Persuasion and Matching: Optimal Productive Transport

Anton Kolotilin

UNSW Business School

Roberto Corrao

Massachusetts Institute of Technology

Alexander Wolitzky

Massachusetts Institute of Technology

We consider general Bayesian persuasion problems where the receiver's utility is single-peaked in a one-dimensional action. We show that a signal that pools at most two states in each realization is always optimal and that such "pairwise" signals are the only solutions under a nonsingularity condition on utilities. Our core results provide conditions under which the induced receiver action is single-dipped or single-peaked on each set of nested signal realizations. We also provide conditions for the optimality of either full disclosure or negative assortative disclosure, where all signal realizations are nested. Methodologically, our results rely on novel duality and complementary slackness theorems. Our analysis extends to a general problem of assigning one-dimensional inputs to productive units, which we call "optimal productive transport." This problem covers additional applications including matching with peer effects (assigning workers to firms, students to schools, or residents to neighborhoods), robust option pricing (assigning future asset prices to price distributions), and partisan gerrymandering (assigning voters to districts).

This paper supersedes earlier papers titled "Persuasion with Non-Linear Preferences," "Persuasion as Matching," and "Assortative Information Disclosure." For helpful comments

Electronically published March 4, 2025

Journal of Political Economy, volume 133, number 4, April 2025.

https://doi.org/10.1086/734095

I. 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 et al. 2017; Kolotilin 2018; Dworczak and Martini 2019; Kleiner, Moldovanu, and Strack 2021)—so that a distribution over states can be summarized by its mean—the general, nonlinear case is far less well understood. 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 nonlinear.

This paper studies persuasion with nonlinear preferences as an instance of a general class of economic models that we call "optimal productive transport." In the persuasion context, we consider a standard setting 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 if and only if his expected marginal utility from increasing his action equals zero: that is, if and only if the receiver's first-order condition holds. This first-order approach is key for tractability. We provide three types of results, all of which have general analogues beyond persuasion.

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 nonsingularity condition on the sender's and receiver's utilities—which we call the "twist condition"—every optimal signal is pairwise.

Second, we ask when it is optimal for the sender to induce higher actions with riskier or safer prospects. That is, when the sender pools two extreme states $x_1 < x_4$ and separately pools two moderate states $x_2 \leq x_3$ such that $x_1 < x_2 \leq x_3 < x_4$, do the extreme states induce a higher action—in which case, we say that disclosure is "single-dipped," as the receiver's action

and suggestions, we thank Jakša Cvitanić, Laura Doval, 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 the Massachusetts Institute of Technology's Sloan's Program on Innovation in Markets and Organizations. Alexander Wolitzky gratefully acknowledges support from a National Science Foundation CAREER Award (1555071) and Sloan Foundation Fellowship (2017–9633).

is single-dipped on the set $\{x_1, x_2, x_3, x_4\}$—or a lower action—in which case, we say that disclosure is "single-peaked"? This question turns out to be key for understanding optimal disclosure patterns with nonlinear preferences. Our core results provide general conditions for the optimality of single-dipped disclosure (and, similarly, single-peaked disclosure). The conditions are based on the following simple idea. If disclosure is not single-dipped, then there must exist a single-peaked triple: a pair of pooled states $x_1 < x_4$ and an intervening state $x_2 \in (x_1, x_4)$ such that the induced action at x_2 (e.g., action y_2) is greater than the induced action at $\{x_1, x_4\}$ (e.g., action y_1). Our conditions ensure that any single-peaked triple can be profitably perturbed in the direction of single-dippedness by shifting weight on x_1 and x_4 from y_1 to y_2, while shifting weight on x_2 in the opposite direction.

Third, we provide conditions for the optimality of either full disclosure, where the state is always disclosed, or (more interestingly) negative assortative disclosure, where all states are paired in a negatively assortative manner, so that all prospects can be ordered from safest to riskiest, 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 pairwise 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 pattern as the solution of a pair of ordinary differential equations and provide examples where these equations admit an explicit solution.

While this paper is mainly motivated by Bayesian persuasion, the theory we develop applies equally to several other applications. We consider three: matching with peer effects (e.g., assigning workers to firms, students to schools, or residents to neighborhoods to maximize welfare), robust option pricing (assigning future asset prices to price distributions to bound the price of a derivative), and partisan gerrymandering (assigning voters to districts to maximize expected seat share). To facilitate the analysis of these applications, in section VII we recast our model in terms of assigning general "inputs" to "productive units." Table 1 explains how our general model maps to each of our applications.

Mathematically, our model combines aspects of a production problem (combining inputs to produce output) and a transportation problem (matching inputs and outputs to generate utility)—hence the name "optimal productive transport." The model is a new kind of optimal transport problem. Our key technical results are duality and complementary slackness theorems for this problem. The closest strand of the optimal transport literature is that on "martingale optimal transport" (e.g., Beiglböck, Henry-Labordère, and Penkner 2013; Galichon, Henry-Labordère,

TABLE 1
Atlas of Our Applications

Application	Input (x)	Productive Unit (μ)	Output (y)	Meaning of Single-Dippedness
Persuasion	State	Posterior	Receiver action	Riskier prospects induce higher actions
Matching with peer effects	Worker, student, or resident with ability x	Firm, school, or neighborhood	Peer effect	Diverse firms, schools, or neighborhoods are more desirable
Option pricing	Period 2 asset price	Period 2 asset price distribution	Period 1 asset price	Riskier assets are more expensive
Gerrymandering	Voter with partisanship x	District	Probability that designer's party wins district	Polarized districts are stronger

and Touzi 2014; Beiglböck and Juillet 2016), which we discuss in section III.A.[1]

The optimal productive transport framework nests a great deal of prior work, both in persuasion and in the contexts of the other applications we cover. Some key prior works include Rayo and Segal (2010), Goldstein and Leitner (2018), and Guo and Shmaya (2019) on persuasion; Arnott and Rowse (1987) and Saint-Paul (2001) on matching; Beiglböck and Juillet (2016) on option pricing; and Friedman and Holden (2008) on gerrymandering. 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—that is, that riskier or safer prospects induce higher actions—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) nonmonotone stress tests, where a regulator designs a test that pools the weakest banks that it wants to receive funding with the strongest banks (and, subsequently, pools less weak banks with less strong ones), 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

[1] A few recent papers apply optimal transport to persuasion, but these works are not very related to ours either methodologically or substantively. Malamud and Schrimpf (2022) focus on the question of when optimal signals partition a multidimensional state space; Arieli, Babichenko, and Sandomirskiy (2024) consider persuasion with multiple receivers; and Lin and Liu (2024) and Perez-Richet and Skreta (2025) consider limited sender commitment.

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.

The paper is organized as follows. Section II presents our model in the context of persuasion. Section III formulates primal and dual versions of our problem and establishes strong duality and complementary slackness. Section IV shows that pairwise signals are without loss. Sections V and VI present our main substantive results: section V provides conditions for single-dipped or single-peaked disclosure to be optimal, and section VI provides conditions for full disclosure or negative assortative disclosure to be optimal. Section VII reframes our model as "optimal productive transport" and applies it to matching, option pricing, and gerrymandering, as well as some specific persuasion problems. Section VIII concludes. Additional results, as well as all proofs, are deferred to the appendixes (apps. D–F are available online).

II. Persuasion with Nonlinear Preferences

For concreteness, we exposit our model and main results in the context of Bayesian persuasion. In section VII.A, we rephrase the model as a general problem of assigning inputs to productive units, which we call optimal productive transport. This more general framing covers our matching, option pricing, and gerrymandering applications.

A. *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 $V(y, x)$ and the receiver's utility $U(y, x)$ depend on the receiver's action $y \in Y := [0, 1]$ and the state of the world $x \in [0, 1]$. The sender and receiver share a common prior $\phi \in \Delta([0, 1])$, with support $X := \text{supp}(\phi)$.[2] A signal $\tau \in \Delta(\Delta(X))$ is a distribution over posterior beliefs $\mu \in \Delta(X)$ such that the average posterior equals the prior: $\mathbb{E}_\tau[\mu] = \phi$ (Aumann and Maschler 1995; Kamenica and Gentzkow 2011). An outcome $\pi \in \Delta(Y \times X)$ is a joint distribution over actions and states. As we will see, it is equivalent to view the sender as choosing a signal τ (the "signal-based problem") or as directly choosing an outcome π subject to an obedience constraint (the "outcome-based problem").

We impose four standard assumptions on preferences, which are similar to those in canonical unidimensional models of communication such

[2] 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 $\mu \in \Delta(X)$, its support $\text{supp}(\mu)$ is the smallest compact set of measure one.

as signaling (Spence 1973), cheap talk (Crawford and Sobel 1982), and hard information disclosure (Seidmann and Winter 1997). First, utilities are smooth.

ASSUMPTION 1. $V(y, x)$ and $u(y, x) := \partial U(y, x)/\partial y$ are three times differentiable.

Apart from the receiver's marginal utility u, we denote partial derivatives with subscripts: for example, $V_y(y, x) = \partial V(y, x)/\partial y$.

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

ASSUMPTION 2. $u(y, x)$ satisfies strict aggregate single-crossing in y: for all posteriors $\mu \in \Delta(X)$,

$$\int_X u(y, x)\mathrm{d}\mu(x) = 0 \Longrightarrow \int_X u_y(y, x)\mathrm{d}\mu(x) < 0.$$

Quah and Strulovici (2012) and Choi and Smith (2017) characterized a weak version of aggregate single-crossing. We provide an analogous characterization of strict aggregate single-crossing in appendix A. A sufficient condition is strict monotonicity of u (or equivalently strict concavity of U): that is, $u_y(y, x) < 0$ for all (y, x). In fact, appendix A shows that strict aggregate single-crossing is equivalent to strict monotonicity up to a normalization.

Third, the receiver's optimal action satisfies an interiority condition.[3]

ASSUMPTION 3. $\min_{x\in[0,1]} u(0, x) = \max_{x\in[0,1]} u(1, x) = 0$.

The key implication of assumptions 1–3 is that for any posterior μ, the receiver's optimal action $\gamma(\mu) := \arg\max_{y\in[0,1]} \mathbb{E}_\mu[U(y, x)]$ is unique and is characterized by the first-order condition

$$\int_X u(\gamma(\mu), x)\mathrm{d}\mu(x) = 0. \tag{1}$$

Our assumptions thus allow a "first-order approach" to the persuasion problem, similar to the approach of Holmström (1979) and Mirrlees (1999) to the classical moral hazard problem.[4]

Uniqueness of the receiver's optimal action implies that any signal τ induces a unique outcome π_τ and that we can define the sender's indirect utility from inducing posterior μ as

$$W(\mu) = \int_X V(\gamma(\mu), x)\mathrm{d}\mu(x).$$

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

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

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

ASSUMPTION 4. $V_y(y, x) > 0$ and $u_x(y, x) > 0$.

Together with assumptions 1–3, assumption 4 ensures that for each action y there is a unique state $\chi(y)$ such that $u(y, \chi(y)) = 0$ (i.e., the receiver's optimal action at $\chi(y)$ is y) and that $\chi(y)$ is a strictly increasing, continuous function with range $[0, 1]$.

A common interpretation of the receiver's action $y \in [0, 1]$ is that the receiver has a private type and makes a binary choice—say, whether to accept or reject a proposal—and y 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.[5]

B. Special Cases

We list some leading special cases of the model, which we return to periodically to illustrate our results:

1. The linear case (Gentzkow and Kamenica 2016): $u(y, x) = x - y$ and $V(y, x) = V(y)$. That is, $\gamma(\mu) = \mathbb{E}_\mu[x]$ and V is state-independent. This is the well-studied case where the sender's indirect utility $W(\mu)$ depends only on $\mathbb{E}_\mu[x]$.[6]
2. The linear receiver case: $u(y, x) = x - y$ 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.[7]

2a. The separable subcase (Rayo and Segal 2010): $V(y, x) = w(x)G(y)$ with $w > 0$ and $G > 0$. An interpretation is that the receiver has a private type t with distribution G and accepts a proposal if and only if $\mathbb{E}_\mu[x] \geq t$ and that the sender's utility when the proposal is accepted is $w(x)$.[8]

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

[6] More generally, utilities can be transformed to fall in the linear case if u and V_y are affine in $m(x)$ for some function m. In this case, $\gamma(\mu) = a(\mathbb{E}_\mu[m(x)])$ for some function a, and $W(\mu) = H(\mathbb{E}_\mu[m(x)]) + \mathbb{E}_\mu[l(x)]$ for $l(x) = V(0, x)$ and $H(y) = \int_0^{a(y)} V_y(\tilde{y}, m^{-1}(y))\mathrm{d}\tilde{y}$. Since $\mathbb{E}_\tau[\mathbb{E}_\mu[l(x)]] = \mathbb{E}_\phi[l(x)]$ for any signal τ, the sender's problem is the same if the state is $\tilde{x} = m(x)$, the receiver's marginal utility is $\tilde{x} - y$, and the sender's utility is $H(y)$.

[7] The assumption that $V_y(y, x) > 0$ is unnecessary in the linear receiver case.

[8] Rayo and Segal focused on the sub-subcase with the uniform distribution $G(y) = y$. They assume that the state (x, z) is two-dimensional, that the sender's and receiver's marginal utilities are $V_y(y, x, z) = z$ and $u(y, x) = x - y$, and that there are finitely many states (x, z), so generically the sender's utility can be written as $V_y(y, x) = w(x)$. Rayo (2013), Nikandrova and Pancs (2017), and Onuchic and Ray (2023) consider the separable

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

3. The state-independent sender case: $V(y, x) = V(y)$, 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: $u(y, x) = I(x)(x - y)$, 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: $u(y, x) = T(x - y)$, 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 $1 - x$, the receiver's ability is $1 - t$, the receiver's "bad luck" ε has distribution J, and the project succeeds if and only if $1 - x \leq 1 - t - \varepsilon$ or, equivalently, $\varepsilon \leq x - t$. This example fits the current subcase with V equal to the distribution of t and $T(x - y) = J(x - y) - \kappa$.

3c. The quantile sub-subcase: $u(y, x) = 1\{x \geq y\} - \kappa$, with $\kappa \in (0, 1)$. This subcase corresponds to the previous example with $J(x - y) = 1\{x \geq y\}$, so the project succeeds if and only if the receiver's ability exceeds the project's difficulty. While u is now discontinuous, this subcase arises as a limit of the translation-invariant case. Kolotilin and Wolitzky (2024b) and Yang and Zentefis (2024) study the quantile sub-subcase.

III. Optimality Conditions

This section establishes optimality conditions that form the basis for our analysis. Section III.A formulates signal-based and outcome-based primal and dual problems and shows that they are equivalent. We will make use of both formulations. Section III.B establishes our key complementary slackness theorem.

---

subcase where x is continuous and (x, z) is supported on the graph of $x \to w(x)$. Rochet and Vila (1994), Tamura (2018), Kramkov and Xu (2022), and Dworczak and Kolotilin (2024) allow more general distributions of $(x, z) \in \mathbb{R}^2$.

A. *Primal and Dual Problems*

The sender's signal-based primal problem is to find a signal $\tau \in \Delta(\Delta(X))$ to

$$\text{maximize} \quad \int_{\Delta(X)} W(\mu)\mathrm{d}\tau(\mu) \tag{P}$$

$$\text{subject to} \quad \int_{\Delta(X)} \mu\mathrm{d}\tau(\mu) = \phi. \tag{BP}$$

Here, the primal constraint (BP) is the usual "Bayes plausibility" constraint (Kamenica and Gentzkow 2011).

Next, let $L(X)$ denote the set of Lipschitz continuous functions on X. The signal-based dual problem is to find a "price function" $p \in L(X)$ to

$$\text{minimize} \quad \int_X p(x)\mathrm{d}\phi(x) \tag{D}$$

$$\text{subject to} \quad \int_X p(x)\mathrm{d}\mu(x) \geq W(\mu), \quad \text{for all } \mu \in \Delta(X). \tag{ZP}$$

The interpretation is that $p(x)$ is the shadow price of state x, and the dual constraint (ZP) is the "zero profit" condition that the sender's indirect utility from inducing any posterior μ cannot exceed the expectation of $p(x)$ under μ. This interpretation will become clearer in the general framework of section VII.A.

A preliminary result is that strong duality holds: solutions to (P) and (D) exist and give the same value.

