Econometrica Supplementary Material

SUPPLEMENT TO "STRONG DUALITY FOR A MULTIPLE-GOOD MONOPOLIST"

(*Econometrica*, Vol. 85, No. 3, May 2017, 735–767)

BY CONSTANTINOS DASKALAKIS, ALAN DECKELBAUM, AND CHRISTOS TZAMOS

APPENDIX A: STRONG MECHANISM DESIGN DUALITY—PROOF OF THEOREM 2

IN THIS SECTION, we give a formal proof of the strong mechanism duality theorem. To carefully prove the statement, we specify that the proof is for Radon measures. A Radon measure is a locally-finite inner-regular Borel measure. We use $\Gamma(X) = \text{Radon}(X)$ (resp. $\Gamma_+(X) = \text{Radon}_+(X)$) as the set of signed (resp. unsigned) Radon measures on X. The transformed measure of a distribution is always a signed Radon measure as it defines a bounded linear functional on the utility function u.[12]

A.1. *A Strong Duality Lemma*

The overall structure of our proof of Theorem 2 is roughly parallel to the proof of Monge–Kantorovich duality presented in Villani (2008), although the technical aspects of our proof are different, mainly due to the added convexity constraint on u. We begin by stating the Legendre–Fenchel transformation and the Fenchel–Rockafellar duality theorem.

DEFINITION 15—Legendre–Fenchel Transform: Let E be a normed vector space and let $\Lambda : E \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The *Legendre–Fenchel transform of* Λ, denoted Λ^*, is a map from the topological dual E^* of E to $\mathbb{R} \cup \{\infty\}$ given by

$$\Lambda^*(z^*) = \sup_{z \in E}\big(\langle z^*, z\rangle - \Lambda(z)\big).$$

CLAIM 3—Fenchel–Rockafellar Duality: *Let E be a normed vector space, E^* its topological dual, and Θ, Ξ two convex functions on E taking values in $\mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be the Legendre–Fenchel transforms of Θ and Ξ, respectively. Assume that there exists $z_0 \in E$ such that $\Theta(z_0) < +\infty$, $\Xi(z_0) < +\infty$ and Θ is continuous at z_0. Then*

$$\inf_{z \in E}\big[\Theta(z) + \Xi(z)\big] = \max_{z^* \in E^*}\big[-\Theta^*(-z^*) - \Xi^*(z^*)\big].$$

LEMMA 4: *Let X be a compact convex subset of $\mathbb{R}^n$, and let $\mu \in \Gamma(X)$ be such that $\mu(X) = 0$. Then*

$$\inf_{\substack{\gamma \in \Gamma_+(X \times X) \\ \gamma_1 \succeq_{\text{cvx}} \mu_+ \\ \gamma_2 \preceq_{\text{cvx}} \mu_-}} \int_{X \times X} \|x - y\|_1 \, d\gamma(x, y) = \sup_{\substack{\phi, \psi \in \mathcal{U}(X) \\ \phi(x) - \psi(y) \leq \|x - y\|_1}} \left(\int_X \phi \, d\mu_+ - \int_X \psi \, d\mu_-\right)$$

and the infimum on the left-hand side is achieved.

[12]More formally, this follows from the Riesz representation theorem.

 DOI: 10.3982/ECTA12618

PROOF: We will apply Fenchel–Rockafellar duality with $E = CB(X \times X)$, the space of continuous (and bounded) functions on $X \times X$ equipped with the $\|\cdot\|_\infty$ norm. Since X is compact, by the Riesz representation theorem $E^* = \Gamma(X \times X)$.

We now define functions Θ, Ξ mapping $CB(X \times X)$ to $\mathbb{R} \cup \{+\infty\}$ by

$$\Theta(f) = \begin{cases} 0, & \text{if } f(x, y) \geq -\|x - y\|_1 \text{ for all } x, y \in X, \\ +\infty, & \text{otherwise,} \end{cases}$$

$$\Xi(f) = \begin{cases} \int_X \psi \, d\mu_- - \int_X \phi \, d\mu_+, & \text{if } f(x, y) = \psi(y) - \phi(x) \text{ for some } \psi, \phi \in \mathcal{U}(X), \\ +\infty, & \text{otherwise.} \end{cases}$$

We note that Ξ is well-defined: If $\psi(x) - \phi(y) = \psi'(x) - \phi'(y)$ for all $x, y \in X$, then $\psi(x) - \psi'(x) = \phi(y) - \phi'(y)$ for all $x, y \in X$. This means that ψ' differs from ψ only by an additive constant, and ϕ differs from ϕ' by the same additive constant, and therefore (since μ_+ and μ_- have the same total mass), $\int_X \psi \, d\mu_- - \int_X \phi \, d\mu_+ = \int_X \psi' \, d\mu_- - \int_X \phi' \, d\mu_+$.

