\rho/\theta_1 + (1-\rho)/\theta_2)$. Thus, if $\theta_1 < \theta_2$ and $\theta_1\theta_2 > (<)1$, we have

$$\varphi''(\rho) = \left(\frac{\rho}{\theta_1} + \frac{1-\rho}{\theta_2}\right)\frac{\mathrm{d}^2}{\mathrm{d}\rho^2}a^\star(\mu) + 2\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right)\frac{\mathrm{d}}{\mathrm{d}\rho}a^\star(\mu) > (<)0,$$

so φ is strictly convex (concave), and $\rho\varphi(1) + (1-\rho)\varphi(0) > (<)\varphi(\rho)$. □

If $\underline{\theta} \geq 1$, then $\theta_1\theta_2 > 1$ for all $\underline{\theta}_1 \leq \theta_1 < \theta_2$, so full disclosure is uniquely optimal by Theorem 4 and Lemma 26. Assume henceforth that $\underline{\theta} \leq 1$.

After some algebra, we get, for all a and $\theta_1 < \theta_2 < \theta_3$,

$$|S| = \frac{(\theta_3-\theta_2)(\theta_3-\theta_1)(\theta_2-\theta_1)(1-\theta_2\theta_3-\theta_1\theta_3-\theta_1\theta_2)}{\theta_1\theta_2\theta_3}$$

If $\overline{\theta} \leq 1/\sqrt{3}$ ($\underline{\theta} \geq 1/\sqrt{3}$), then $|S| > (<)0$ for all $\theta_1 < \theta_2 < \theta_3 \leq \overline{\theta}$ ($\underline{\theta} \leq \theta_1 < \theta_2 < \theta_3$), so $\Gamma^\star$ is pairwise by Theorem 1. Proposition 4 in Zhang and Zhou (2016) derives a version of this result for a finite set Θ.

Moreover, if $\overline{\theta} \leq 1/\sqrt{3}$ ($\underline{\theta} \geq 1/\sqrt{3}$), then Γ is single-dipped (-peaked), as follows from Theorem 2 with

$$y = \begin{pmatrix} u(a_2,\theta_3)u(a_1,\theta_2) - u(a_2,\theta_2)u(a_1,\theta_3) \\ u(a_2,\theta_3)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_3) \\ u(a_2,\theta_2)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_2) \end{pmatrix}$$
$$\left(y = -\begin{pmatrix} u(a_2,\theta_3)u(a_1,\theta_2) - u(a_2,\theta_2)u(a_1,\theta_3) \\ u(a_2,\theta_3)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_3) \\ u(a_2,\theta_2)u(a_1,\theta_1) - u(a_2,\theta_1)u(a_1,\theta_2) \end{pmatrix} \right),$$

because, for $a < a'$ and $\theta < \theta'$ with $\theta\theta' < 1$, we have

$$u(a', \theta')u(a, \theta) - u(a', \theta)u(a, \theta') = (a' - a)(\theta' - \theta)(1 - \theta\theta') > 0,$$

and

$$Ry = \begin{pmatrix} (a_2 - a_1)^2|S| \\ 0 \\ 0 \end{pmatrix} \gneq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \left(Ry = \begin{pmatrix} -(a_2 - a_1)^2|S| \\ 0 \\ 0 \end{pmatrix} \gneq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right).$$

Since $\Gamma^\star$ is pairwise and Γ is single-dipped (-peaked) if $\overline{\theta} \leq 1/\sqrt{3}$ ($\underline{\theta} \geq 1/\sqrt{3}$), it follow that $\Gamma^\star$ is single-dipped (-peaked) if $\overline{\theta} \leq 1/\sqrt{3}$ ($\underline{\theta} \geq 1/\sqrt{3}$). Finally, since, by Lemma 26, (5) holds for all $p \in (0, 1)$, Theorem 6 yields that, if $\overline{\theta} \leq 1/\sqrt{3}$ ($\underline{\theta} \geq 1/\sqrt{3}$), then $\Gamma^\star$ is single-dipped (-peaked) negative assortative disclosure, and the optimal outcome is unique.