LEMMA 1. There exists $\tau \in \Delta(\Delta(X))$ that solves (P); there exists $p \in L(X)$ that solves (D); and the values of (P) and (D) are equal: for any solutions τ of (P) and p of (D), we have

$$\int_{\Delta(X)} W(\mu)\mathrm{d}\tau(\mu) = \int_X p(x)\mathrm{d}\phi(x).$$

Lemma 1 follows by showing that $W(\mu)$ is Lipschitz continuous and applying corollary 2 of Dworczak and Kolotilin (2024), which in turn generalizes theorem 2 of Dworczak and Martini (2019) from linear persuasion problems to nonlinear ones.[9]

Next, the outcome-based primal problem is to find an outcome $\pi \in \Delta(Y \times X)$ to

[9] Corollary 2 of Dworczak and Kolotilin (2024) is proved using strong duality in an optimal transport problem, as in Villani (2009). Further duality results for persuasion problems include those of Dizdar and Kováč (2020), Kramkov and Xu (2022), Smolin and Yamashita (2023), and Galperti, Levkun, and Perego (2024).

$$\text{maximize} \quad \int_{Y\times X} V(y,x)\mathrm{d}\pi(y,x) \tag{P'}$$

$$\text{subject to} \quad \int_{Y\times\tilde{X}} \mathrm{d}\pi(y,x) = \int_{\tilde{X}} \mathrm{d}\phi(x), \quad \text{for all measurable } \tilde{X}\subset X, \tag{BP'}$$

$$\int_{\tilde{Y}\times X} u(y,x)\mathrm{d}\pi(y,x) = 0, \quad \text{for all measurable } \tilde{Y}\subset Y. \tag{OB}$$

Here, (BP') is an outcome-based version of Bayes plausibility, which says that the marginal of π on X equals the prior ϕ; and (OB) is the "obedience" constraint that the receiver's action at each posterior μ is $\gamma(\mu)$. A joint distribution π that violates (OB) is inconsistent with optimal play by the receiver, as there exists $\tilde{Y}\subset Y$ such that the receiver's play is suboptimal conditional on the event $\{y\in\tilde{Y}\}$. Conversely, for any joint distribution π that satisfies (BP') and (OB), if the sender designs a mechanism that draws (y, x) according to π and recommends action y to the receiver, it is optimal for the receiver to obey the recommendation. We therefore say that an outcome π is "implementable" if and only if it satisfies (BP') and (OB) and "optimal" if and only if it solves (P').

Finally, letting $B(Y)$ denote the set of bounded, measurable functions on Y, the outcome-based dual problem is to find $p\in L(X)$ and $q\in B(Y)$ to

$$\text{minimize} \quad \int_X p(x)\mathrm{d}\phi(x) \tag{D'}$$

$$\text{subject to} \quad p(x) \geq V(y,x) + q(y)u(y,x), \quad \text{for all } (y,x)\in Y\times X. \tag{ZP'}$$

The interpretation is that $p(x)$ is the shadow price of state x; $q(y)$ is the value of relaxing the obedience constraint at action y; and (ZP') says that $p(x)$ is no less than the sender's value from assigning state x to any action y, where this value is the sum of the sender's utility, $V(y, x)$, and the product of $q(y)$ and the amount by which obedience at y is relaxed when state x is assigned to action y, $u(y, x)$.

We now establish that the price functions in the signal-based and outcome-based formulations coincide. Hence, by lemma 1, strong duality holds in the outcome-based formulation, as well as the signal-based one.[10]

LEMMA 2. A price function $p\in L(X)$ is feasible (optimal) for (D) if and only if there exists $q\in B(Y)$ such that (p, q) is feasible (optimal) for (D').

[10] Strong duality in the outcome-based formulation is established under slightly different assumptions in lemmas 1 and 2 of Kolotilin (2018). However, a key step in the proof—that q can be taken to be bounded—is incomplete in Kolotilin (2018).

B. *Complementary Slackness*

Letting p be the optimal price function (which we will see in remark 1 is unique), define the set

$$\Lambda = \left\{\mu \in \Delta(X) : \int_X p(x)\mathrm{d}\mu(x) = W(\mu)\right\}. \tag{2}$$

Note that Λ is compact, because $\mathbb{E}_\mu[p(x)]$ and $W(\mu)$ are continuous in μ.

By lemma 1, together with (BP), a signal τ is optimal if and only if

$$\int_{\Delta(X)}\left(\int_X p(x)\mathrm{d}\mu(x) - W(\mu)\right)\mathrm{d}\tau(\mu) = 0.$$

Hence, since the integrand is nonnegative by (ZP) and Λ is compact, τ is optimal if and only if $\mathrm{supp}(\tau) \subset \Lambda$. Any posterior $\mu \notin \Lambda$ is thus excluded from the support of any optimal signal. In analogy with the optimal transport literature (e.g., sec. 3 of Ambrosio, Brué, and Semola 2021), we refer to the set Λ as the "contact set."

The following is our main technical result.

Theorem 1. There exists $q \in B(Y)$ such that

1. (p, q) is optimal for (D');
2. for all μ in Λ (and, thus, in the support of any optimal signal τ), we have

$$q(\gamma(\mu)) = -\frac{\int_X V_y(\gamma(\mu), x)\mathrm{d}\mu(x)}{\int_X u_y(\gamma(\mu), x)\mathrm{d}\mu(x)}; \tag{3}$$

3. for all nondegenerate μ in Λ (and, thus, in the support of any optimal signal τ), the function q has derivative $q'(\gamma(\mu))$ at $\gamma(\mu)$ satisfying, for all $x \in \mathrm{supp}(\mu)$,

$$V_y(\gamma(\mu), x) + q(\gamma(\mu))u_y(\gamma(\mu), x) + q'(\gamma(\mu))u(\gamma(\mu), x) = 0. \tag{4}$$

Theorem 1 is our key tool for characterizing optimal signals. Intuitively, by complementary slackness, the support of any optimal outcome π is contained in the set of points (y, x) that satisfy (ZP') with equality. Thus, if it is ever optimal to induce action y at state x—that is, if y maximizes $V(y, x) + q(y)u(y, x)$—then y must satisfy the first-order condition

$$V_y(y, x) + q(y)u_y(y, x) + q'(y)u(y, x) = 0,$$

which is just (4) with $\gamma(\mu) = y$. Moreover, taking the expectation of this equation with respect to μ yields (3). This equation simply says that $q(y)$

equals the product of the sender's expected marginal utility at y and the rate at which y increases as obedience is relaxed, where the latter term equals $-1/\mathbb{E}_{\mu}[u_y(y, x)]$ by the implicit function theorem applied to obedience. Note that in the linear case, $q(y)$ simply equals $V'(y)$, a useful point that also appears implicitly in Dworczak and Martini (2019) and explicitly in Dworczak and Kolotilin (2024).

As shown in appendix F, another implication of theorem 1 is as follows.

REMARK 1. There is a unique solution p to (D).

Lemmas 1 and 2 and theorem 1 can be compared to results in optimal transport. In standard 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 our problem, the marginal distribution over states is given (by the prior ϕ), and the problem is to find an optimal joint distribution with this marginal, where for each action the conditional distribution over states satisfies obedience. Strong duality and complementary slackness theorems are likewise key tools in optimal transport (e.g., Villani 2009, theorem 5.10), but the relevant versions of these results differ from ours.[11]

The most relevant strand of the optimal transport literature is that on martingale optimal transport (MOT). The MOT problem is to find an optimal joint distribution of two variables (e.g., y and x) with given marginals, subject to the martingale constraint that the expectation of x given y equals y. This problem coincides with our linear receiver case but with an exogenously fixed distribution of the receiver's action. Motivated by problems in mathematical finance, Beiglböck, Henry-Labordère, 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 (or in standard optimal transport). 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.[12]

[11] For example, in standard optimal transport, both dual variables appear in the dual objective function, and they are both uniquely determined.

[12] The MOT literature uses the contact set of an outcome-based dual problem. See Kolotilin, Corrao, and Wolitzky (2022) for an alternative development of the results in the current paper that relies on the contact set of our outcome-based dual, (D'). The approach in the current version, which is based on the contact set Λ of the signal-based dual, (D), turns out to be simpler.

IV. Pairwise Disclosure and the Twist Condition

Our first substantive result is that there is always an optimal signal that pools at most two states in every realized posterior and that under an additional condition every optimal signal has this property. This result simplifies the persuasion problem to a generalized matching problem, where the sender chooses what pairs of states to match together and with what weights.

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

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 x—the identity of the undefended castle—is uniformly distributed. Suppose the receiver can attack any two castles and that payoffs are $(-1, +1)$ for the sender and receiver, respectively, if the receiver attacks the undefended castle and $(+1, -1)$ 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 (4) implies that if it is optimal to induce the same action y at three states x_1, x_2, and x_3, then the vector $(V_y(y, x_1), V_y(y, x_2), V_y(y, x_3))$ must be a linear combination of the vectors $(u(y, x_1), u(y, x_2), u(y, x_3))$ and $(u_y(y, x_1), u_y(y, x_2), u_y(y, x_3))$. 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.[15]

Twist condition.—For any action y and any triple of states $x_1 < x_2 < x_3$ such that $x_1 < \chi(y) < x_3$, we have $|S| \neq 0$,[16] where

[13] See fig. 2. The "disclose-pair" pattern in fig. 2*D* is reminiscent of this example but with different weights on the states in each pair.

[14] Pairwise signals are also suboptimal in the price-discrimination problem of Bergemann, Brooks, and Morris (2015), as well as in Brzutowski (2024), where $U(y, x) = \mathbf{1}\{y \geq x\} - y$. In these models, the receiver's utility is not quasi-concave. However, a variant of Bergemann, Brooks, and Morris (2015) with smooth demand and concave monopoly profit would fit our assumptions, so pairwise signals would be optimal.

[15] The term "twist condition" is in analogy to optimal transport, where the twist condition is an analogous nonsingularity condition (e.g., definition 1.16 in Santambrogio 2015).

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

$$S := \begin{pmatrix} V_y(y, x_1) & V_y(y, x_2) & V_y(y, x_3) \\ u(y, x_1) & u(y, x_2) & u(y, x_3) \\ u_y(y, x_1) & u_y(y, x_2) & u_y(y, x_3) \end{pmatrix}. \tag{5}$$

We will apply this condition extensively in section V.

Theorem 2. 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 the contact set is pairwise and hence so is any optimal signal.

The intuition for the first part of the theorem is that for any posterior μ, 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 outcome. Figure 1 illustrates this argument for a posterior with weight on three states.[17]

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 μ with $\text{supp}(\mu) = \{x_1, x_2, x_3\}$. We can split μ into posteriors μ' and μ'' with at most binary support that both induce action $\gamma(\mu)$. For example, suppose that $\text{supp}(\mu') = \{x_1, x_2\}$ and $\text{supp}(\mu'') = \{x_1, x_3\}$. Now consider a perturbation that moves probability mass ε on x_1 from μ' to μ''. This perturbation induces nonzero marginal changes in the action at μ' and μ''. Under the twist condition, these changes have a nonzero marginal effect on the sender's expected utility, by the implicit function theorem. Therefore, either this perturbation or the reverse perturbation, where ε is replaced with $-\varepsilon$, is strictly profitable.[18]

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 2 easily implies these earlier results.[19] Note

[17] More formally, for a given posterior μ, another posterior μ' induces the same action as μ if and only if the action $\gamma(\mu)$ satisfies the first-order condition (1) at posterior μ'. Since the first-order condition is a moment condition, the set of posteriors that induce action $\gamma(\mu)$ is the set of probability distributions that satisfy one moment condition. By the Richter-Rogosinsky theorem, the extreme points of this set have at most binary support. Hence, by Choquet's theorem, μ can be written as an expectation, with respect to some measure $\sigma_\mu \in \Delta(\Delta(X))$, of distributions with at most binary support that all induce action $\gamma(\mu)$. Finally, by the measurable selection theorem, the mapping from μ to σ_μ can be taken to be measurable and can thus be used to define a pairwise signal that induces the same distribution on $Y \times X$ as any given signal τ.

[18] Formally, the second part of theorem 2 directly follows from theorem 1.

[19] Proposition 4 in Alonso and Câmara (2016) states that if $u(y, x) = x - y$ and there do not exist $\zeta \leq 0$ and $\iota \in \mathbb{R}$ such that $V_y(y, x_i) = \zeta x_i + \iota$ for $i = 1, 2, 3$, then it is not optimal to induce action y at states x_1, x_2, and x_3. This result is too strong as stated, and it is not correct

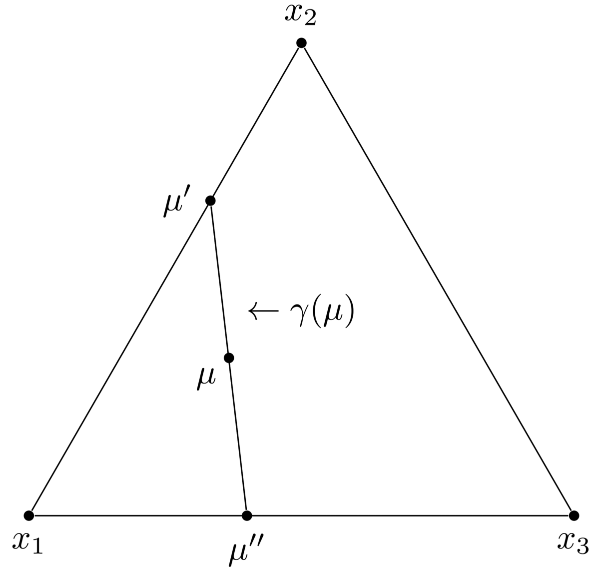

FIG. 1.—Pairwise signals are without loss. The optimal action at any posterior on the line between μ' and μ'' equals $\gamma(\mu)$, so splitting μ into μ' and μ'' eliminates a nonbinary-support posterior without changing the outcome.

that the twist condition always fails in the linear case, where $|S| = 0$. Hence, in the linear case, theorem 2 never rules out pooling multiple states, and indeed pooling multiple states is often optimal (e.g., Kolotilin et al. 2017).[20]

An immediate corollary of theorem 2 is that no disclosure is generically suboptimal when there are at least three states, because for a fixed action y, a generic vector $(V_y(y, x))_{x\in X}$ with $|X| \geq 3$ coordinates cannot be expressed as a linear combination of two fixed vectors $(u(y, x))_{x\in X}$ and $(u_y(y, x))_{x\in X}$ (for any standard notion of genericity, e.g., Hunt, Sauer, and Yorke 1992, 222), as is required by (4). Moreover, in the linear receiver and state-independent sender cases, if no disclosure is optimal for all priors then the sender's and receiver's utilities must take a particular nongeneric form: they must fall in the linear case, as defined in footnote 6, with a concave V.

COROLLARY 1. For any prior ϕ with $|\text{supp}(\phi)| \geq 3$ and any u, no disclosure is suboptimal for generic V_y. Moreover, in the linear receiver

unless ζ is also allowed to be positive. Theorem 2 implies this corrected version of Alonso and Câmara's result.

[20] Of course, theorem 2 shows that even when pooling multiple states is optimal, there also exists an optimal pairwise signal, where the "multistate pool" is split into pairs. Conversely, if multiple posteriors all induce the same action, they can be pooled without affecting the outcome.

and state-independent sender cases, no disclosure is optimal for all priors $\phi \in \Delta([0,1])$ if and only if there exist $m, l: [0,1] \to \mathbb{R}$ and concave $H: \mathbb{R} \to \mathbb{R}$ such that

$$W(\mu) = H(\mathbb{E}_\mu[m(x)]) + \mathbb{E}_\mu[l(x)], \quad \text{for all } \mu \in \Delta([0,1]). \tag{6}$$

Given Kamenica and Gentzkow's concavification result, corollary 1 implies that, for generic utilities, the sender's indirect utility is not concave in the posterior when there are more than two states. Note that corollary 1 allows the case where u and V_y 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.

V. 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 riskier or safer prospects to induce higher actions: that is, when optimal signals are "single-dipped" or "single-peaked."[21] As we will see, this question unifies and generalizes much of what is known about special cases of the persuasion problem with nonlinear preferences, as well as other models that fit our optimal productive transport framework.[22]

A. Single-Dippedness/-Peakedness

A signal τ is "single-dipped" ("-peaked") if for any $\mu_1, \mu_2 \in \operatorname{supp}(\tau)$ such that $\operatorname{supp}(\mu_1)$ contains $x_1 < x_3$ and $\operatorname{supp}(\mu_2)$ contains $x_2 \in (x_1, x_3)$, we have $\gamma(\mu_1) \geq (\leq) \gamma(\mu_2)$. Similarly, τ is "strictly single-dipped" ("-peaked") if for any $\mu_1, \mu_2 \in \operatorname{supp}(\tau)$ such that $\operatorname{supp}(\mu_1)$ contains $x_1 < x_3$ and $\operatorname{supp}(\mu_2)$ contains $x_2 \in (x_1, x_3)$, we have $\gamma(\mu_1) > (<) \gamma(\mu_2)$. We also apply these definitions to an arbitrary set of posteriors $M \subset \Delta(X)$ by replacing $\operatorname{supp}(\tau)$ with M in the definitions. In particular, a pairwise signal is single-dipped if the induced receiver action is single-dipped on each set of nested pairs of states.