It is clear that $\Theta(f)$ is convex, since any convex combination of two functions for which $f(x, y) \geq -\|x - y\|_1$ will yield another function for which the inequality is satisfied. It is furthermore clear that Ξ is convex, since we can take convex combinations of the ψ and ϕ functions as appropriate. (Notice that $\mathcal{U}(X)$ is closed under addition and positive scaling of functions.)

Consider the function $z_0 \in CB(X \times X)$ which takes the constant value of 1. It is clear that $\Theta(z_0) = 0$ and $\Xi(z_0) = \mu_-(X) < \infty$. Furthermore, $\Theta(z) = 0$ for any $z \in CB(X \times X)$ with $\|z - z_0\|_\infty < 1$, and therefore Θ is continuous at z_0. We can thus apply the Fenchel–Rockafellar duality theorem.

We compute, for any $\gamma \in \Gamma(X \times X)$:

$$\begin{aligned} \Theta^*(-\gamma) &= \sup_{f \in CB(X \times X)} \left[\int_{X \times X} f(x, y) \, d\big(-\gamma(x, y)\big) - \begin{cases} 0, & \text{if } f(x, y) \geq -\|x - y\|_1 \forall x, y \in X, \\ +\infty, & \text{otherwise,} \end{cases} \right] \\ &= \sup_{\substack{f \in CB(X \times X) \\ f(x,y) \geq -\|x-y\|_1}} \left(-\int_{X \times X} f(x, y) \, d\gamma(x, y) \right) \\ &= \sup_{\substack{\tilde{f} \in CB(X \times X) \\ \tilde{f}(x,y) \leq \|x-y\|_1}} \left(\int_{X \times X} \tilde{f}(x, y) \, d\gamma(x, y) \right). \end{aligned}$$

We claim therefore that

$$\Theta^*(-\gamma) = \begin{cases} \int_{X \times X} \|x - y\|_1 \, d\gamma(x, y), & \text{if } \gamma \in \Gamma_+(X \times X), \\ \infty, & \text{otherwise.} \end{cases}$$

Indeed, if γ is a positive linear functional, then the result follows from monotonicity, since $\|x - y\|_1$ is the point-wise greatest function $\tilde{f}$ satisfying the constraint $\tilde{f}(x, y) \leq \|x - y\|_1$, and $\|x-y\|_1$ is continuous. Suppose instead that γ is a signed Radon measure which is not positive everywhere. Then there exists a continuous nonnegative function $g : X \times X \to \mathbb{R}$ such that $\int g\, d\gamma = -\varepsilon$ for some $\varepsilon > 0$.[13] Since $g(x, y) \geq 0$, it follows that $-kg(x, y) \leq 0 \leq \|x - y\|_1$ for any $k \geq 0$. Therefore,

$$\sup_{\substack{\tilde{f} \in CB(X \times X) \\ \tilde{f}(x,y) \leq \|x-y\|_1}} \left(\int_{X \times X} \tilde{f}(x, y)\, d\gamma(x, y) \right)$$
$$\geq \int -kg(x, y)\, d\gamma(x, y) = k\varepsilon.$$

The claim follows, since $k > 0$ is arbitrary.

We similarly compute, for any $\gamma \in \Gamma(X \times X)$:

$$\begin{aligned} \Xi^*(\gamma) = \sup_{f \in CB(X \times X)} \Bigg[& \int_{X \times X} f(x, y)\, d\gamma(x, y) \\ & - \begin{cases} \int_X \psi\, d\mu_- - \int_X \phi\, d\mu_+, & \text{if } f(x, y) = \psi(y) - \phi(x) \text{ and } \psi, \phi \in \mathcal{U}(X), \\ +\infty, & \text{otherwise,} \end{cases} \Bigg] \\ = \sup_{\psi, \phi \in \mathcal{U}(X)} & \left[\int_{X \times X} \big(\psi(y) - \phi(x)\big)\, d\gamma(x, y) - \int_X \psi\, d\mu_- + \int_X \phi\, d\mu_+ \right]. \end{aligned}$$