F.11. **Proof of Proposition 2.** Let

$$y = \begin{cases} \left(0, \frac{1}{(\theta_2 - \theta_0)g(a_2|\theta_2)}, \frac{1}{(\theta_2 - \theta_0)g(a_2|\theta_3)}\right), & \theta_2 > \theta_0, \\ (0, 1, 0), & \theta_2 = \theta_0, \\ \left(\frac{1}{(\theta_0 - \theta_1)g(a_1|\theta_1)}, \frac{1}{(\theta_0 - \theta_2)g(a_1|\theta_2)}, 0\right), & \theta_2 < \theta_0, \end{cases}$$

where $\theta_1 < \theta_2 < \theta_3$, $a_2 < a_1$, and $\theta_1 \leq \theta_0 \leq \theta_3$. We focus on the case $\theta_0 < \theta_2$, as the other cases are analogous. The above perturbation increases action a_1, because, by strict log-submodularity of g,

$$u(a_1, \theta_2)y_2 - u(a_1, \theta_3)y_3 = \frac{g(a_1|\theta_2)}{g(a_2|\theta_2)} - \frac{g(a_1|\theta_3)}{g(a_2|\theta_3)} > 0.$$

Moreover, the same perturbation also increases the sender's expected utility for fixed a_1, a_2. This follows because

$$\begin{aligned} &(V(a_1, \theta_2) - V(a_2, \theta_2))y_2 - (V(a_1, \theta_3) - V(a_2, \theta_3))y_3 \\ =&\left(\frac{G(a_1|\theta_2) - G(a_2|\theta_2)}{(\theta_2 - \theta_0)g(a_2|\theta_2)} - \frac{G(a_1|\theta_3) - G(a_2|\theta_3)}{(\theta_3 - \theta_0)g(a_2|\theta_3)}\right) \\ >&\frac{1}{(\theta_2 - \theta_0)}\left(\frac{G(a_1|\theta_2) - G(a_2|\theta_2)}{g(a_2|\theta_2)} - \frac{G(a_1|\theta_3) - G(a_2|\theta_3)}{g(a_2|\theta_3)}\right) \\ =&\frac{1}{(\theta_2 - \theta_0)}\int_{a_2}^{a_1}\left(\frac{g(t|\theta_2)}{g(a_2|\theta_2)} - \frac{g(t|\theta_3)}{g(a_2|\theta_3)}\right)\mathrm{d}t \geq 0, \end{aligned}$$

where the first inequality is by $\theta_0 < \theta_2 < \theta_3$ and the second inequality is by log-submodularity of g. Thus, every optimal outcome is single-peaked.

F.12. **Proof of Proposition 3.** As shown by Kamenica and Gentzkow (2011), there exists an optimal outcome with a finite support. Suppose the support contains a strictly single-peaked triple (a_1, θ_1), (a_2, θ_2), (a_1, θ_3), with $\theta_1 < \theta_2 < \theta_3$, $a_1 < a_2$, and $\theta_1 < a_1 < \theta_3$. Notice that $V(a_1, \theta_3) \neq -\infty$ (so $a_1 \geq \sigma(\theta_3)$), as otherwise the sender's expected utility would be $-\infty$, which cannot be optimal. Taking into account that $\sigma(\theta) = \theta$ for $\theta \leq \theta_0$ gives $a_1 > \theta_0$. Thus, the first row in R is zero. Consider a perturbation that shifts weights $y_1 = (\theta_3 - \theta_2)\varepsilon$ and $y_3 = (\theta_2 - \theta_1)\varepsilon$ on θ_1 and θ_3 from a_1 to a_2 and shifts weight $y_2 = (\theta_3 - \theta_1)\varepsilon$ from a_2 to a_1, where ε takes the maximum value such that $y_1 \leq \pi(\{(a_1, \theta_1\})$, $y_2 \leq \pi(\{(a_2, \theta_2\})$, $y_3 \leq \pi(\{(a_1, \theta_3\})$, so that a strictly single-peaked triple is removed. This perturbation holds fixed a_1 and a_2 and thus does not change the sender's expected utility, since the first row in R is zero. Repeating such perturbations until all strictly single-peaked triples are removed (a finite number of times since $\operatorname{supp}(\pi)$ is finite) yields a single-dipped outcome that is weakly preferred by the sender.