[21] Recall that we refer to "riskiness" in terms of the range of a posterior's support: if $x_1 < x_2 \leq x_3 < x_4$, then a posterior with support $\{x_1, x_4\}$ is riskier than one with support $\{x_2, x_3\}$.

[22] In the MOT context, 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 with respect to first-order stochastic dominance, while single-dippedness/-peakedness corresponds to monotonicity with respect to a variability order that depends on u; when $u(y, x) = x - y$, this variability order is the usual convex order.

An equivalent definition is that a signal τ is single-dipped if it never induces a strictly single-peaked triple (y_1, x_1), (y_2, x_2), (y_1, x_3), with $x_1 < x_2 < x_3$ and $y_1 < y_2$, in that there exist $\mu_1, \mu_2 \in \text{supp}(\tau)$ such that x_1, $x_3 \in \text{supp}(\mu_1)$ and $y_1 = \gamma(\mu_1)$, and $x_2 \in \text{supp}(\mu_2)$ and $y_2 = \gamma(\mu_2)$. (Otherwise, such a triple would witness a violation of single-dippedness.) Correspondingly, we say that a set $\Gamma \subset X \times Y$ is single-dipped if it does not contain a strictly single-peaked triple; and an outcome π is single-dipped if $\pi(\Gamma) = 1$ for some single-dipped set $\Gamma \subset X \times Y$.[23]

Each panel in figure 2 illustrates a signal in the linear receiver case ($u(y, x) = x - y$). Figure 2*A* is full disclosure, which is trivially strictly single-dipped, as no states are paired. Figure 2*B* is no disclosure, which is single-dipped but not strictly single-dipped. Panels C–E of figure 2 are all strictly single-dipped. Figure 2*C* is an example of negative assortative disclosure, where state $x = 1/3$ is disclosed and the other states are paired with weight 2/3 on the higher state in each pair. Figure 2*D* shows a signal 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, whereas disclose-pair splits up the pooling region in upper-censorship to obtain strict single-dippedness. Figure 2*E* shows a more complicated strictly single-dipped signal. While strict single-dippedness implies that each action is induced at two states at most, this panel shows that more than two actions can be induced at a single state (here, state 2/5).[24] Finally, figure 2*F* shows "matching across the median" (e.g., Kremer and Maskin 1996), which is not single-dipped, for example, because it contains the strictly single-peaked triple $\{(1/4, 1/2), (1/2, 3/4), (3/4, 1/2)\}$.

All of our results (and all proofs, except for the proof of theorem 4) are symmetric between the single-dipped and single-peaked cases. We thus present our results and proofs only for the single-dipped case (except for the proof of theorem 4), omitting the analogous results for the single-peaked case.

Remark 2. A strictly single-dipped set can be described by two functions χ_1 and χ_2 that specify the states $\chi_1(y)$ and $\chi_2(y)$, which are pooled together to induce each action y. Specifically, for any strictly single-dipped set Λ, there exist unique functions χ_1 and χ_2 from $Y_\Lambda = \{\gamma(\mu) : \mu \in \Lambda\}$ to X such that $\text{supp}(\mu) = \{\chi_1(\gamma(\mu)), \chi_2(\gamma(\mu))\}$ for all

[23] These definitions extend naturally to strict single-dippedness and (strict) single-peakedness.

[24] Also, while the function χ_2 defined in remark 2 is always monotone under strict single-dippedness, figure 2*E* shows that the function χ_1 can be nonmonotone.

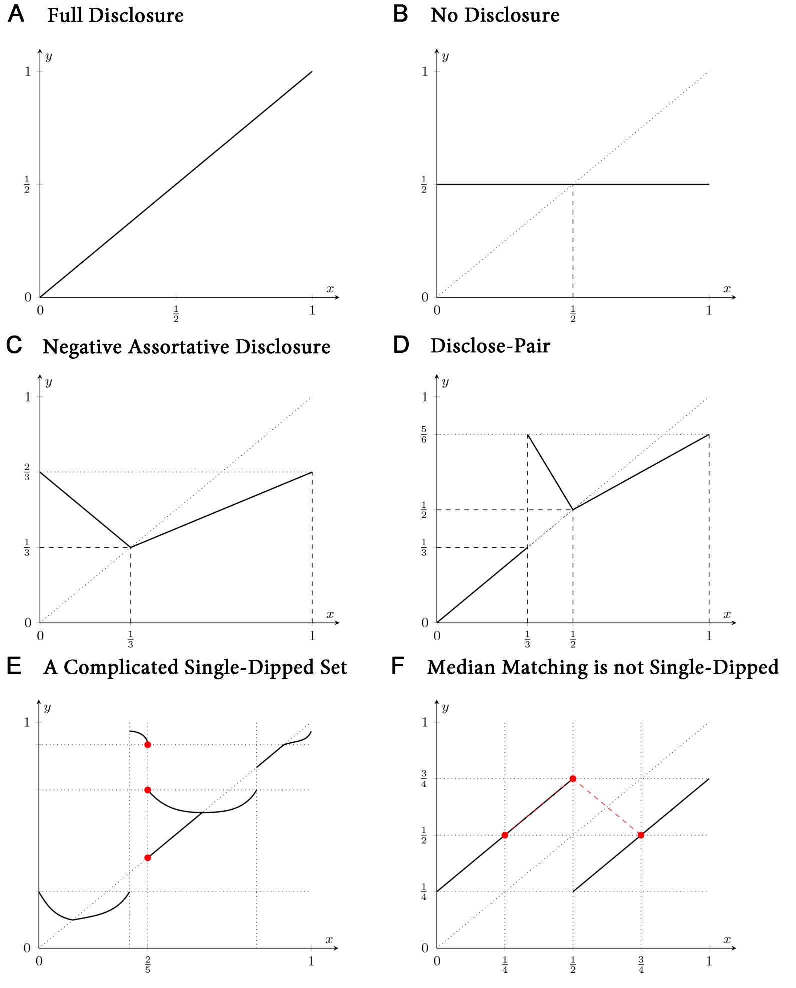

FIG. 2.—Some single-dipped disclosure patterns. Each panel displays, for the indicated signal τ (e.g., full disclosure in panel A) the set of points $(x, \gamma(\mu))$ where $x \in \operatorname{supp}(\mu)$ and $\mu \in \operatorname{supp}(\tau)$.

$\mu \in \Lambda$, $\chi_1(y) = \chi(y) = \chi_2(y)$ or $\chi_1(y) < \chi(y) < \chi_2(y)$ for all y, and $\chi_2(y) \leq \chi_2(y')$ and $\chi_1(y') \notin (\chi_1(y), \chi_2(y))$ for all $y < y'$.[25]

B. Variational Theorem

The next result captures the core economic logic behind single-dippedness. It is also our key tool for determining when optimal signals

[25] This remark follows from corollary 1.6 and lemma A.9 of Beiglböck and Juillet (2016). For completeness, we provide a simple self-contained proof.

are single-dipped: we use it to establish our main sufficient condition for single-dipped disclosure to be optimal (theorem 4, in sec. V.C), and use it directly to study some applications in appendix E.[26]

Theorem 3. Suppose that for any pair of actions $y_1 < y_2$ and any triple of states $x_1 < x_2 < x_3$ such that $x_1 < \chi(y_1) < x_3$, there exists a vector $\beta \geq 0$ such that $R\beta \geq 0$ and $R\beta \neq 0$, where

$$R := \begin{pmatrix} V(y_2, x_1) - V(y_1, x_1) & -(V(y_2, x_2) - V(y_1, x_2)) & V(y_2, x_3) - V(y_1, x_3) \\ -u(y_1, x_1) & u(y_1, x_2) & -u(y_1, x_3) \\ u(y_2, x_1) & -u(y_2, x_2) & u(y_2, x_3) \end{pmatrix}.$$

Then the contact set is single-dipped and hence so is every optimal signal.

The intuition behind theorem 3 is very simple and is illustrated in figure 3. The condition in theorem 3 says that a signal that induces a strictly single-peaked triple (y_1, x_1), (y_2, x_2), (y_1, x_3) with positive probability can be improved by reallocating mass β_1 on x_1 and mass β_3 on x_3 from y_1 to y_2, while reallocating mass β_2 on x_2 from y_2 to y_1. This reallocation is profitable for the sender, because the sender's expected utility increases when y_1 and y_2 are held fixed (i.e., the first coordinate of $R\beta$ is nonnegative); the receiver's marginal utility conditional on being recommended y_1 increases (i.e., the second coordinate of $R\beta$ is nonnegative), which increases the receiver's action and hence increases the sender's expected utility; and the receiver's marginal utility conditional on being recommended y_2 also increases (i.e., the third coordinate of $R\beta$ is nonnegative), which again increases the sender's expected utility. Moreover, at least one of these improvements is strict (i.e., $R\beta \neq 0$). The same logic also applies for any signal that induces a strictly single-peaked triple, even if this triple occurs with 0 probability, except now mass must be reallocated from small intervals around x_1, x_2, and x_3.[27]

C. *Sufficient Conditions for Single-Dipped Disclosure*

We can now give our main sufficient condition on utilities for single-dipped 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 4. If $u_{yx}(y, x)/u_x(y, x)$ and $V_{yx}(y_2, x)/u_x(y_1, x)$ are increasing in x for any y and $y_1 \leq y_2$, then there exists an optimal single-dipped signal.

[26] Theorem 3 provides conditions under which every optimal signal is single-dipped. In addition, lemma 9 in app. C provides weaker conditions under which some optimal signal has this property.

[27] Formally, this step relies on our complementary slackness theorem, theorem 1.

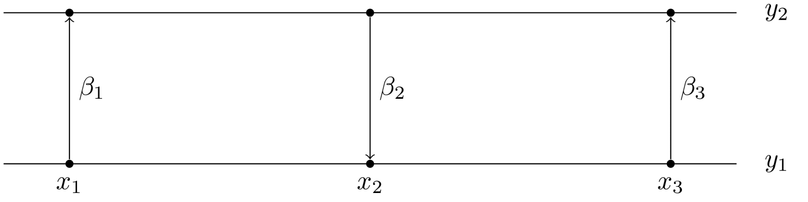

FIG. 3.—Profitable perturbation of a nonsingle-dipped signal. The figure shows a perturbation that shifts weights β_1 and β_3 on x_1 and x_3 from a posterior inducing action y_1 to a posterior inducing action y_2 and shifts weight β_2 on x_2 in the opposite direction. This perturbation is profitable if it increases the receiver's expected marginal utility at y_1 and y_2 and increases the sender's expected utility for fixed y_1 and y_2.

If in addition either $u_{yx}(y, x)/u_x(y, x)$ or $V_{yx}(y_2, x)/u_x(y_1, x)$ is strictly increasing in x for any y and $y_1 \leq y_2$, then the contact set is strictly single-dipped and hence so is every optimal signal.

The proof establishes single-dippedness by constructing perturbations that satisfy the conditions in theorem 3 and further establishes strict single-dippedness by verifying the twist condition from theorem 2.

The intuition for theorem 4 is relatively straightforward in the linear receiver and state-independent sender cases. (See fig. 4.) In the linear receiver case, we have $u_{yx}(y, x)/u_x(y, x) = 0$ and $V_{yx}(y_2, x)/u_x(y_1, x) = V_{yx}(y_2, x)$, so our sufficient conditions for single-dipped disclosure to

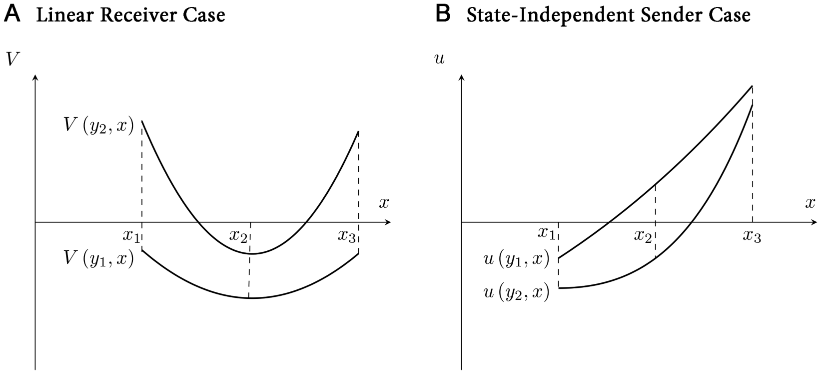

FIG. 4.—Intuition for theorem 4 in two special cases. *A*, In the linear receiver case, when the sender's utility increment $V(y_2, x) - V(y_1, x)$ is convex in the state, more extreme states should induce higher actions. *B*, In the state-independent sender case, when the receiver's marginal utility $u(y, x)$ is more convex in the state at higher actions, more extreme states should induce higher actions.

be optimal are satisfied if and only if V_y is convex in x.[28] To see why, note that for any strictly single-peaked triple (y_1, x_1), (y_2, x_2), (y_1, x_3), the perturbation that moves mass on x_1 and x_3 from y_1 to y_2 and moves mass on x_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 y_2 and concentrating the state conditional on action y_1. This perturbation is profitable when the difference $V(y_2, x) - V(y_1, x)$ is convex in x, which holds whenever V_y is convex in x.

In the state-independent sender case, we have $V_{yx}(y_2, x)/u_x(y_1, x) = 0$, so our sufficient conditions for single-dipped disclosure to be optimal are satisfied if and only if u_x is log-supermodular in (y, x), or equivalently u is more convex in x at higher actions y.[29] To see why, note that for any strictly single-peaked triple (y_1, x_1), (y_2, x_2), (y_1, x_3) , the perturbation that moves mass on x_1 and x_3 from y_1 to y_2 and moves mass on x_2 in the opposite direction, so as to hold fixed the receiver's marginal utility conditional on being recommended y_1 as well as the total probability of each recommendation, has the effect of increasing the receiver's marginal utility conditional on being recommended y_2. This follows because, by log-supermodularity of u_x, for the receiver's expected marginal utility the marginal rate of substitution between "shifting weight from x_1 to x_2" and "shifting weight from x_2 to x_3" is higher at y_1 than y_2. Finally, when V is state-independent, this perturbation increases the sender's expected utility.[30]

As we explain in section VII, there are close antecedents to the conditions in theorem 4 for the linear receiver and state-independent sender cases, in nonpersuasion settings that nonetheless fall in our general optimal productive transport framework. In particular, results in the MOT literature (e.g., theorem 6.1 of Beiglböck and Juillet 2016) can be translated to our framework to imply the linear receiver case of theorem 4, while results in the gerrymandering literature (lemma 1 of Friedman and Holden 2008) can be translated to our framework to imply the state-independent sender case of theorem 4. Theorem 4 thus unifies and generalizes these disparate contributions.

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

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

[30] In the linear receiver and state-independent sender cases, the sufficient conditions for the optimality of strictly single-dipped disclosure in theorem 4 are "almost necessary," because the condition $|S| \neq 0$ on $Y \times [0, 1]$ implies that $|S|$ has a constant sign on $Y \times [0, 1]$, which can be shown to be equivalent to strict convexity of V_y in the linear receiver case, and to strict log-supermodularity of u_x in the state-independent sender case. By theorem 2, a necessary condition for the optimality of strictly single-dipped disclosure is that $|S| \neq 0$ on the restricted domain where $x_1 < \chi(y) < x_3$.

We also establish an additional result in appendix B: under our conditions for strictly single-dipped disclosure to be optimal (and a regularity condition), the optimal signal is unique.[31]

VI. Full Disclosure and Negative Assortative Disclosure

While single-dippedness is an important property, it remains important to fully characterize optimal signals when this is tractable.[32] The current section does this for the polar cases of "maximum" and "minimal" pairwise disclosure. The former case corresponds to full disclosure, where each state is disclosed; while the latter case corresponds to negative assortative disclosure, where all states are paired in a negatively assortative manner, so all posteriors 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.[33]

A. *Full Disclosure*

Full disclosure is the (unique) signal τ where every $\mu \in \text{supp}(\tau)$ is degenerate.

If for all states x_1 and x_2, and all probabilities ρ, the sender prefers to split the posterior $\mu = \rho\delta_{x_1} + (1 - \rho)\delta_{x_2}$ into degenerate posteriors δ_{x_1} and δ_{x_2}, then the sender prefers full disclosure to any pairwise signal. Since pairwise signals are without loss by theorem 2, full disclosure is then optimal. Conversely, if the sender strictly prefers not to split $\mu = \rho\delta_{x_1} + (1 - \rho)\delta_{x_2}$ into δ_{x_1} and δ_{x_2} for some states x_1 and x_2 and some probability ρ, then the sender strictly prefers the pairwise signal that differs from full disclosure only in that it pools states x_1 and x_2 into μ; so full disclosure is not optimal.[34] Recalling that belief $\mu = \rho\delta_{x_1} + (1 - \rho)\delta_{x_2}$ induces action $\gamma(\mu)$ satisfying $\rho u(\gamma(\mu), x_1) + (1 - \rho) u(\gamma(\mu), x_2) = 0$, we obtain the following result.