We notice that $\Xi^*(\gamma) \geq 0$ for all $\gamma \in \Gamma(X \times X)$ by setting $\psi = \phi = 0$ and thus $\Theta^*(-\gamma) + \Xi^*(\gamma) = \infty$ if $\gamma \notin \Gamma_+(X \times X)$. Moreover, when $\gamma \in \Gamma_+(X \times X)$,

$$\begin{aligned} \Xi^*(\gamma) &= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[\int_{X \times X} \big(\psi(y) - \phi(x)\big)\, d\gamma(x, y) - \int_X \psi\, d\mu_- + \int_X \phi\, d\mu_+ \right] \\ &= \sup_{\psi, \phi \in \mathcal{U}(X)} \left[\int_X \psi\, d(\gamma_2 - \mu_-) + \int_X \phi\, d(\mu_+ - \gamma_1) \right] \\ &= \begin{cases} 0, & \text{if } \gamma_1 \succeq_{\text{cvx}} \mu_+ \text{ and } \gamma_2 \preceq_{\text{cvx}} \mu_-, \\ \infty, & \text{otherwise.} \end{cases} \end{aligned}$$

The last equality is true because if $\gamma_1 \succeq_{\text{cvx}} \mu_+$ does not hold, we can find a function $\phi \in \mathcal{U}(X)$ such that $\int_X \phi\, d(\mu_+ - \gamma_1) > 0$. Since we are allowed to scale ϕ arbitrarily, we can make the inside quantity as large as we want. The same holds when $\mu_- \not\succeq_{\text{cvx}} \gamma_2$.

We now apply Fenchel–Rockafellar duality:

$$\inf_{f \in CB(X \times X)} \big[\Theta(f) + \Xi(f)\big] = \max_{\gamma \in \Gamma(X \times X)} \big[-\Theta^*(-\gamma) - \Xi^*(\gamma)\big],$$

$$\inf_{\substack{f(x,y) \geq -\|x-y\|_1 \\ f(x,y) = \psi(y) - \phi(x) \\ \psi, \phi \in \mathcal{U}(X)}} \left(\int_X \psi\, d\mu_- - \int_X \phi\, d\mu_+ \right) = \max_{\gamma \in \Gamma_+(X \times X)} \left[- \int_{X \times X} \|x - y\|_1\, d\gamma(x, y) - \Xi^*(\gamma) \right],$$

[13]Formally, we have used Lusin's theorem to find such a g which is continuous, as opposed to merely measurable.

$$\inf_{\substack{\psi,\phi\in\mathcal{U}(X)\\ \phi(x)-\psi(y)\le\|x-y\|_1}}\left(\int_X \psi\, d\mu_- - \int_X \phi\, d\mu_+\right) = \max_{\substack{\gamma\in\Gamma_+(X\times X)\\ \gamma_1\succeq_{\text{cvx}}\mu_+\\ \gamma_2\preceq_{\text{cvx}}\mu_-}}\left(-\int_{X\times X}\|x-y\|_1\, d\gamma(x,y)\right),$$

$$\sup_{\substack{\psi,\phi\in\mathcal{U}(X)\\ \phi(x)-\psi(y)\le\|x-y\|_1}}\left(\int_X \phi\, d\mu_+ - \int_X \psi\, d\mu_-\right) = \min_{\substack{\gamma\in\Gamma_+(X\times X)\\ \gamma_1\succeq_{\text{cvx}}\mu_+\\ \gamma_2\preceq_{\text{cvx}}\mu_-}}\left(\int_{X\times X}\|x-y\|_1\, d\gamma(x,y)\right).$$

Q.E.D.

A.2. *From Two Convex Functions to one*

LEMMA 5: *Let* $X=\prod_{i=1}^n[x_i^{\text{low}},x_i^{\text{high}}]$ *for some* $x_i^{\text{low}},x_i^{\text{high}}\ge 0$, *and let* $\mu\in\Gamma(X)$ *such that* $\mu(X)=0$. *Then*

$$\sup_{\substack{\phi,\psi\in\mathcal{U}(X)\\ \phi(x)-\psi(y)\le\|x-y\|_1}}\left(\int_X \phi\, d\mu_+ - \int_X \psi\, d\mu_-\right) = \sup_{u\in\mathcal{U}(X)\cap\mathcal{L}_1(X)}\left(\int_X u\, d\mu_+ - \int_X u\, d\mu_-\right).$$

Furthermore, if the supremum of one side is achieved, then so is the supremum of the other side.

PROOF: Given any feasible u for the right-hand side of Lemma 5, we observe that $\phi=\psi=u$ is feasible for the left-hand side, and therefore the left-hand side is at least as large as the right-hand side. It therefore suffices to prove the reverse direction of the inequality. Let ϕ and ψ be feasible for the left-hand side. Given ϕ, it is clear that ψ must satisfy $\psi(y)\ge\sup_x[\phi(x)-\|x-y\|_1]$.

Set $\bar\psi(y)=\sup_x[\phi(x)-\|x-y\|_1]$. Since ψ exists, this supremum indeed has finite value. Since $\bar\psi\le\psi$ point-wise, it follows that $\int_X\bar\psi\, d\mu_-\le\int_X\psi\, d\mu_-$. We must now prove that $\bar\psi\in\mathcal{U}(X)$, thereby showing that $\phi,\bar\psi$ is feasible for the left-hand side and that replacing ψ by $\bar\psi$ does not decrease the objective value.

CLAIM 4: $\bar\psi\in\mathcal{U}(X)$ *and* $\bar\psi\in\mathcal{L}_1(X)$.

PROOF: We will first show that $\bar\psi\in\mathcal{U}(X)$. We need to show continuity, monotonicity, and convexity.

- *Continuity.* Continuity of $\bar\psi$ follows from the Maximum Theorem since both ϕ and $\|\cdot\|_1$ are uniformly continuous.
- *Monotonicity.* Let $y\le y'$ coordinate-wise and let x be arbitrary. We must show that there exists an x' such that $\phi(x)-\|x-y\|_1\le\phi(x')-\|x'-y'\|_1$. Set $x_i'=\max\{x_i,y_i'\}$. Since $x\le x'$, we have $\phi(x)\le\phi(x')$. We notice that if $x_i\ge y_i'$, then $x_i'=x_i$ and thus $|x_i'-y_i'|\le|x_i-y_i|$, while if $x_i\le y_i'$, then $|x_i'-y_i'|=0$. Therefore, we have that $\|x-y\|_1\ge\|x'-y'\|_1$ and thus $\phi(x)-\|x-y\|_1\le\phi(x')-\|x'-y'\|_1$, as desired.
- *Convexity.* Let y,y',y'' be collinear points in X such that $y=\frac{y'+y''}{2}$. Then, given any x, we must show that there exist x' and x'' such that

$$\phi(x')-\|x'-y'\|_1+\phi(x'')-\|x''-y''\|_1\ge 2\phi(x)-2\|x-y\|_1.$$

We define x_i' and x_i'' as follows:

- If $y_i'\ge y_i''$, set $x_i'=\max\{x_i,y_i'\}$ and $x_i''=\max\{2x_i-x_i',y_i''\}$.
- If $y_i'<y_i''$, set $x_i''=\max\{x_i,y_i'\}$ and $x_i'=\max\{2x_i-x_i'',y_i'\}$.

Notice that $x' + x'' \geq 2x$, and thus (since ϕ is convex and monotone) we have $\phi(x') + \phi(x'') \geq 2\phi(x)$.

Suppose without loss of generality that $y'_i \geq y''_i$. We now consider two cases:

- $y'_i \geq x_i$. We then have $x'_i = y'_i$ and $x''_i = \max\{2x_i - y'_i, y''_i\}$. Therefore, $|y'_i - x'_i| = 0$ and $|y''_i - x''_i| \leq |y''_i - 2x_i + y'_i| = 2|y_i - x_i|$ since $y'_i + y''_i = 2y_i$.
- $y'_i < x_i$. We now have $x'_i = x_i$ and $x''_i = \max\{x_i, y''_i\} = x_i$. Therefore, $|y''_i - x''_i| + |y'_i - x'_i|$ is equal to $|y'_i + y''_i - 2x_i|$, which equals $|2y_i - 2x_i|$.

Therefore, we have that $|y'_i - x'_i| + |y''_i - x''_i| \leq |2y_i - 2x_i|$ for all i, which implies that $\|x' - y'\|_1 + \|x'' - y''\|_1 \leq 2\|x - y\|_1$.