[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 sec. 1.3 in Santambrogio 2015).

[32] Recall that many different disclosure patterns can be single-dipped, as illustrated in fig. 2.

[33] In the context of partisan gerrymandering, Kolotilin and Wolitzky (2024a) provide conditions under which the disclose-pair pattern illustrated in fig. 2*D* is optimal in the state-independent sender case.

[34] This argument is valid when ϕ has finite support. The general case (theorem 5) uses duality and is adapted from part 2 of proposition 1 in Kolotilin (2018); we give a slightly simpler proof and also establish uniqueness.

Theorem 5. Full disclosure is optimal if and only if, for all $\mu = \rho\delta_{x_1} + (1-\rho)\delta_{x_2}$ with $x_1 < x_2$ in X and $\rho \in (0,1)$, we have

$$\rho V(\gamma(\mu), x_1) + (1-\rho)V(\gamma(\mu), x_2) \leq \rho V(\gamma(\delta_{x_1}), x_1) + (1-\rho)V(\gamma(\delta_{x_2}), x_2). \tag{7}$$

Moreover, full disclosure is uniquely optimal if (7) holds with strict inequality for all such μ.

In the linear case, condition (7) holds if and only if V is convex in y. In the state-independent sender case, condition (7) simplifies as follows:

Corollary 2. In the state-independent sender case, full disclosure is optimal if and only if, for all $\mu = \rho\delta_{x_1} + (1-\rho)\delta_{x_2}$ with $x_1, x_2 \in X$ and $\rho \in (0,1)$, we have

$$V(\gamma(\mu)) \leq \rho V(\gamma(\delta_{x_1})) + (1-\rho)V(\gamma(\delta_{x_2})). \tag{8}$$

In a classical one-to-one matching model, Becker (1973) showed that if the utility from matching two types $h(x_1, x_2)$ 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 if and only if $h(x_1, x_1) + h(x_2, x_2) \geq 2h(x_1, x_2)$ for all x_1, x_2 (which is a strictly weaker property than supermodularity). In the persuasion setting, segregation corresponds to full disclosure. Note that if we fix $\rho = 1/2$ and let $h(x_1, x_2) = V(\gamma(\delta_{x_1}/2 + \delta_{x_2}/2))$, then (8) 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 (7).

Corollary 2'. In the linear receiver case, full disclosure is optimal if $V(y, x)$ is convex in y and satisfies $V(x_1, x_2) + V(x_2, x_1) \leq V(x_1, x_1) + V(x_2, x_2)$ for all $x_1, x_2 \in X$.

A sufficient condition for $V(x_1, x_2) + V(x_2, x_1) \leq V(x_1, x_1) + V(x_2, x_2)$ is supermodularity of V: for all $x_1 < x_2$ and $y_1 < y_2$, $V(y_1, x_1) + V(y_2, x_2) \geq V(y_1, x_2) + V(y_2, x_1)$. Thus, in the linear receiver case, full disclosure is optimal whenever the sender's utility is convex in y and supermodular in (y, x). This sufficient condition for full disclosure generalizes that given by Rayo and Segal (2010) for the separable subcase.[35]

In addition, in appendix B we show that when the prior has full support and the twist condition holds, full disclosure is uniquely optimal whenever it is optimal.

[35] Their condition is that w is increasing in x and G is convex in y, where $V(y, x) = w(x)G(y)$. In the sub-subcase with $G(y) = y$, (7) holds if and only if w is increasing in x, because (7) simplifies to $\rho(1-\rho)(w(x_2) - w(x_1))(x_2 - x_1) \geq 0$.

B. *Negative Assortative Disclosure*

A pairwise signal τ is "negative assortative" if the supports of any μ, $\mu' \in \text{supp}(\tau)$ are nested: that is, denoting $\text{supp}(\mu) = \{x_1, x_2\}$ and $\text{supp}(\mu') = \{x'_1, x'_2\}$, where without loss $x_1 \leq x_2$ and $x'_1 \leq x'_2$, we have either $x_1 \leq x'_1 \leq x'_2 \leq x_2$ or $x'_1 \leq x_1 \leq x_2 \leq x'_2$. We also apply the same definition to an arbitrary set of posteriors $M \subset \Delta(X)$ by replacing $\text{supp}(\tau)$ with M. In particular, a strictly single-dipped contact set Λ is negative assortative if $\chi_1 : Y_\Lambda \to X$ is decreasing and $\chi_2 : Y_\Lambda \to X$ is increasing.[36]

The main result of this section is that if strictly single-dipped disclosure is optimal and the sender strictly prefers to pool any two states, then negative assortative disclosure is optimal. Moreover, if the prior has a density, then the optimal signal is unique (by theorem 7 in app. B) 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 $\{x_1, x_3\}$ and $\{x'_1, x'_3\}$ with $x_1 < x_3$, $x'_1 < x'_3$, and $x_1 \leq x'_1$, must be either ordered (i.e., $x_1 < x_3 \leq x'_1 < x'_3$) or nested (i.e., $x_1 \leq x'_1 < x'_3 \leq x_3$). This follows because if the pairs overlap (i.e., $x_1 < x'_1 < x_3 < x'_3$), then either (x_1, x'_1, x_3) or (x'_1, x_3, x'_3), together with the corresponding actions, would form a single-peaked triple. Hence, for any pair of pooled states (x_1, x_3), there must exist a disclosed state $x_2 \in (x_1, x_3)$: intuitively, there must exist pairs of pooled states in the interval (x_1, x_3) that are closer and closer together, until the pair degenerates into a single disclosed state. Therefore, if any two pairs of pooled states $\{x_1, x_3\}$ and $\{x'_1, x'_3\}$ are ordered, there would exist two distinct disclosed states $x_2 \in (x_1, x_3)$ and $x'_2 \in (x'_1, x'_3)$. 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.[37]

To derive the equations for the optimal signal, note that if χ_1 and χ_2 are differentiable then the posterior μ that induces $y = \gamma(\mu)$ equals[38]

$$\mu = \frac{-f(\chi_1(y))\chi'_1(y)}{-f(\chi_1(y))\chi'_1(y) + f(\chi_2(y))\chi'_2(y)}\delta_{\chi_1(y)} + \frac{f(\chi_2(y))\chi'_2(y)}{-f(\chi_1(y))\chi'_1(y) + f(\chi_2(y))\chi'_2(y)}\delta_{\chi_2(y)}.$$

[36] Recall that, as defined in remark 2, $\chi_1(y)$ and $\chi_2(y)$ are the smaller and larger states that are pooled together to induce action $y \in Y_\Lambda$.

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

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

Hence, by (1) and (BP), we have

$$u(y,\chi_1(y))f(\chi_1(y))\chi_1'(y) = u(y,\chi_2(y))f(\chi_2(y))\chi_2'(y),$$

or more generally (allowing that χ_1 and χ_2 may not be differentiable)

$$u(y,\chi_1(y))(-\mathrm{d}\phi([0,\chi_1(y)])) + u(y,\chi_2(y))\mathrm{d}\phi([0,\chi_2(y)]) = 0. \quad (9)$$

In addition, the sender's first-order condition (4) gives

$$v(y,\chi_1(y)) + q(y)u_y(y,\chi_1(y)) + q'(y)u(y,\chi_1(y)) = 0,$$
$$v(y,\chi_2(y)) + q(y)u_y(y,\chi_2(y)) + q'(y)u(y,\chi_2(y)) = 0,$$

and solving this system of equations gives

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{V_y(y,\chi_1(y))u(y,\chi_2(y)) - V_y(y,\chi_2(y))u(y,\chi_1(y))}{u(y,\chi_1(y))u_y(y,\chi_2(y)) - u(y,\chi_2(y))u_y(y,\chi_1(y))}\right)\\ &= \frac{V_y(y,\chi_1(y))u_y(y,\chi_2(y)) - V_y(y,\chi_2(y))u_y(y,\chi_1(y))}{u_y(y,\chi_1(y))u(y,\chi_2(y)) - u_y(y,\chi_2(y))u(y,\chi_1(y))}.\end{aligned} \quad (10)$$

Finally, the solution to the differential equations (9) and (10) must satisfy the boundary conditions

$$(\chi_1(\bar{y}),\chi_1(\underline{y}),\chi_2(\underline{y}),\chi_2(\bar{y})) = (0,\chi(\underline{y}),\chi(\underline{y}),1), \quad (11)$$

where $\underline{y} = \min Y_\Lambda$ and $\bar{y} = \max Y_\Lambda$, because the lowest induced action $\underline{y}$ is induced at the disclosed state $\chi(\underline{y}) = \chi_1(\underline{y}) = \chi_2(\underline{y})$, and the highest induced action $\bar{y}$ is induced at states $0 = \chi_1(\bar{y})$ and $1 = \chi_2(\bar{y})$.[39]

Theorem 6. Assume that $X = [0,1]$. If Λ is strictly single-dipped and for all $x_1 < x_2$ there exists $\rho \in (0,1)$ such that

$$\begin{aligned}&\rho V(\gamma(\mu),x_1) + (1-\rho)V(\gamma(\mu),x_2)\\ &> \rho V(\gamma(\delta_{x_1}),x_1) + (1-\rho)V(\gamma(\delta_{x_2}),x_2),\end{aligned} \quad (12)$$

with $\mu = \rho\delta_{x_1} + (1-\rho)\delta_{x_2}$, then Λ is negative assortative. Moreover, if the prior ϕ has a density f, then the optimal signal is unique, and the

[39] In the linear receiver case, (10) simplifies to

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}y}\left(V_y(y,\chi_1(y))\frac{\chi_2(y)-y}{\chi_2(y)-\chi_1(y)} + V_y(y,\chi_2(y))\frac{y-\chi_1(y)}{\chi_2(y)-\chi_1(y)}\right)\\ &= -\frac{V_y(y,\chi_2(y)) - V_y(y,\chi_1(y))}{\chi_2(y)-\chi_1(y)}.\end{aligned}$$

Geometrically, this says that the slope of the curve $\gamma(\mu) \mapsto \mathbb{E}_\mu[V_y(\gamma(\mu),x)]$ is equal to the negative of the slope of the secant passing through the points $(\chi_1(\gamma(\mu)), V_y(\gamma(\mu),\chi_1(\gamma(\mu))))$ and $(\chi_2(\gamma(\mu)), V_y(\gamma(\mu),\chi_2(\gamma(\mu))))$. Nikandrova and Pancs (2017) derive this condition for the separable sub-subcase with $V_y(y,x) = w(x)$.

functions χ_1 and χ_2 are continuous and solve the system of differential equations (9) and (10), with boundary conditions (11).

Like equation (7) in section VI.A, equation (12) simplifies in special cases. In the linear case, (12) holds if and only if V is strictly concave in y.[40] In the state-independent sender case, it holds if and only if $V(\gamma(\mu)) > \rho V(\gamma(\delta_{x_1})) + (1-\rho)V(\gamma(\delta_{x_2}))$. In the linear receiver case, it holds if $V(y, x)$ is concave in y and satisfies $V(x_1, x_2) + V(x_2, x_1) > V(x_1, x_1) + V(x_2, x_2)$ for all $x_1 < x_2$; 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.[41]

We can also give primitive conditions on V and u for (12) to hold and hence for the unique optimal signal to be negative assortative.

Corollary 3. Assume that the condition for strict single-dippedness given in theorem 4 holds. Then for all $x_1 < x_2$, there exists $\rho \in (0,1)$ such that (12) holds if and only if for all $y \in Y$,

$$V_{yy}(y,\chi(y)) \le \frac{V_y(y,\chi(y))\,u_{yy}(y,\chi(y))}{u_y(y,\chi(y))} + 2\,\frac{V_{yx}(y,\chi(y))\,u_y(y,\chi(y)) - V_y(y,\chi(y))\,u_{yx}(y,\chi(y))}{u_x(y,\chi(y))}. \tag{13}$$

Equation (13) is a local necessary condition for (12): if (13) fails, then (12) also fails for $x_1 < x_2$ sufficiently close to $\chi(y)$. When the condition for strict single-dippedness holds, this local necessary condition turns out to be globally sufficient for (12). Equation (13) simplifies dramatically in some special cases. In the linear receiver case, (13) simplifies to $V_{yy}(y,y) + 2V_{yx}(y,y) \le 0$; in the translation-invariant subcase of the linear receiver case, this simplifies further to $P''(0) \ge 0$. In the separable (respectively, translation-invariant) subcase of the state-independent sender case, (13) simplifies to $V_{yy}(y)/V_y(y) \le 2I'(y)/I(y)$ (respectively, $V_{yy}(y)/V_y(y) \le T''(0)/T'(0)$).

In appendix D, we give three examples of optimal single-dipped negative assortative disclosure. Example 1 illustrates how the differential equations (9) and (10) can sometimes be explicitly solved to find the optimal signal. Example 2 characterizes the optimal signal in the quantile sub-subcase. In the quantile sub-subcase, our sufficient conditions for strictly single-dipped disclosure to be optimal are not satisfied, and there are multiple optimal signals; however, one optimal signal is strictly

[40] In the linear case, V is strictly concave if and only if no disclosure is uniquely optimal for all priors, by corollary 1 in Kolotilin, Mylovanov, and Zapechelnyuk (2022).

[41] Their condition is that w is strictly decreasing in x and G is concave in y, where $V(y, x) = w(x)G(y)$. In the sub-subcase with $G(y) = y$, (12) holds if and only if w is strictly decreasing in x.

single-dipped negative assortative. Finally, example 3 illustrates that in some cases the unique optimal signal randomizes conditional on the state, even when the prior is atomless.[42]

VII. Optimal Productive Transport: Theory and Applications

We now give a general interpretation of our framework in terms of assigning inputs to productive units and describe the implications of our results for matching, option pricing, and partisan gerrymandering, as well as some specific persuasion models. We pay particular attention to the implications of single-dippedness summarized in table 1 in the introduction.

A. *Optimal Productive Transport*

Our signal-based primal problem (P) may be generalized as follows: given a distribution $\phi \in \Delta([0,1])$ with support $\mathrm{supp}(\phi) = X$ of inputs x, find a distribution $\tau \in \Delta(\Delta(X))$ of productive units $\mu \in \Delta(X)$ to

$$\begin{aligned} &\text{maximize} \quad \int_{\Delta(X)} \int_X V(\tilde{\gamma}(\mu), x)\mathrm{d}\mu(x)\mathrm{d}\tau(\mu) \\ &\text{subject to} \quad \int_{\Delta(X)} \mu \mathrm{d}\tau(\mu) = \phi, \end{aligned}$$

where $\tilde{\gamma} : \Delta(X) \to Y$ is an arbitrary production function that specifies the output y produced by unit μ.[43] This optimal productive transport problem is the same as (P), except that the value of an arbitrary production function $\tilde{\gamma}(\mu)$ is not necessarily given by the first-order condition (1) for some function u.

But now suppose that $\tilde{\gamma}$ satisfies the following two properties:

1. *Betweenness.*—For any $\mu, \eta \in \Delta(X)$ satisfying $\tilde{\gamma}(\mu) < \tilde{\gamma}(\eta)$, and any $\rho \in (0,1)$, we have $\tilde{\gamma}(\mu) < \tilde{\gamma}(\rho\eta + (1-\rho)\mu) < \tilde{\gamma}(\eta)$.
2. *Continuity.*—$\tilde{\gamma}$ is continuous on $\Delta(X)$.

The key property here is the first one, which says that mixing two units produces an output in between those produced by each of them in isolation. The following result—which adapts proposition A.1 of Dekel (1986)—shows that these properties ensure the existence of a function u such

[42] In contrast, Zeng (2023) shows that there is always a deterministic optimal signal in the separable subcase of the linear receiver or state-independent sender case.

[43] The interpretation of the assumption that the domain of $\tilde{\gamma}$ is probability measures on X rather than arbitrary measures is that production exhibits constant returns to scale, so nothing is gained by varying the scale of a productive unit.

that $\tilde{\gamma}$ is given by (1). Thus, under betweenness and continuity, the optimal productive transport problem is the same as our persuasion problem.

PROPOSITION 1. A function $\tilde{\gamma}$ satisfies betweenness and continuity if and only if there exists a continuous function $u: Y \times X \to \mathbb{R}$ such that, for any $\mu \in \Delta(X)$,

$$\int_X u(y, x)\mathrm{d}\mu(x) = (<)0 \Longleftrightarrow y = (>)\tilde{\gamma}(\mu).$$

In the persuasion context, the function u in proposition 1 is the receiver's marginal utility. More generally, $u(y, x)$ can be viewed as a measure of the "efficacy" of input x in the production of output y. For the remainder of this section, we assume betweenness and continuity, as well as that some such u satisfies assumptions 1–4.

The general interpretation of the signal-based primal problem (P) is that a planner obtains utility $V(y, x)$ from using input x in the production of output y and assigns inputs to productive units according to a production plan $\tau \in \Delta(\Delta(X))$ to maximize the expectation of $V(\gamma(\mu), x)$. The corresponding interpretation of the signal-based dual problem (D) is that there is a decentralized economy with price $p(x)$ attached to input x, where the zero-profit condition (ZP) says that an entrepreneur who obtains utility $V(y, x)$ from using input x in the production of output y cannot create a unit $\mu \in \Delta(X)$ that leaves her with a positive utility after paying for its inputs.

More precisely, a pair $(\tau, p) \in \Delta(\Delta(X)) \times L(X)$ is a competitive equilibrium if

(i) all inputs are assigned to productive units: $\int_X \mu \mathrm{d}\tau(\mu) = \phi$;
(ii) operating units make zero profits: $\int_X p(x)\mathrm{d}\mu(x) = W(\mu)$ for all $\mu \in \mathrm{supp}(\tau)$;
(iii) no entrant can make strictly positive profits: $\int_X p(x)\mathrm{d}\mu(x) \geq W(\mu)$ for all $\mu \in \Delta(X)$.

Then, by strong duality (lemma 1), we have

REMARK 3. A competitive equilibrium exists, and a pair (τ, p) is a competitive equilibrium if and only if τ solves (P) and p solves (D).

We will make use of the interpretation of optimal plans and prices as competitive equilibria in the matching context in section VII.B.

Similarly, the general interpretation of the outcome-based primal problem (P') is that a planner chooses an outcome $\pi \in \Delta(Y \times X)$ to maximize her expected utility, subject to the constraints that all inputs are utilized and that each output y is produced by a unit μ satisfying $\gamma(\mu) = y$. The corresponding dual can again be interpreted as a decentralized economy, where now the zero-profit condition (ZP') says that an entrepreneur who produces any output y cannot profitably employ any input in the production of this output, after accounting for the input's price, $p(x)$, and the marginal effect of its use on the output, $q(y)u(y, x)$. Finally, complementary slackness (theorem 1) says that any entrepreneur who breaks even must employ only inputs that satisfy (ZP') with equality.

B. Matching with Peer Effects

Assigning workers with heterogeneous abilities to firms with workplace peer effects (i.e., intrafirm spillovers) is an important topic in labor economics (Kremer and Maskin 1996; Saint-Paul 2001; Eeckhout 2018; Boerma, Tsyvinski, and Zimin 2025). In Saint-Paul (2001), there is a continuum of workers indexed by ability $x \in [0, 1]$. The population distribution of x is ϕ, with support $X \subset [0, 1]$. Workers sort into firms, which are ex ante homogeneous and face constant returns to scale, so that a firm can be identified with its distribution of workers $\mu \in \Delta(X)$. Workplace peer effects depend on the mean worker ability in a firm, $\gamma(\mu) = \mathbb{E}_\mu[x]$, so the output of a worker with ability x in firm μ can be written as $V(\gamma(\mu), x)$. The planner's problem of assigning workers to firms to maximize total output is thus precisely (P), in the linear receiver case where $\gamma(\mu) = \mathbb{E}_\mu[x]$. Moreover, the problem of finding competitive equilibrium wages $p(x)$ for workers with ability x (where, as in sec. VII.A, a competitive equilibrium is an assignment of workers to firms and wages such that operating firms make zero profits and no entrant can make strictly positive profits) is precisely (D). In this context, lemma 1 says that a competitive equilibrium exists and maximizes total output.

Saint-Paul (2001) considered the special case of this model where $V(y, x) = xG(y)$ for an increasing function G. In this case, the total output of a firm μ equals $\mathbb{E}_\mu[x]G(\mathbb{E}_\mu[x])$. Since this is a function only of $\mathbb{E}_\mu[x]$, Saint-Paul (2001) coincides with the linear case of our model. We now summarize the implications of our results for the general linear receiver case of our model in the worker assignment context. We pay special attention to the separable case where $V(y, x) = w(x)G(y)$ for increasing functions w and G, which may be particularly natural in this context.

In the worker assignment context, an assignment is pairwise if each firm contains at most two worker types, and a pairwise assignment is strictly single-dipped if firms with more heterogeneous workers have higher average worker ability. Since we are in the linear receiver case, theorem 4 implies that a strictly single-dipped assignment is optimal whenever V_y is strictly convex in x—or, in the separable case, w is strictly convex. Intuitively, V_y is the marginal benefit of having higher-ability coworkers, so when this is convex in a worker's own ability, it is optimal for firms where the distribution of worker abilities is more spread out to have higher average worker ability. Moreover, when in addition V is strictly increasing in both arguments and convex in x, firms with more heterogeneous workers also produce higher output, as if $\mathbb{E}_\mu[x] = y < y' = \mathbb{E}_{\mu'}[x]$ and $\mathrm{supp}(\mu')$ is more spread out than $\mathrm{supp}(\mu)$, then[44]

$$\mathbb{E}_\mu[V(y, x)] < \mathbb{E}_\mu[V(y', x)] < \mathbb{E}_{\mu'}[V(y', x)].$$

Thus, in this case firms with more heterogeneous workforces are more productive and hence also pay higher average wages (by zero profit).

Our conditions for full disclosure (i.e., segregation, where each firm has a homogeneous workforce) and negative assortative disclosure/matching (where all firms can be ordered from least to most heterogeneous) are also interesting in the worker assignment context. By corollary 2', segregation is optimal if V is convex in y and supermodular—or, in the separable case, G is convex. On the other hand, by corollary 3, negative assortative matching is optimal if V_y is strictly convex in x and $V_{yy} + 2V_{yx} \le 0$—or, in the separable case, w is strictly convex and $w(x)G''(y) + 2w'(x)G'(y) \le 0$. Intuitively, these results say that segregation is optimal if peer effects are convex and that negative assortative matching is optimal if single-dipped assignment is optimal and peer effects are sufficiently concave.

We also note that $p(x)$ is convex whenever V is convex in x—or, in the separable case, w is convex. This follows because $p(x) = \sup_{y \in Y} V(y, x) + q(y)(x - y)$, so if $V_{xx} > 0$ then p is the supremum of a set of convex functions. Recalling that $p(x)$ is the equilibrium wage of a worker with ability x, this says that wages rise more than one-for-one with ability. This result reflects the fact that higher-ability workers are not only better workers, but also better coworkers.

A model that is equivalent to worker-firm matching can capture the assignment of students to schools with peer effects, or more generally the assignment of heterogeneous agents to clubs. In Arnott and Rowse (1987), there is a continuum of students indexed by ability x, who must be assigned to ex ante identical schools, which can be identified with their student bodies $\mu \in \Delta(X)$. A student with ability x who attends a school μ attains an education that she values at $V(\gamma(\mu), x)$, where again $\gamma(\mu) = \mathbb{E}_\mu[x]$. Arnott and Rowse (1987) study the planner's problem of assigning students to schools to maximize total educational attainment: this problem is equivalent to the linear receiver case of (P). The "decentralized" version of this problem is considered by Epple and Romano (1998), who study competitive equilibrium in a market for private schooling, where a school μ with mean student ability $y = \gamma(\mu)$ charges tuition $t(y, x)$ to students of ability x. Here, a competitive equilibrium may be

[44] Here, the first inequality follows because $V_y > 0$, and the second inequality follows because V_x, $V_{xx} > 0$ and μ' can be obtained from μ by increasing its mean and then taking a mean-preserving spread (i.e., μ' is greater than μ in the increasing convex order).

defined precisely as in section VII.A, with the interpretation that $p(x)$ is the equilibrium utility of a student with ability x and that the tuition charged to a student with ability x to attend a school with mean student ability y is $t(y, x) = V(y, x) - p(x)$. (The assumption that schools take student utility levels as given when setting tuition is called "utility taking" in the literature on club economies (e.g., Ellickson et al. 1999.) In this context, lemma 1 says that a competitive equilibrium exists and maximizes total educational attainment.[45]

The conditions on V under which an optimal assignment of students to schools is single-dipped, segregation, or negative assortative matching are the same as in the worker assignment context. Indeed, bearing in mind that $p(x)$ is the equilibrium utility of an agent with ability x in either model—so that the wage of a worker with ability x is $p(x)$, and the tuition paid by a student with ability x who attends a school with mean ability y is $V(y, x) - p(x)$—the models are identical. In particular, if V_y is strictly convex in x, then a strictly single-dipped assignment is optimal, so that schools with more heterogeneous student bodies are more desirable for all students.

The outcome-based dual (D') has a particularly natural interpretation in the student assignment/club economy setting. In a competitive equilibrium, a school with mean student ability y charges tuition $t(y, x) = q(y)(y - x)$ to students with ability x. (Thus, a school subsides students with above-average ability and charges students with below-average ability.) A student with ability x attends the school y that gives her the highest utility, $p(x) = \sup_{y \in Y} V(y, x) - q(y)(y - x)$. All operating schools break even, and no entrepreneur can turn a positive profit by starting a new school.

Yet another interpretation of the model covers peer effects in residential choice: de Bartolome (1990), Benabou (1996), Becker and Murphy (2000), and Durlauf (2004) consider residents with binary ability x who choose to live in one of two neighborhoods of fixed size. A resident with ability x who lives in a neighborhood with a fraction y of high-ability residents obtains utility $V(y, x)$. These papers study the planner's problem of assigning residents to neighborhoods to maximize total utility and contrast the solution with the competitive equilibrium outcome when real estate prices cannot depend on the purchaser's ability. Our model extends this setting to the case of a continuum of ability levels and (potentially) a continuum of neighborhoods, where a resident with ability x who lives in a neighborhood with a distribution of residents $\mu \in \Delta(X)$ obtains

[45] Arnott and Rowse (1987) additionally endogenize public spending on schools, while Epple and Romano (1998) let students differ in income as well as ability. Arnott and Rowse (1987) focus on the Cobb–Douglas utility function $V(y, x) = x^{\alpha} y^{\beta}$ and provide conditions for the optimality of segregation ("perfect streaming") or no disclosure ("complete mixing").

utility $V(\gamma(\mu), x)$, with $\gamma(\mu) = \mathbb{E}_\mu[x]$. Our condition for segregation to be optimal—that V is convex in y and supermodular—sharpens results in this literature.[46] Our condition for strict single-dippedness—that V_y is strictly convex in x—has no analogue in the literature (which assumes binary types). Finally, remark 3 shows that a competitive equilibrium is efficient if real estate prices can depend on the purchaser's ability. In such an equilibrium, the price for a resident with ability x to buy a house in a neighborhood with mean ability y is $V(y, x) - p(x) = q(y)(y - x)$.[47]

C. *Option Pricing*

In mathematical finance, the literature on martingale optimal transport (e.g., Beiglböck, Henry-Labordère, and Penkner 2013; Galichon, Henry-Labordère, and Touzi 2014; Beiglböck and Juillet 2016) studies the following problem. An underlying asset will be marketed in two future periods, 1 and 2. In period 0, an exotic option is for sale, which will pay $V(y, x)$ if the realized asset price is y in period 1 and x in period 2. An analyst knows the marginal distributions of y and x, but her only information regarding their joint distribution is that it satisfies $\mathbb{E}[x|y] = y$ for every y. The interpretation of this assumption is that there are liquid markets for European call options on the asset price in each period, from which the analyst can infer the marginal distributions of the asset price (by Breeden and Litzenberger 1978); and the analyst believes that the market satisfies no-arbitrage, which implies that the asset price is a martingale under the risk-neutral measure. The analyst's problem is to find the joint distribution $\pi \in \Delta(Y \times X)$ that maximizes the expected value of the option (and, thus, the maximum option price consistent with no-arbitrage), subject to the two marginal constraints and the martingale constraint.

Now consider the variant of this problem where the marginal distribution of y is also unknown. The interpretation is that there is a liquid market for call options only on the period 2 asset price: for example, perhaps the asset is a share in a firm that is expected to go public after period 1, and there are only liquid options markets for the prices of publicly traded firms. Then the analyst's problem of determining the maximum option price, subject to constraint that the marginal distribution of the period 2 price x is ϕ, and the martingale constraint $\mathbb{E}[x|y] = y$, is precisely (P'), in

[46] The closest point in the literature seems to be an observation by Benabou (1996, 249) that $V_{yy} > 0$ and $V_{yx} > 0$ both favor segregation.

[47] The above models all feature linear peer effects: $\gamma(\mu) = \mathbb{E}_\mu[x]$. Boucher et al. (2024) consider a model of nonlinear peer effects where $\gamma(\mu) = h^{-1}(\mathbb{E}_\mu[h(x)])$ and h is a power function. In our setting, this model—along with the more general one where $u(y, x) = h(x) - h(y)$ for any strictly increasing function h, which yields peer effect $\gamma(\mu) = h^{-1}(\mathbb{E}_\mu[h(x)])$—is equivalent to the linear peer effect (linear receiver) case, up to the change of variables $\tilde{x} = h(x)$ and $\tilde{y} = h(y)$.

the linear receiver case where $\gamma(\mu) = \mathbb{E}_\mu[x]$. Lemma 1 establishes strong duality for this problem.[48]

In this context, the optimal dual variables (p, q) have an important interpretation. Recall that the option to be priced pays $V(y, x)$ when the asset price is y in period 1 and x in period 2. An alternative to buying this exotic option is to buy a simple option that pays $p(x)$ when the period 2 asset price is x, and in addition to plan to sell $q(y)$ units of the asset itself in period 1 when the period 1 asset price is y. Since selling $q(y)$ units at price y yields a profit of $q(y)(y - x)$ when the period 2 asset price turns out to be x, this alternative strategy is sure to outperform—or "super-replicate"—the exotic option if and only if

$$p(x) + q(y)(y - x) \geq V(y, x) \quad \text{for all } (y, x) \in Y \times X.$$

Note that this condition is precisely (ZP'). Thus, lemma 1 implies that the maximum option price can be calculated as either $\mathbb{E}_\pi[V(y, x)]$ under the joint distribution of asset prices π that solves (P') (i.e., the maximum expected value of the exotic option), or as $\mathbb{E}_\phi[p(x)]$, for the simple option payouts $p(x)$ that solve (D') (i.e., the price of the cheapest strategy that super-replicates the exotic option).

In the option pricing context, a joint distribution of asset prices is pairwise if it is a binomial tree: each period 1 price y can be followed by at most two distinct period 2 prices x. A pairwise joint distribution is strictly single-dipped if more dispersed period 2 prices follow higher period 1 prices: that is, if riskier assets are more expensive. Since we are in the linear receiver case, theorem 4 implies that the option price is maximized by a strictly single-dipped distribution whenever V_y is strictly convex in x. This condition is known as the martingale Spence-Mirrlees condition in the MOT literature, which Beiglböck and Juillet (2016), Henry-Labordère and Touzi (2016), and Beiglböck, Henry-Labordère, and Touzi (2017) show implies that a strictly single-dipped distribution (which they call a "left-curtain coupling") is optimal in the standard MOT problem (where the period 1 asset price distribution is fixed exogenously).[49] Moreover, by corollary 2', full disclosure (where $x = y$ with probability 1) is optimal if V is strictly convex in y and supermodular; while corollary 3 implies that

[48] The possibility that the period 1 marginal may be unknown and the resulting problem (P') are briefly considered in corollary 1.5 of Acciaio et al. (2016). That result establishes weak duality and primal attainment but not dual attainment, which, as we discuss, is an important issue.

[49] Specifically, 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); while 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_y is strictly convex in x. All these papers concern the MOT context, where the distribution of y is fixed exogenously.

negative assortative matching (where higher period 1 prices are always followed by more dispersed period 2 prices, so more expensive assets are riskier) is optimal if V_y is strictly convex in x and $V_{yy} + 2V_{yx} \leq 0$.

The formula for $q(y)$ also has an interesting interpretation in the option pricing context. By equation (3), for every period 1 price y in the support of the marginal of an optimal joint distribution π, we have

$$q(y) = \mathbb{E}_{\pi}[V_y(y, x)|y].$$

This is a version of Shephard's lemma: the amount of the asset sold at period 1 price y under the cheapest super-replicating strategy equals the derivative of the option price with respect to y. In addition, in finance the derivative of the option price with respect to the underlying asset price is known as the option's "Delta." Thus, in the option pricing context, $q(y)$ is simply Delta.