We have thus shown that $\bar{\psi} \in \mathcal{U}(X)$. We will now show that $\bar{\psi} \in \mathcal{L}_1(X)$. We have

$$\begin{aligned}\bar{\psi}(x) - \bar{\psi}(y) &= \sup_z \inf_w \big(\phi(z) - \|z - x\|_1 - \phi(w) + \|w - y\|_1\big)\\ &\leq \sup_z \big(\phi(z) - \|z - x\|_1 - \phi(z) + \|z - y\|_1\big)\\ &= \sup_z \big(\|z - y\|_1 - \|z - x\|_1\big) \leq \|x - y\|_1.\end{aligned}$$

Q.E.D.

Since $\phi, \bar{\psi}$ are a feasible pair of functions for the left-hand side of Lemma 5, we know that ϕ satisfies the inequality $\phi(x) \leq \inf_y[\bar{\psi}(y) + \|x - y\|_1]$. We now set $\bar{\phi}(x) = \inf_y[\bar{\psi}(y) + \|x - y\|_1]$. It is clear that the value of the left-hand objective function under $\bar{\phi}, \bar{\psi}$ is at least as large as its value under $\phi, \bar{\psi}$.

We claim that not only is $\bar{\phi}$ continuous, monotonic, and convex, but in fact that $\bar{\phi} = \bar{\psi}$. We notice that $\bar{\phi}(x) \leq \bar{\psi}(x) + \|x - x\|_1 = \bar{\psi}(x)$. To prove the other direction of the inequality, we compute

$$\bar{\phi}(x) = \inf_y \big[\bar{\psi}(y) + \|x - y\|_1\big] = \bar{\psi}(x) + \inf_y \big[\bar{\psi}(y) - \bar{\psi}(x) + \|x - y\|_1\big] \geq \bar{\psi}(x),$$

where the last inequality holds since $\bar{\psi}(x) - \bar{\psi}(y) \leq \|x - y\|_1$. Therefore, $\bar{\phi} = \bar{\psi}$, and thus $\bar{\phi} \in \mathcal{U}(X)$. Since $\bar{\phi}$ satisfies the inequality $\bar{\phi}(x) - \bar{\phi}(y) \leq \|x - y\|_1$, it is feasible for the right-hand side of Lemma 5, and the value of the right-hand objective under $\bar{\phi}$ is at least as large as the value of the left-hand objective under ϕ, ψ. We notice finally that if ϕ, ψ are optimal for the left-hand side, then $\bar{\phi}$ is optimal for the right-hand side. *Q.E.D.*

A.3. *Proof of Theorem 2*

By combining Lemma 1, Lemma 4, and Lemma 5, we have

$$\begin{aligned}\inf_{\substack{\gamma\in\Gamma_+(X\times X)\\ \gamma_1-\gamma_2\succeq_1\mu}} \int_{X\times X} \|x - y\|_1 \, d\gamma &\geq \sup_{u\in\mathcal{U}(X)\cap\mathcal{L}_1(X)} \int_X u \, d\mu\\ &= \sup_{\substack{\phi,\psi\in\mathcal{U}(X)\\ \phi(x)-\psi(y)\leq\|x-y\|_1}} \left(\int_X \phi \, d\mu_+ - \int_X \psi \, d\mu_-\right)\\ &= \inf_{\substack{\gamma\in\Gamma_+(X\times X)\\ \gamma_1\succeq_{\mathrm{cvx}}\mu_+\\ \gamma_2\preceq_{\mathrm{cvx}}\mu_-}} \int_{X\times X} \|x - y\|_1 \, d\gamma(x, y).\end{aligned}$$

By Lemma 4, the last minimization problem above achieves its infimum for some γ^*. We notice that γ^* is also feasible for the first minimization problem above, and therefore the inequality is actually an equality and γ^* is optimal for the first minimization problem. In addition, since γ^* is feasible for the last minimization problem, it satisfies $\gamma_1^*(X) = \gamma_2^*(X) = \mu_+(X)$. All that remains is to prove that the supremum to the maximization problem is achieved for some u^*. A proof of this fact is in Appendix A.4.