D. Partisan Gerrymandering

Partisan gerrymandering—where a partisan designer assigns voters to districts to maximize her party's seat share—is an important feature of American politics. Kolotilin and Wolitzky (2024a) develop and calibrate a model of partisan gerrymandering, which generalizes the leading earlier models of Owen and Grofman (1988), Friedman and Holden (2008), and Gul and Pesendorfer (2010). In this model, there is a continuum of voters indexed by their partisanship $x \in [0, 1]$. The population distribution of x is ϕ, with support $X \subset [0, 1]$. The designer chooses a districting plan $\tau \in \Delta(\Delta(X))$ that assigns voters to equipopulous districts $\mu \in [0, 1]$, prior to the realization of an aggregate shock $y \in \mathbb{R}$ with distribution V. The share of type-x voters who vote for the designer's party when the aggregate shock takes value y is deterministic and is denoted by $v(y, x) \in [0, 1]$.[50] The function $v(y, x)$ is assumed to be strictly decreasing in y and strictly increasing in x: that is, higher aggregate shocks are less favorable for the designer, while voters with higher partisanship are more favorable. The designer wins a district μ if and only if she receives a majority of the district vote. Thus, defining $u(y, x) \coloneqq v(y, x) - 1/2$, note that the designer wins a district μ if and only if $y \leq \gamma(\mu)$, where $\gamma(\mu)$ is given by (1). The designer thus wins a district μ with probability $V(\gamma(\mu))$. Finally, the designer chooses τ to maximize her expected seat share, subject to the constraint that all voters are assigned to equipopulous districts: that is, $\mathbb{E}_{\tau}[\mu] = \phi$.[51] The designer's problem is thus

[50] Among other notational differences, the order of the arguments of v is reversed in Kolotilin and Wolitzky (2024a).

[51] As discussed in Kolotilin and Wolitzky (2024a), the equipopulation constraint is strictly enforced in practice, while other constraints on districting (e.g., geographic continuity of districts) are often relatively slack and are thus neglected in much of the gerrymandering literature.

precisely (P), in the state-independent sender case where $V(y, x) = V(y)$. Note that the designer's preferences are state-independent because she cares only about the probability of winning each district and not directly about a district's composition.

In the gerrymandering context, a districting plan is pairwise if each district contains at most two voter types, and a pairwise districting plan is strictly single-dipped if more polarized districts are more favorable for the designer (i.e., if $\gamma(\mu) > \gamma(\mu')$ for all $\mu, \mu' \in \mathrm{supp}(\tau)$ such that supp (μ) contains $x_1 < x_3$ and supp(μ') contains $x_2 \in (x_1, x_3)$.) Since V is state-independent, theorem 4 implies that strictly single-dipped districting is optimal whenever u_x is strictly log-supermodular. This result generalizes a main result of Friedman and Holden (2008; their lemma 1), which shows that strictly single-dipped districting is optimal under an "informative signal property" that is equivalent to log-supermodularity of u_x.[52] As explained in Kolotilin and Wolitzky (2024a), the intuition for this result is that log-supermodularity of u_x means that moderate voters "swing more" with the aggregate shock y than more extreme voters, so a marginal voter is less likely to be pivotal in a district consisting of moderates than in a district that is evenly divided between left-wing and right-wing extremists. The designer then optimally exploits this difference in pivot probabilities by assigning more favorable marginal voters to more polarized districts: that is, by creating a single-dipped districting plan.

Kolotilin and Wolitzky (2024a) go on to apply the duality and complementary slackness developed in the current paper to derive further properties of optimal districting plans. In particular, they give conditions under which optimal districting segregates the strongest opposing voters (as in "pack-and-crack" districting or the disclose-pair plan illustrated in fig. 2*D*) or more moderate voters (as in an alternative plan proposed by Friedman and Holden [2008], which resembles negative assortative disclosure with an interval of disclosed intermediate states).[53]

E. Specific Persuasion Models

Our analysis covers most persuasion models with nonlinear preferences considered to date, including Zhang and Zhou's (2016) model of information disclosure in contests; Guo and Shmaya's (2019) model of persuading a receiver with affiliated private information; and Goldstein and Leitner's (2018) model of optimal stress tests. The main results in the latter two papers show that, respectively, single-peaked and single-dipped negative assortative disclosure are optimal. In appendix E, we describe how

[52] Friedman and Holden (2008) additionally assume that there is a finite number of districts and that u satisfies a "central unimodality" condition.

[53] As argued by Cox and Holden (2011), this is a key question for assessing the likely consequences of restrictions on districting such as those instituted by the Voting Rights Act of 1965.

our analysis covers these prior models and provides some additional results.

VIII. Conclusion

This paper has developed a general model of assigning inputs to productive outputs, which we call optimal productive transport. Our leading application is Bayesian persuasion, but the model also covers other applications including matching, option pricing, and partisan gerrymandering. In the persuasion context, our substantive results provide conditions for all optimal signals to be pairwise, for riskier or safer prospects to induce higher actions, 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 novel duality and complementary slackness theorems, which form the basis of all of our proofs.

We mention a few open issues. First, while the persuasion literature has made progress by allowing unrestricted disclosure policies, the pairwise signals that we highlight 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) and Onuchic and Ray (2023). An interesting observation is that, at least in the separable subcase of our model considered by Rayo (2013) and Onuchic and Ray (2023), our condition (12) 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 if and only if 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 II, our analysis pertains to disclosure mechanisms that do not first elicit the receiver's type, or "public persuasion" in the language of Kolotilin et al. (2017). Public persuasion turns out to be without loss in Kolotilin et al. (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.

Third, while we have taken some steps toward fully characterizing optimal signals by deriving the differential equations (9) and (10) and solving them in a couple examples, much more remains to be done. Equations (9) and (10) are closely related to the optimality and Monge-Ampere equations in optimal transport (e.g., sec. 1.7.6 of Santambrogio 2015). The rich mathematical literature on these equations may hold some insights for fully characterizing optimal signals in certain settings.

Finally, our model could be generalized to allow multidimensional states or actions. We suspect that our results on duality (lemmas 1 and 2), complementary slackness (theorem 1), and pairwise signals (theorem 2) generalize up to some technicalities.[54] Generalizing our other results would require a more general notion of single-dippedness. 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 (2024).

Appendix A

Characterization of Strict Aggregate Single-Crossing

We present two alternative conditions that are equivalent to strict aggregate single-crossing of u. Condition 2 is analogous to the "signed-ratio monotonicity" conditions for weak aggregate single-crossing 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 app. F). Condition 3 is novel. It corresponds to strict monotonicity of u (i.e., $u_y(y, x) < 0$), up to a normalizing factor $g(y) > 0$.

Lemma 3. Let assumption 1 hold. The following statements are equivalent:

1. Assumption 2 holds.
2. For all x, x', and y, we have

$$u(y, x) = 0 \Longrightarrow u_y(y, x) < 0, \tag{14}$$

$$u(y, x) < 0 < u(y, x') \Longrightarrow u(y, x')u_y(y, x) - u(y, x)u_y(y, x') < 0. \tag{15}$$

3. There exists a differentiable function $g(y) > 0$ such that $\tilde{u}(y, x) = u(y, x)/g(y)$ satisfies $\tilde{u}_y(y, x) < 0$ for all (y, x).

Appendix B

Uniqueness

This appendix presents a notable technical result: under a regularity condition, strict single-dippedness implies that there is a unique optimal signal. It also shows that conditions for uniqueness are much weaker when full disclosure is optimal.

[54] For example, our proof of theorem 1 is facilitated by the existence of a bijection between actions y and states $\chi(y)$ such that $u(y, \chi(y)) = 0$ (cf. assumption 4).

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

We say that a strictly single-dipped set Λ is "regular" if for each $y \in Y_\Lambda$, there exists $\varepsilon > 0$ such that either (i) $\chi_1(\tilde{y}) = \chi_2(\tilde{y})$ for all $\tilde{y} \in (y - \varepsilon, y) \cap Y_\Lambda$ or (ii) $\chi_1(\tilde{y}) < \chi_2(\tilde{y})$ for all $\tilde{y} \in (y - \varepsilon, y) \cap Y_\Lambda$. This regularity condition rules out pathological cases where states switch infinitely many times from being disclosed to being paired. This condition is satisfied in every example in the literature that we know of.

Theorem 7. If $X = [0, 1]$, ϕ has a density, and Λ is strictly single-dipped and regular, then there is a unique optimal signal.

In martingale optimal transport, the optimal solution is unique under the martingale Spence-Mirrlees condition (e.g., proposition 3.5 in Beiglböck, Henry-Labordère, and Touzi 2017), which coincides with our condition for the optimality of strict single-dippedness in the linear receiver case. The key implication of theorem 7 is that the optimal marginal distribution of actions is unique. There is no analogue of this result in martingale optimal transport, where this marginal distribution is fixed.

To see the intuition for theorem 7, consider the case where ϕ is discrete and χ_2 is strictly increasing. Let $\bar{y} = \max_{\mu \in \Lambda} \gamma(\mu)$ be the highest action that can be optimally induced. Since Λ is strictly single-dipped, there is a unique posterior $\bar{\mu}$ in Λ inducing action $\bar{y}$, namely, $\bar{\mu} = \delta_{\chi_2(\bar{y})}$ if $\chi_1(\bar{y}) = \chi_2(\bar{y})$ and $\bar{\mu} = \bar{\rho}\delta_{\chi_1(\bar{y})} + (1 - \bar{\rho})\delta_{\chi_2(\bar{y})}$ where $\bar{\rho} \in (0, 1)$ is uniquely determined by (1) if $\chi_1(\bar{y}) < \chi_2(\bar{y})$. Since χ_2 is strictly increasing, for any optimal signal τ, the state $\chi_2(\bar{y})$ can only induce action $\bar{y}$ and thus $\tau(\bar{\mu}) = \phi(\chi_2(\bar{y}))$ if $\chi_1(\bar{y}) = \chi_2(\bar{y})$ and $\tau(\bar{\mu}) = \phi(\chi_2(\bar{y}))/(1 - \bar{\rho})$ if $\chi_1(\bar{y}) < \chi_2(\bar{y})$. Working our way through the support of ϕ from the highest state to the lowest in this fashion, we obtain the unique value of $\tau(\mu)$ for each $\mu \in \Lambda$. When ϕ has a density, the possibility that χ_2 may be only weakly increasing does not threaten uniqueness of the optimal signal, because the set of states corresponding to flat regions of χ_2 is at most countable and thus has ϕ-measure 0. Finally, our regularity condition ensures that the above argument extends easily from the discrete case to the continuous one.

In addition, when the prior has full support and the contact set is pairwise (e.g., 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 x_1 and x_2 to induce an action y. Then the signal that discloses all other states while pooling x_1 and x_2 to induce y is also optimal. But then the signal that discloses all other states while pooling x_1, x_2, and the third state $\chi(y) \neq x_1, x_2$ to induce y would also be optimal—but this signal is not pairwise, which is a contradiction.

Remark 4. Assume that $X = [0, 1]$. If the contact set is pairwise and full disclosure is optimal, then it is uniquely optimal.

Appendix C

Proofs

C1. Proof of Lemma 2

The proof of lemma 2 remains valid without assumption 4 and when X is an arbitrary compact metric space.

One direction is obvious. If (p, q) is feasible for (D'), then, for all $\mu \in \Delta(X)$, we have

$$\int_X p(x)\mathrm{d}\mu(x) \geq \int_X (V(\gamma(\mu), x) + q(\gamma(\mu))u(\gamma(\mu), x))\mathrm{d}\mu(x)$$
$$= \int_X V(\gamma(\mu), x)\mathrm{d}\mu(x) = W(\mu),$$

so p is feasible for (D).

Suppose now that p is feasible for (D). By (ZP) for $\mu = \delta_x$, with $x \in X$, we have

$$p(x) \geq V(\gamma(\delta_x), x), \quad \text{for all } x \in X. \tag{16}$$

For $x_1, x_2 \in X$ and $y \in Y$ such that $u(y, x_1) < 0 < u(y, x_2)$, let

$$\mu = \frac{u(y, x_2)}{u(y, x_2) - u(y, x_1)}\delta_{x_1} + \frac{-u(y, x_1)}{u(y, x_2) - u(y, x_1)}\delta_{x_2}.$$

Then $\gamma(\mu) = y$ by (1), so, by (ZP),

$$\frac{p(x_1) - V(y, x_1)}{u(y, x_1)} \leq \frac{p(x_2) - V(y, x_2)}{u(y, x_2)}, \quad \text{if } u(y, x_1) < 0 < u(y, x_2). \tag{17}$$

By (16), it suffices to show that there exists a bounded, measurable q such that (ZP') holds for all $y \neq \gamma(\delta_x)$ or, equivalently, $q(y) \in [\underline{q}(y), \bar{q}(y)]$, for all $y \in Y$, where $\underline{q}$ and $\bar{q}$ are defined by

$$\underline{q}(y) = \begin{cases} \sup_{x_1 \in X: u(y, x_1) < 0} \dfrac{p(x_1) - V(y, x_1)}{u(y, x_1)}, & y > 0, \\ -\infty, & y = 0, \end{cases}$$

$$\bar{q}(y) = \begin{cases} \inf_{x_2 \in X: u(y, x_2) > 0} \dfrac{p(x_2) - V(y, x_2)}{u(y, x_2)}, & y < 1, \\ +\infty, & y = 1. \end{cases}$$

That is, it suffices to show that the correspondence Q given by $Q(y) = [\underline{q}(y), \bar{q}(y)]$, for all $y \in Y$, admits a bounded, measurable selection q. We now show that there exists $C > 0$ such that $Q(y) \cap [-C, C]$ is nonempty valued for all $y \in Y$, so q given by $q(y) = \arg\min_{r \in Q(y)} |r|$, for all $y \in Y$, is a required selection, by the measurable maximum theorem (theorem 18.19 in Aliprantis and Border 2006). By (17), $Q(y)$ is nonempty for all $y \in (0, 1)$, so it suffices to show that there exists $C > 0$ such that $\underline{q}(y) \leq C$ and $\bar{q}(y) \geq -C$.

Define

$$\tilde{q}(y, x) = \begin{cases} \dfrac{V_y(y, x)}{-u_y(y, x)}, & u(y, x) = 0, \\ \dfrac{V(\gamma(\delta_x), x) - V(y, x)}{u(y, x)}, & u(y, x) \neq 0. \end{cases}$$

Recall that assumption 2 requires that $u_y(y, x) < 0$ when $u(y, x) = 0$; so $\tilde{q}(y, x)$ is well-defined. It suffices to show that there exists $C > 0$ such that $|\tilde{q}(y, x)| \leq C$ for all $(y, x) \in Y \times X$, because then, by (16), we have

$$\frac{p(x_2) - V(y, x_2)}{u(y, x_2)} \geq \frac{V(\gamma(\delta_{x_2}), x_2) - V(y, x_2)}{u(y, x_2)} = \tilde{q}(y, x_2), \qquad \text{if } u(y, x_2) > 0,$$

$$\frac{p(x_1) - V(y, x_1)}{u(y, x_1)} \leq \frac{V(\gamma(\delta_{x_1}), x_1) - V(y, x_1)}{u(y, x_1)} = \tilde{q}(y, x_1), \qquad \text{if } u(y, x_1) < 0,$$

so, by the definition of $\bar{q}$ and $\underline{q}$, we get that $\bar{q}(y) \geq -C$ and $\underline{q}(y) \leq C$.

Finally, to show that there exists $C > 0$ such that $|\tilde{q}(y, x)| \leq C$ for all $(y, x) \in Y \times X$, it suffices to show that $\tilde{q}$ is continuous on the compact set $Y \times X$. By Berge's theorem, $\gamma(\delta_x)$ is continuous in x, as it is a unique maximizer of a continuous function $U(y, x) = \int_0^y u(\tilde{y}, x)\mathrm{d}\tilde{y}$. Note that $\tilde{q}$ is continuous at each (y, x) such that $u(y, x) \neq 0$, because V, u, and γ are continuous. Next, consider (y, x) such that $u(y, x) = 0$, or equivalently $y = \gamma(\delta_x)$. For each $(y', x') \in Y \times X$, there exists $\hat{y}$ between $\gamma(\delta_{x'})$ and y' such that

$$[V(\gamma(\delta_{x'}), x') - V(y', x')]u_y(\hat{y}, x') = -V_y(\hat{y}, x')u(y', x'),$$

by the mean value theorem applied to the function

$$[V(\gamma(\delta_{x'}), x') - V(\tilde{y}, x')]u(y', x') - [V(\gamma(\delta_{x'}), x') - V(y', x')]u(\tilde{y}, x'),$$

where the argument $\tilde{y}$ is between $\gamma(\delta_{x'})$ and y'. Thus,

$$\tilde{q}(y', x') - \tilde{q}(y, x) = \frac{V_y(\hat{y}, x')}{-u_y(\hat{y}, x')} - \frac{V_y(y, x)}{-u_y(y, x)}.$$

If $(y', x') \to (y, x)$, then $(\hat{y}, x') \to (y, x)$, because $\gamma(\delta_x)$ is continuous in x. Hence, $\tilde{q}(y', x') \to \tilde{q}(y, x)$, because V_y and u_y are continuous. This shows that $\tilde{q}$ is continuous on $Y \times X$. QED

C2. Proof of Theorem 1

The proof of theorem 1 remains valid if assumption 4 is replaced with the weaker requirement that $u(y, x)$ satisfies strict single-crossing in x: for all y and $x < x'$, we have $u(y, x) \geq 0 \Longrightarrow u(y, x') > 0$.

By lemma 2, there exists $q \in B(Y)$ such that (p, q) is feasible for (D'). Recall that, under strict single-crossing of $u(y, x)$ in x, for each action y, there is a unique state $\chi(y)$ such that $u(y, \chi(y)) = 0$. First, redefine $q(y) = -V_y(y, \chi(y))/u_y(y, \chi(y))$ for all $y \in Y$ such that $p(\chi(y)) = V(y, \chi(y))$. We now show that (p, q) is still feasible for (D'). Fix any y such that $p(\chi(y)) = V(y, \chi(y))$ and any $x \in X$. For any $\varepsilon \in (0, 1)$, define $y_\varepsilon \in Y$ as a unique solution to $(1 - \varepsilon)u(y_\varepsilon, \chi(y)) + \varepsilon u(y_\varepsilon, x) = 0$. By the implicit function theorem,

$$\lim_{\varepsilon \downarrow 0} \frac{y_\varepsilon - y}{\varepsilon} = \frac{u(y, x)}{-u_y(y, \chi(y))}.$$

By (ZP'), we have

$$V(y, \chi(y)) \geq V(y_\varepsilon, \chi(y)) + q(y_\varepsilon)u(y_\varepsilon, \chi(y)) \quad \text{and}$$
$$p(x) \geq V(y_\varepsilon, x) + q(y_\varepsilon)u(y_\varepsilon, x).$$

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

$$p(x) \geq V(y, x) + \frac{(1-\varepsilon)[V(y_\varepsilon, \chi(y)) - V(y, \chi(y))] + \varepsilon[V(y_\varepsilon, x) - V(y, x)]}{\varepsilon}.$$

Taking the limit $\varepsilon \to 0$ gives

$$p(x) \geq V(y, x) + \frac{V_y(y, \chi(y))}{-u_y(y, \chi(y))}u(y, x),$$

showing that (p, q) is still feasible with redefined q.

Thus, (3) holds, by construction, for all degenerate $\mu \in \Lambda$. Since, for nondegenerate $\mu \in \Lambda$, (4) integrated over μ yields (3), it remains to show that (4) holds for each nondegenerate $\mu \in \Lambda$.

Fix a nondegenerate $\mu \in \Lambda$, so that

$$\int_X (p(x) - V(\gamma(\mu), x) - q(\gamma(\mu))u(\gamma(\mu), x))\mathrm{d}\mu(x) = 0.$$

Since the integrand is nonnegative and continuous in x, it follows that

$$p(x) = V(\gamma(\mu), x) + q(\gamma(\mu))u(\gamma(\mu), x), \quad \text{for all } x \in \operatorname{supp}(\mu). \tag{18}$$

Since μ is nondegenerate, strict single-crossing of $u(y, x)$ in x implies that there exist $x_1, x_2 \in \operatorname{supp}(\mu)$ such that $x_1 < \chi(\gamma(\mu)) < x_2$. Thus, by (ZP'), for every $\tilde{y} \in Y$, we have

$$p(x_1) = V(\gamma(\mu), x_1) + q(\gamma(\mu))u(\gamma(\mu), x_1) \geq V(\tilde{y}, x_1) + q(\tilde{y})u(\tilde{y}, x_1).$$

Therefore, for every $\tilde{y} > \gamma(\mu)$, we have

$$\frac{q(\tilde{y}) - q(\gamma(\mu))}{\tilde{y} - \gamma(\mu)} \geq \frac{1}{-u(\tilde{y}, x_1)}\left[\frac{V(\tilde{y}, x_1) - V(\gamma(\mu), x_1)}{\tilde{y} - \gamma(\mu)} + q(y)\frac{u(\tilde{y}, x_1) - u(\gamma(\mu), x_1)}{\tilde{y} - \gamma(\mu)}\right].$$

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

$$\underline{q}'_+(\gamma(\mu)) := \liminf_{\tilde{y} \downarrow \gamma(\mu)} \frac{q(\tilde{y}) - q(\gamma(\mu))}{\tilde{y} - \gamma(\mu)} \geq C_1,$$

where

$$C_1 = -\frac{1}{u(\gamma(\mu), x_1)}[V_y(\gamma(\mu), x_1) + q(\gamma(\mu))u_y(\gamma(\mu), x_1)].$$

Applying a similar argument for $x = x_1$ and $\tilde{y} < \gamma(\mu)$, we get

$$\bar{q}'_-(\gamma(\mu)) := \limsup_{\tilde{y} \uparrow \gamma(\mu)} \frac{q(\tilde{y}) - q(\gamma(\mu))}{\tilde{y} - \gamma(\mu)} \leq C_1.$$

Similarly, considering $x = x_2$ with $\tilde{y} > \gamma(\mu)$ and $\tilde{y} < \gamma(\mu)$, we get

$$\overline{q}'_+(\gamma(\mu)) \coloneqq \limsup_{\tilde{y}\downarrow\gamma(\mu)} \frac{q(\tilde{y}) - q(\gamma(\mu))}{\tilde{y} - \gamma(\mu)} \leq C_2,$$
$$\underline{q}'_-(\gamma(\mu)) \coloneqq \liminf_{\tilde{y}\uparrow\gamma(\mu)} \frac{q(\tilde{y}) - q(\gamma(\mu))}{\tilde{y} - \gamma(\mu)} \geq C_2,$$

where

$$C_2 = -\frac{1}{u(\gamma(\mu), x_2)}[V_y(\gamma(\mu), x_2) + q(\gamma(\mu))u_y(\gamma(\mu), x_2)].$$

In sum, we have

$$C_1 \leq \underline{q}'_+(\gamma(\mu)) \leq \overline{q}'_+(\gamma(\mu)) \leq C_2 \quad \text{and} \quad C_2 \leq \underline{q}'_-(\gamma(\mu)) \leq \overline{q}'_-(\gamma(\mu)) \leq C_1.$$

We see that $C_1 = C_2$ and all four Dini derivatives of q at $\gamma(\mu)$ coincide, so q has a derivative $q'(\gamma(\mu))$ at $\gamma(\mu)$ that satisfies $q'(\gamma(\mu)) = C_1 = C_2$.

Since $x_1, x_2 \in \text{supp}(\mu)$ are arbitrary, (4) holds for all $x \in \text{supp}(\mu)$ with $x \neq \chi(\gamma(\mu))$. For $x \in \text{supp}(\mu)$ with $x = \chi(\gamma(\mu))$, (4) holds because, as shown above, we have that $q(\gamma(\mu)) = -V_y(\gamma(\mu), \chi(\gamma(\mu)))/u_y(\gamma(\mu), \chi(\gamma(\mu)))$. QED

C3. Proof of Theorem 3

The proof of theorem 3 remains valid if the condition $u_x(y, x) > 0$ in assumption 4 is replaced with strict single-crossing of $u(y, x)$ in x.

We will prove that Λ is single-dipped, which implies that every optimal signal is single-dipped. We start with an appropriate version of the theorem of alternative.

LEMMA 4. Exactly one of the following two alternatives holds.

1. There exists $\alpha > 0$ such that $\alpha R \leq 0$.
2. There exists $\beta \geq 0$ such that $R\beta \geq 0$ and $R\beta \neq 0$.

Proof. Clearly, statements 1 and 2 cannot both hold, because premultiplying $R\beta \geq 0$ with $R\beta \neq 0$ by $\alpha > 0$ yields $\alpha R\beta > 0$, whereas postmultiplying $\alpha R \leq 0$ by $\beta \geq 0$ yields $\alpha R\beta \leq 0$.

Now suppose that statement 1 does not hold. Then there does not exist $\alpha \geq 0$ such that $\alpha(R \;\; -I) \leq (0 \;\; -e)$ 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 $\beta \geq 0$ and $\gamma \geq 0$ such that $(R \;\; -I)^T(\beta \;\; \gamma) \geq 0$ and $-e\gamma < 0$, which in turn shows that statement 2 holds. QED

We prove the theorem by contraposition. Suppose that Λ is not single-dipped, so there exist $\mu_1, \mu_2 \in \Lambda$ and $x_1 < x_2 < x_3$ such that $x_1, x_3 \in \text{supp}(\mu_1)$, $x_2 \in \text{supp}(\mu_2)$, and $\gamma(\mu_1) < \gamma(\mu_2)$. By strict single-crossing of $u(y, x)$ in x, without loss, we can assume that $x_1 < \chi(\gamma(\mu_1)) < x_3$, by redefining $x_1 = \min \text{supp}(\mu_1)$ and $x_3 = \max \text{supp}(\mu_1)$, if necessary.

By (ZP') and theorem 1, we have

$$V(\gamma(\mu_1), x_1) + q(\gamma(\mu_1))u(\gamma(\mu_1), x_1) \geq V(\gamma(\mu_2), x_1) + q(\gamma(\mu_2))u(\gamma(\mu_2), x_1),$$

$$V(\gamma(\mu_2), x_2) + q(\gamma(\mu_2))u(\gamma(\mu_2), x_2) \geq V(\gamma(\mu_1), x_2) + q(\gamma(\mu_1))u(\gamma(\mu_1), x_2),$$

$$V(\gamma(\mu_1), x_3) + q(\gamma(\mu_1))u(\gamma(\mu_1), x_3) \geq V(\gamma(\mu_2), x_3) + q(\gamma(\mu_2))u(\gamma(\mu_2), x_3).$$

By (3), we have, for $i = 1, 2$, $q(\gamma(\mu_i)) = -\mathbb{E}_{\mu_i}[V_y(\gamma(\mu_i), x)]/\mathbb{E}_{\mu_i}[u_y(\gamma(\mu_i), x)] > 0$, where the inequality follows from assumptions 2 and 4. Thus, the vector $\alpha = (1, q(\gamma(\mu_1)), q(\gamma(\mu_2)))$ is strictly positive and satisfies $\alpha R \leq 0$. By lemma 4, there does not exist a vector $\beta \geq 0$ such that $R\beta \geq 0$ and $R\beta \neq 0$, as desired. QED

C4. Proof of Theorem 4

The proof uses the following five lemmas, whose proofs are deferred to appendix F. We start with the second part of the theorem, and we show that Λ is strictly single-dipped (-peaked), which implies that every optimal signal is strictly single-dipped (-peaked).

LEMMA 5. If $u_{yx}(y, x)/u_x(y, x)$ and $V_{yx}(y, x)/u_x(y, x)$ are increasing (decreasing) in x for all y, with at least one of them strictly increasing (decreasing), then $|S| > (<)0$ for all y and $x_1 < x_2 < x_3$ such that $x_1 < \chi(y) < x_3$.

LEMMA 6. If $u_{yx}(y, x)/u_x(y, x)$ and $V_{yx}(y_2, x)/u_x(y_1, x)$ are increasing (decreasing) in x for all y and $y_2 \geq (\leq) y_1$, with at least one of them strictly increasing (decreasing), then $|R| > (<)0$ for all $x_1 < x_2 < x_3$ and all $y_2 > (<) y_1$ such that $x_1 < \chi(y_1) < x_3$.

LEMMA 7. If $u_{yx}(y, x)/u_x(y, x)$ is increasing in x for all y, then for all $x_1 < x_2 < x_3$ and all $y_2 > y_1$ such that $x_1 < \chi(y_1) < x_3$, we have

$$u(y_2, x_3)u(y_1, x_1) > u(y_2, x_1)u(y_1, x_3),$$

$$u(y_2, x_2)u(y_1, x_1) > u(y_2, x_1)u(y_1, x_2),$$

$$u(y_2, x_3)u(y_1, x_2) > u(y_2, x_2)u(y_1, x_3).$$

LEMMA 8. If $V_{yx}(y_2, x)/u_x(y_1, x)$ is decreasing in x for all $y_2 \leq y_1$, then for all $x_1 < x_2 < x_3$ and all $y_2 < y_1$ such that $x_1 < \chi(y_1) < x_3$, we have

$$\frac{u(y_1, x_1)}{V(y_1, x_1) - V(y_2, x_1)} < \frac{u(y_1, x_2)}{V(y_1, x_2) - V(y_2, x_2)} < \frac{u(y_1, x_3)}{V(y_1, x_3) - V(y_2, x_3)}.$$

LEMMA 9. Suppose that V^n is a sequence of functions satisfying assumption 1 such that V_y^n converges uniformly to V_y, and suppose that the contact sets Λ^n under V^n are single-dipped (-peaked). Then there exists a single-dipped (-peaked) optimal signal under V.

Now, the set Λ is single-dipped (-peaked) by theorem 3 with

$$\beta = \begin{pmatrix} u(y_2, x_3)u(y_1, x_2) - u(y_2, x_2)u(y_1, x_3) \\ u(y_2, x_3)u(y_1, x_1) - u(y_2, x_1)u(y_1, x_3) \\ u(y_2, x_2)u(y_1, x_1) - u(y_2, x_1)u(y_1, x_2) \end{pmatrix} \quad \left(\beta = \begin{pmatrix} \dfrac{u(y_2, x_1)}{V(y_2, x_1) - V(y_1, x_1)} \\ \dfrac{u(y_2, x_2)}{V(y_2, x_2) - V(y_1, x_2)} \\ \dfrac{u(y_2, x_3)}{V(y_2, x_3) - V(y_1, x_3)} \end{pmatrix} \right),$$

as follows from lemmas 6 and 7 (lemmas 6 and 8). Moreover, $|\text{supp}(\mu)| \leq 2$ for all $\mu \in \Lambda$ by theorem 2 and lemma 5, showing that Λ is strictly single-dipped (-peaked).

Finally, we prove the first part of the theorem. Consider V^n such that

$$V_y^n(y, x) = V_y(y, x) + \int_0^x \frac{\tilde{v}(x)}{n} u_x(y, \tilde{x}) \mathrm{d}\tilde{x},$$

where $\tilde{v}(x)$ is a continuous, strictly positive, and strictly increasing (decreasing) function on $[0, 1]$. Then $V_y^n(y, x) > 0$ because $V_y(y, x) > 0$ and $u_x(y, x) > 0$ for all (y, x), by assumption 4. Moreover, for all $y_2 \geq (\leq) y_1$,

$$\frac{V_{yx}^n(y_2, x)}{u_x(y_1, x)} = \frac{V_{yx}(y_2, x)}{u_x(y_1, x)} + \frac{\tilde{v}(x)}{n} \frac{u_x(y_2, x)}{u_x(y_1, x)}$$

is strictly increasing (decreasing) in x, because $\tilde{v}(x)$ is strictly positive and strictly increasing (decreasing) in x; $V_{yx}(y_2, x)/u_x(y_1, x)$ is increasing (decreasing) in x; and $u_x(y_2, x)/u_x(y_1, x)$ is increasing in x, since $u_{yx}(y, x)/u_x(y, x)$ is increasing (decreasing) in x. Thus, by lemma 9, there exists an optimal single-dipped (-peaked) signal. QED

C5. Proof of Theorem 6

The proof of theorem 6 remains valid if assumption 4 is replaced with strict single-crossing of $u(y, x)$ in x. Since $X = [0, 1]$, Λ is strictly single-dipped, and for all $x_1 < x_2$ there exists $\rho \in (0, 1)$ such that (12) holds, it follows that $\chi_1(y_2) \leq \chi_1(y_1)$ for all $y_1 < y_2$ in Y_Λ, and thus Λ is negative assortative. Suppose by contradiction that there exist $y_1 < y_2$ in Y_Λ such that $\chi_1(y_1) < \chi_1(y_2)$. Then $\chi_2(y_1) \leq \chi_1(y_2)$, as otherwise there would exist $\mu_1, \mu_2 \in \Lambda$ such that $\gamma(\mu_1) = y_1 < y_2 = \gamma(\mu_2)$, and $\chi_1(y_1) < \chi_1(y_2) < \chi_2(y_1)$ contradicting that Λ is single-dipped. By lemma 13, there exist $y_1' \leq y_1$ and $y_2' \leq y_2$ in Y_Λ such that $\chi_1(y_1) \leq \chi_1(y_1') = \chi_2(y_1') \leq \chi(y_1) \leq \chi_2(y_1) \leq \chi_1(y_2) \leq \chi_1(y_2') = \chi_2(y_2') \leq \chi_2(y_2)$, and thus $y_1' \leq y_2'$. In fact, $y_1' < y_2'$, as otherwise we would have $\chi_1(y_1) \leq \chi(y_1) = \chi_2(y_1) = \chi_1(y_2)$, which implies $\chi_1(y_1) = \chi_2(y_1) = \chi_1(y_2)$, contradicting $\chi_1(y_1) < \chi_1(y_2)$. Thus, denoting $x_1 = \chi(y_1') < \chi(y_2') = x_2$, we have $\delta_{x_1}, \delta_{x_2} \in \Lambda$. For any $\mu = \rho\delta_{x_1} + (1 - \rho)\delta_{x_2}$ with $\rho \in (0, 1)$, we have

$$p(x_1) = V(\gamma(x_1), x_1) \geq V(\gamma(\mu), x_1) + q(\gamma(\mu))u(\gamma(\mu), x_1),$$

$$p(x_2) = V(\gamma(x_2), x_2) \geq V(\gamma(\mu), x_2) + q(\gamma(\mu))u(\gamma(\mu), x_2),$$

by (ZP) and the definition of Λ. Adding the first inequality multiplied by ρ and the second inequality multiplied by $1 - \rho$, we obtain that (12) fails for all $\rho \in (0, 1)$, yielding a contradiction.