A.4. *Existence of Optimal Mechanism*

We now prove that the supremum of the maximization problem of Theorem 2 is achieved for some u^*. Consider a sequence of feasible functions $u_1, u_2, \dots \in \mathcal{U}(X) \cap \mathcal{L}_1(X)$ such that $\int_X u_i \, d\mu$ converges monotonically to the supremum value V, which we have proven is finite.[14] Since $\mu(X) = 0$, we may without loss of generality assume that $u_i(0^n) = 0$ for all u_i. Since all of the functions are bounded by $\|x^{\text{high}}\|_1$ and are 1-Lipschitz (which implies equicontinuity), the Arzelà–Ascoli theorem implies that there exists a uniformly converging subsequence. Let u^* be the limit of that subsequence. Since the convergence is uniform, the function u^* is 1-Lipschitz, nondecreasing, and convex and thus feasible for the mechanism design problem. Moreover, since the objective is linear, the revenue of the mechanism with that utility is equal to V and thus the supremum is achieved.

A.5. *Omitted Proofs From Section 5—Example 2*

It is straightforward to verify that the mechanism is IC and IR. All that remains is to prove that the utility function u^* induced by the mechanism is optimal.

The transformed measure μ of the type distribution is composed of:

- a point mass of $+1$ at $(4, 4)$.
- mass -3 distributed throughout the rectangle (Density $-\frac{1}{12}$).
- mass $+\frac{7}{3}$ distributed on upper edge of rectangle (Linear density $+\frac{7}{36}$).
- mass $-\frac{4}{3}$ distributed on lower edge of rectangle (Linear density $-\frac{1}{9}$).
- mass $+\frac{4}{3}$ distributed on right edge of rectangle (Linear density $+\frac{4}{9}$).
- mass $-\frac{1}{3}$ distributed on left edge of rectangle (Linear density $-\frac{1}{9}$).

We claim that $\mu(Z) = \mu(Y) = \mu(W) = 0$, which is straightforward to verify.

We will construct an optimal γ^* for the dual program of Theorem 2, using the intuition of Remark 1. Our γ^* will be decomposed into $\gamma^* = \gamma^Z + \gamma^Y + \gamma^W$ with $\gamma^Z \in \Gamma_+(Z \times Z)$, $\gamma^Y \in \Gamma_+(Y \times Y)$, and $\gamma^W \in \Gamma_+(W \times W)$. To ensure that $\gamma_1^* - \gamma_2^* \succeq_{\text{cvx}} \mu$, we will show that

$$\gamma_1^Z - \gamma_2^Z \succeq_{\text{cvx}} \mu|_Z; \qquad \gamma_1^Y - \gamma_2^Y \succeq_{\text{cvx}} \mu|_Y; \qquad \gamma_1^W - \gamma_2^W \succeq_{\text{cvx}} \mu|_W.$$

We will also show that the conditions of Corollary 1 hold for each of the measures γ^Z, γ^Y, and γ^W separately, namely, $\int u^* \, d(\gamma_1^A - \gamma_2^A) = \int_A u^* \, d\mu$ and $u^*(x) - u^*(y) = \|x - y\|_1$ hold γ^A-almost surely for $A = Z, Y$, and W.

Construction of γ^Z. Since $\mu_+|_Z$ is a point mass at $(4, 4)$ and $\mu_-|_Z$ is distributed throughout a region which is coordinate-wise greater than $(4, 4)$, we notice that $\mu|_Z \preceq_{\text{cvx}} 0$. We therefore set γ^Z to be the zero measure, and the relation $\gamma_1^Z - \gamma_2^Z = 0 \succeq_{\text{cvx}} \mu|_Z$, as well as the two necessary equalities from Corollary 1, are trivially satisfied.

Construction of γ^W. We will construct $\gamma^W \in \Gamma(\mu_+|_W, \mu_-|_W)$ such that $x \geq y$ component-wise holds $\gamma^W(x, y)$ almost surely. Geometrically, we view this as "transporting" $\mu_+|_W$ into $\mu_-|_W$ by moving mass downwards and leftwards. Indeed, since both items are allocated

[14]Finiteness is also obvious because X is bounded and the infimum problem is feasible.

with probability 1 in W, being able to transport both downwards and leftwards is in line with our interpretation of the second condition of Corollary 1, as explained in Remark 1.[15]

We notice that $\mu_+|_W$ consists of mass distributed on the top and right edges of W, while $\mu_-|_W$ consists of mass on the interior and bottom of W. We first match the μ_+ mass on $[8, 16] \times \{7\}$ with the μ_- mass on $[8, 16] \times [\frac{14}{3}, 7]$ by moving mass downwards, then we match the μ_+ mass on $\{16\} \times [4, \frac{14}{3}]$ with the μ_- mass on $[\frac{32}{3}, 16] \times (4, \frac{14}{3}]$ by moving mass to the left, and we finally match the μ_+ mass on $\{16\} \times [\frac{14}{3}, 7]$ with the remaining negative mass arbitrarily. Noticing that $u^*(x) = \|x\|_1 - 12$ for all $x \in W$, it is straightforward to verify the desired properties from Corollary 1.