Now assuming that ϕ has a density f and Λ is negative assortative, we will show that the functions χ_1 and χ_2 are continuous and satisfy the differential equations (9) and (10) and the boundary condition (11). Since the closure of $X^\star = \cup_{\mu \in \Lambda} \text{supp}(\mu)$ is $X = [0, 1]$, it follows that the closure of the union of the images of the functions χ_1 and χ_2 must also be equal to $[0, 1]$. Since χ_1 is decreasing and χ_2 is increasing on the compact domain Y_Λ, and since $\chi_1(y) \leq \chi(y) \leq \chi_2(y)$ for all $y \in Y_\Lambda$, it follows that χ_1 and χ_2 are continuous functions such that $\chi_1(\underline{y}) = \chi(\underline{y}) = \chi_2(\underline{y})$, $\chi_1(y) < \chi(y) < \chi_2(y)$ for all $y > \underline{y}$ in Y_Λ, $\chi_1(\bar{y}) = 0$,

$\chi_2(\bar{y}) = 1$, and $(\chi_1(\underline{z}_i), \chi_2(\underline{z}_i)) = (\chi_1(\bar{z}_i), \chi_2(\bar{z}_i))$ for all i, where $\{(\underline{z}_i, \bar{z}_i)\}_i$ is an at most countable set of disjoint open intervals comprising the set $[\underline{y}, \bar{y}] \setminus Y_\Lambda$. Since ϕ has a density, the measure of the endpoints of these intervals is zero, and hence the set of optimal signals is unaffected if we extend the domain of χ_1 and χ_2 to $[\underline{y}, \bar{y}]$ by setting $\chi_1(y) = \chi_1(\underline{z}_i) = \chi_1(\bar{z}_i)$ and $\chi_2(y) = \chi_2(\underline{z}_i) = \chi_2(\bar{z}_i)$ for all $y \in (\underline{z}_i, \bar{z}_i)$. In sum, without loss of generality, we can assume that χ_1 and χ_2 are continuous monotone functions defined on $[\underline{y}, \bar{y}]$ that satisfy (11) and $\chi_1(y) < \chi(y) < \chi_2(y)$ for all $y \in (\underline{y}, \bar{y}]$.

Since χ_1 is continuously decreasing on $[\underline{y}, \bar{y}]$, χ_2 is continuously increasing on $[\underline{y}, \bar{y}]$, and ϕ has a density, we can rewrite (OB) for $\tilde{Y} = [y, y']$, with $\underline{y} \le y < y' \le \bar{y}$, as

$$\int_y^{y'} u(\tilde{y}, \chi_1(\tilde{y}))(-\mathrm{d}\phi([0, \chi_1(\tilde{y})])) + \int_y^{y'} u(\tilde{y}, \chi_2(\tilde{y}))\mathrm{d}\phi([0, \chi_2(\tilde{y})]) = 0.$$

Taking the limit $y' \downarrow y$, we obtain (9) for all $y \in [\underline{y}, \bar{y}]$.

By theorem 1, for all $y > \underline{y}$ in Y_Λ,

$$V_y(y, \chi_1(y)) + q(y)u_y(y, \chi_1(y)) + q'(y)u(y, \chi_1(y)) = 0,$$
$$V_y(y, \chi_2(y)) + q(y)u_y(y, \chi_2(y)) + q'(y)u(y, \chi_2(y)) = 0.$$

Solving for $q(y)$ and $q'(y)$, we get, for all $y > \underline{y}$ in Y_Λ,

$$q(y) = \frac{V_y(y, \chi_1(y))u(y, \chi_2(y)) - V_y(y, \chi_2(y))u(y, \chi_1(y))}{u(y, \chi_1(y))u_y(y, \chi_2(y)) - u(y, \chi_2(y))u_y(y, \chi_1(y))},$$
$$q'(y) = \frac{V_y(y, \chi_1(y))u_y(y, \chi_2(y)) - V_y(y, \chi_2(y))u_y(y, \chi_1(y))}{u_y(y, \chi_1(y))u(y, \chi_2(y)) - u_y(y, \chi_2(y))u(y, \chi_1(y))},$$

where the denominators in the expressions for $q(y)$ and $q'(y)$ are not equal to 0, by assumption 2. Noting that q' is the derivative of q gives (10) for all $y > \underline{y}$ in Y_Λ. QED

References

Acciaio, B., M. Beiglbock, F. Penkner, and W. Schachermayer. 2016. "A Model-Free Version of the Fundamental Theorem of Asset Pricing and the Super-Replication Theorem." *Math. Finance* 26 (2): 233–51.

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

Alonso, R., and O. Câmara. 2016. "Bayesian Persuasion with Heterogeneous Priors." *J. Econ. Theory* 165:672–706.

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

Arieli, I., Y. Babichenko, and F. Sandomirskiy. 2024. "Feasible Conditional Belief Distributions." Arxiv working paper, https://arxiv.org/abs/2307.07672.

Arnott, R., and J. Rowse. 1987. "Peer Group Effects and Educational Attainment." *J. Public Econ.* 32 (3): 287–305.

Aumann, R. J., and M. Maschler. 1995. *Repeated Games with Incomplete Information.* Cambridge, MA: MIT Press.

Becker, G., and K. Murphy. 2000. *Social Economics: Market Behavior in a Social Environment.* Cambridge, MA: Belknap.

Becker, G. S. 1973. “A Theory of Marriage: Part I.” *J.P.E.* 81 (4): 813–46.
Beiglböck, M., P. Henry-Labordère, 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): 3005–13.
Beiglböck, M., and N. Juillet. 2016. “On a Problem of Optimal Transport under Marginal Martingale Constraints.” *Ann. Probability* 44 (1): 42–106.
Beiglböck, M., M. Nutz, and N. Touzi. 2017. “Complete Duality for Martingale Optimal Transport on the Line.” *Ann. Probability* 45 (5): 3038–74.
Benabou, R. 1996. “Equity and Efficiency in Human Capital Investment: The Local Connection.” *Rev. Econ. Studies* 63 (2): 237–64.
Bergemann, D., B. Brooks, and S. Morris. 2015. “The Limits of Price Discrimination.” *A.E.R.* 105 (3): 921–57.
Boerma, J., A. Tsyvinski, and A. Zimin. 2025. “Sorting with Teams.” *J.P.E.* 133 (2): 421–91.
Boucher, V., M. Rendall, P. Uschev, and Y. Zenou. 2024. “Toward a General Theory of Peer Effects.” *Econometrica* 92 (2): 543–65.
Breeden, D., and R. Litzenberger. 1978. “Prices of State-Contingent Claims Implicit in Option Prices.” *J. Bus.* 51 (5): 621–51.
Brenier, Y. 1991. “Polar Factorization and Monotone Rearrangement of Vector-Valued Functions.” *Communications Pure and Appl. Math.* 44 (4): 375–417.
Brzutowski, T. 2024. “Encouraging a Go-Getter.” Proceedings of the 25th ACM Conference on Economics and Computation.
Choi, M., and L. Smith. 2017. “Ordinal Aggregation Results via Karlin’s Variation Diminishing Property.” *J. Econ. Theory* 168:1–11.
Cox, A. B., and R. R. Holden. 2011. “Reconsidering Racial and Partisan Gerrymandering.” *Univ. Chicago Law Rev.* 78:553–604.
Crawford, V., and J. Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50 (6): 1431–51.
de Bartolome, C. 1990. “Equilibrium and Inefficiency in a Community Model with Peer Group Effects.” *J.P.E.* 98 (1): 110–33.
De March, H., and N. Touzi. 2019. “Irreducible Convex Paving for Decomposition of Multidimensional Martingale Transport Plans.” *Ann. Probability* 47 (3): 1726–74.
Dekel, E. 1986. “An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom.” *J. Econ. Theory* 40 (2): 304–18.
Dizdar, D., and E. Kováč. 2020. “A Simple Proof of Strong Duality in the Linear Persuasion Problem.” *Games and Econ. Behavior* 122:407–12.
Durlauf, S. 2004. “Neighborhood Effects.” In *Handbook of Regional and Urban Economics*, vol. 4, *Cities and Geography*, edited by J. Vernon Henderson and Jacques-François Thisse, 2173–242. Amsterdam: Elsevier.
Dworczak, P., and A. Kolotilin. 2024. “The Persuasion Duality.” *Theoretical Econ.* 19 (4): 1701–55.
Dworczak, P., and G. Martini. 2019. “The Simple Economics of Optimal Persuasion.” *J.P.E.* 127 (5): 1993–2048.
Eeckhout, J. 2018. “Sorting in the Labor Market.” *Ann. Rev. Econ.* 10 (1): 1–29.
Ellickson, B., B. Grodal, S. Scotchmer, and W. Zame. 1999. “Clubs and the Market.” *Econometrica* 67 (5): 1185–217.
Epple, D., and R. Romano. 1998. “Competition between Private and Public Schools, Vouchers, and Peer-Group Effects.” *A.E.R.* 88 (1): 33–62.

Friedman, J. N., and R. T. Holden. 2008. “Optimal Gerrymandering: Sometimes Pack, but Never Crack.” *A.E.R.* 98 (1): 113–44.

Gale, D. 1989. *The Theory of Linear Economic Models*. Chicago: Univ. Chicago Press.

Galichon, A., P. Henry-Labordère, and N. Touzi. 2014. “A Stochastic Control Approach to No-Arbitrage Bounds Given Marginals, with an Application to Lookback Options.” *Ann. Appl. Probability* 24 (1): 312–36.

Galperti, S., A. Levkun, and J. Perego. 2024. “The Value of Data Records.” *Rev. Econ. Studies* 91 (2): 1007–38.

Gangbo, W., and R. McCann. 1996. “The Geometry of Optimal Transportation.” *Acta Mathematica* 177 (2): 113–61.

Gentzkow, M., and E. Kamenica. 2016. “A Rothschild-Stiglitz Approach to Bayesian Persuasion.” *A.E.R. Papers and Proc.* 106 (5): 597–601.

Ghoussoub, N., Y.-H. Kim, and T. Lim. 2019. “Structure of Optimal Martingale Transport Plans in General Dimensions.” *Ann. Probability* 47 (1): 109–64.

Goldstein, I., and Y. Leitner. 2018. “Stress Tests and Information Disclosure.” *J. Econ. Theory* 177:34–69.

Gul, F., and W. Pesendorfer. 2010. “Strategic Redistricting.” *A.E.R.* 100 (4): 1616–41.

Guo, Y., and E. Shmaya. 2019. “The Interval Structure of Optimal Disclosure.” *Econometrica* 87 (2): 653–75.

Henry-Labordère, P., and N. Touzi. 2016. “An Explicit Martingale Version of the One-Dimensional Brenier Theorem.” *Finance and Stochastics* 20 (3): 635–68.

Holmström, B. 1979. “Moral Hazard and Observability.” *Bell J. Econ.* 10 (1): 74–91.

Hunt, B., T. Sauer, and J. Yorke. 1992. “Prevalence: A Translation-Invariant ‘Almost Every’ on Infinite-Dimensional Spaces.” *Bull. American Math. Soc.* 27 (2): 217–38.

Kamenica, E., and M. Gentzkow. 2011. “Bayesian Persuasion.” *A.E.R.* 101 (6): 2590–615.

Kleiner, A., B. Moldovanu, and P. Strack. 2021. “Extreme Points and Majorization: Economic Applications.” *Econometrica* 89 (4): 1557–93.

Kolotilin, A. 2018. “Optimal Information Disclosure: A Linear Programming Approach.” *Theoretical Econ.* 13 (2): 607–36.

Kolotilin, A., R. Corrao, and A. Wolitzky. 2022. “Persuasion with Non-Linear Preferences.” Arxiv working paper, https://arxiv.org/abs/2206.09164.

Kolotilin, A., T. Mylovanov, and A. Zapechelnyuk. 2022. “Censorship as Optimal Persuasion.” *Theoretical Econ.* 17 (2): 561–85.

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li. 2017. “Persuasion of a Privately Informed Receiver.” *Econometrica* 85 (6): 1949–64.

Kolotilin, A., and A. Wolitzky. 2024a. “The Economics of Partisan Gerrymandering.” SSRN working paper, https://ssrn.com/abstract=4998165.

———. 2024b. “Distributions of Posterior Quantiles via Matching.” *Theoretical Econ.* 19 (4): 1399–413.

Kramkov, D., and Y. Xu. 2022. “An Optimal Transport Problem with Backward Martingale Constraints Motivated by Insider Trading.” *Ann. Appl. Probability* 32 (1): 294–326.

Kremer, M., and E. Maskin. 1996. “Wage Inequality and Segregation by Skill.” Working Paper no. 5718, NBER, Cambridge, MA.

Legros, P., and A. F. Newman. 2002. “Monotone Matching in Perfect and Imperfect Worlds.” *Rev. Econ. Studies* 69 (4): 925–42.

Lin, X., and C. Liu. 2024. “Credible Persuasion.” *J.P.E.* 132 (4): 2228–73

Malamud, S., and A. Schrimpf. 2022. “Persuasion by Dimension Reduction.” Arxiv working paper, https://arxiv.org/abs/2110.08884.

Mirrlees, J. A. 1999. “The Theory of Moral Hazard and Unobservable Behaviour: Part I.” *Rev. Econ. Studies* 66 (1): 3–21.
Nikandrova, A., and R. Pancs. 2017. “Conjugate Information Disclosure in an Auction with Learning.” *J. Econ. Theory* 171:174–212.
Onuchic, P., and D. Ray. 2023. “Conveying Value via Categories.” *Theoretical Econ.* 18 (4): 1407–39.
Owen, G., and B. Grofman. 1988. “Optimal Partisan Gerrymandering.” *Polit. Geography Q.* 7 (1): 5–22.
Perez-Richet, E., and V. Skreta. 2025. “Falsification-Proof Non-Market Allocation Mechanisms.” CEPR working paper.
Quah, J., and B. Strulovici. 2012. “Aggregating the Single Crossing Property.” *Econometrica* 80 (5): 2333–48.
Rayo, L. 2013. “Monopolistic Signal Provision.” *BE J. Theoretical Econ.* 13 (1): 27–58.
Rayo, L., and I. Segal. 2010. “Optimal Information Disclosure.” *J.P.E.* 118 (5): 949–87.
Rochet, J.-C., and J.-L. Vila. 1994. “Insider Trading without Normality.” *Rev. Econ. Studies* 61 (1): 131–52.
Saint-Paul, G. 2001. “On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks.” *J.P.E.* 109 (1): 1–37.
Santambrogio, F. 2015. *Optimal Transport for Applied Mathematicians*, vol. 55. Birkäuser, NY: Springer.
Seidmann, D., and E. Winter. 1997. “Strategic Information Transmission with Verifiable Messages.” *Econometrica* 65 (1): 163–69.
Smolin, A., and T. Yamashita. 2023. “Information Design in Smooth Games.” Arxiv working paper, https://arxiv.org/abs/2202.10883.
Spence, M. 1973. “Job Market Signaling.” *Q.J.E.* 87 (3): 355–74.
Tamura, W. 2018. “Bayesian Persuasion with Quadratic Preferences.” SSRN working paper, https://ssrn.com/abstract=1987877.
Villani, C. 2009. *Optimal Transport: Old and New*, vol. 338. Berlin: Springer.
Yang, K. H., and A. Zentefis. 2024. “Monotone Function Intervals: Theory and Applications.” *A.E.R.* 114 (8): 2239–70.
Zeng, Y. 2023. “Derandomization of Persuasion Mechanisms.” *J. Econ. Theory* 212:105690.
Zhang, J., and J. Zhou. 2016. “Information Disclosure in Contests: A Bayesian Persuasion Approach.” *Econ. J.* 126 (597): 2197–217.