Construction of γ^Y. This is the most involved step of the proof. Since item 2 is allocated with 100% probability in region Y, by Remark 1 we would like to transport the positive mass $\mu_+|_Y$ into $\mu_-|_Y$ by moving mass straight downwards. However, this is impossible without first "shuffling" $\mu|_Y$, due to the negative mass on the left boundary of Y. Therefore, we first "shuffle" the positive part of $\mu|_Y$ (on the top boundary) to push positive mass onto the point $(4, 7)$ (the top-left corner of Y), and only then do we transport the positive part of the shuffled measure into the negative part by sending mass downwards. Since the positive and negative parts of $\mu|_Y$ must be matchable by only sending mass downwards, we know how the post-shuffling measure should look. In particular, on every vertical line in region Y the net post-shuffling mass should be zero.

So rather than constructing γ^Y with $\gamma_1^Y - \gamma_2^Y$ equal to $\mu|_Y$, we will have $\gamma_1^Y - \gamma_2^Y = \mu|_Y + \alpha$, where the "shuffling" measure $\alpha = \alpha_+ - \alpha_- \succeq_{\text{cvx}} 0$. As discussed above, we set α to have density function

$$f_\alpha(z_1, z_2) = \mathbb{I}_{z_2=7} \cdot \left(\frac{1}{9}\mathbb{I}_{z_1=4} + \frac{1}{24}\left(z_1 - \frac{20}{3}\right)\right) \cdot \mathbb{I}_{z\in Y}.$$

The measure α is supported on the line $[4, 8] \times \{7\}$ and consists of a point mass of $\frac{1}{9}$ at $(4, 7)$ followed by allocating mass along the one-dimensional upper boundary of Y according to a density function which begins negative and increases linearly. It is straightforward to verify that $\alpha \succeq_{\text{cvx}} 0$,[16] which we need for feasibility, and that $\int_Y u^* \, d\alpha = 0$, which we need to satisfy complementary slackness.

We are now ready to define $\gamma^Y \in \Gamma(\mu_+|_Y + \alpha_+, \mu_-|_Y + \alpha_-)$. We construct γ^Y so that $x_1 = y_1$ and $x_2 \geq y_2$ hold $\gamma^Y(x, y)$ almost surely. Since $\mu_+|_Y + \alpha_+$ only assigns mass to the upper boundary of Y, to show that γ^Y can be constructed so that all mass is transported "vertically downwards" we need only verify that $\mu_+|_Y + \alpha_+$ and $\mu_-|_Y + \alpha_-$ assign the same density to any vertical "strip" in Y. Indeed,

$$\begin{aligned}(\mu_-|_Y + \alpha_-)\big(\{4\} \times [6, 7]\big) &= \mu_-|_Y\big(\{4\} \times [6, 7]\big) \\ &= \frac{1}{9} = \alpha_+\big(\{4\} \times [6, 7]\big) = (\mu_+|_Y + \alpha_+)\big(\{4\} \times [6, 7]\big)\end{aligned}$$

[15]To prove the existence of such a map, it is equivalent by Strassen's theorem to prove that $\mu_+|_W$ stochastically dominates $\mu_-|_W$ in the first order, but in this example we will directly define such a map.

[16]Since α is supported on a one-dimensional line, this verification uses a property analogous to the standard characterization of one-dimensional second-order stochastic dominance via the cumulative density function. Informally, we can argue that $\alpha \succeq_{\text{cvx}} 0$ by considering integrals of one-dimensional test functions (by restricting our attention to the line $z_2 = 7$) and noticing that, since $\alpha(Y) = 0$, we need only consider test functions h which have value 0 at $z_1 = 4$. We then use the fact that all linear functions integrate to 0 under α and that (ignoring the point mass at $z_1 = 4$, since h is 0 at this point) the density of α is monotonically increasing.

and, for all $z_1 \pm \varepsilon \in (4, 8]$, we compute the following, using the fact that the surface area of $Y \cap ([z_1 - \varepsilon, z_1 + \varepsilon] \times [4, 7])$ is $2\varepsilon \cdot (\frac{z_1}{2} - 1)$:

$$\begin{aligned}&(\mu_-|_Y - \alpha|_Y)([z_1 - \varepsilon, z_1 + \varepsilon] \times [4, 7])\\ &\quad = \frac{1}{12} \cdot \left(2\varepsilon \cdot \left(\frac{z_1}{2} - 1\right)\right) - \frac{1}{24} \int_{z_1-\varepsilon}^{z_1+\varepsilon} \left(z - \frac{20}{3}\right) dz\\ &\quad = \frac{\varepsilon z_1}{12} - \frac{\varepsilon}{6} - \frac{1}{24}\left(2\varepsilon z_1 - \frac{40\varepsilon}{3}\right) = \frac{7\varepsilon}{18} = \mu_+|_Y([z_1 - \varepsilon, z_1 + \varepsilon] \times [4, 7]).\end{aligned}$$

Since u^* has the property that $u^*(z_1, a) - u^*(z_1, b) = a - b$ for all $(z_1, a), (z_1, b) \in Y$ (as the second good is received with probability 1), it follows that γ^Y satisfies the necessary conditions of Corollary 1.

APPENDIX B: PROOF OF STOCHASTIC CONDITIONS OF SECTION 6

Our goal in this section is to prove Theorem 3. We begin by presenting some useful probabilistic tools that will be essential for the proof.

B.1. *Probabilistic Lemmas*

We first present a useful result about convex dominance of random variables. For more information about this result, see Theorem 7.A.2 of Shaked and Shanthikumar (2010).

LEMMA 6—Strassen's Theorem: *Let A and B be random vectors. Then $A \preceq_{\text{cvx}} B$ if and only if there exist random vectors $\hat{A}$ and $\hat{B}$, defined on the same probability space, such that $\hat{A} =_{\text{st}} A$, $\hat{B} =_{\text{st}} B$, and $\mathbb{E}[\hat{B}|\hat{A}] \geq \hat{A}$ almost surely, where the final inequality is componentwise and where $=_{\text{st}}$ denotes equality in distribution.*

It is easy to extend the above result to convex dominance with respect to a vector $\vec{v}$ as defined in Definition 8.

LEMMA 7—Extended Strassen's Theorem: *Let A and B be random vectors. Then $A \preceq_{\text{cvx}(\vec{v})} B$ if and only if there exist random vectors $\hat{A}$ and $\hat{B}$, defined on the same probability space, with $\hat{A} =_{\text{st}} A$, $\hat{B} =_{\text{st}} B$, such that (almost surely):*

- *if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i$.*
- *if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i$.*
- *if $v_i = -1$, then $E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$.*

We now state a multivariate variant of Jensen's inequality along with the necessary condition for equality to hold. The proof of this result is standard and straightforward, and thus is omitted.

LEMMA 8—Jensen's Inequality: *Let V be a vector-valued random variable with values in $[0, M]^n$ and let u be a convex Lipschitz-continuous function mapping $[0, M]^n \to \mathbb{R}$. Then $\mathbb{E}[u(V)] \geq u(\mathbb{E}[V])$. Furthermore, equality holds if and only if, for every a in the subdifferential of u at $\mathbb{E}[V]$, the equality $u(V) = a \cdot (V - \mathbb{E}[V]) + u(\mathbb{E}[V])$ holds almost surely.*

The following lemma is a conditional variant of Lemma 8, based on the multivariate conditional Jensen's inequality, as in Theorem 10.2.7 of Dudley (2002). This lemma is used as a tool for Lemma 10, the main result of this subsection.

LEMMA 9: *Let $(\Omega, \mathcal{A}, P)$ be a probability space, V be a random variable on Ω with values in X where $X = \prod_{i=1}^n [x_i^{\text{low}}, x_i^{\text{high}}]$, and $u : X \to \mathbb{R}$ be convex and Lipschitz continuous. Let $\mathcal{C}$ be any sub-σ-algebra of $\mathcal{A}$ and suppose that $\mathbb{E}[u(V)|\mathcal{C}] = u(\mathbb{E}[V|\mathcal{C}])$ almost surely. Then, for almost all $x \in \Omega$, the equality $u(y) = a_{y_x} \cdot (y - y_x) + u(y_x)$ holds almost surely with respect to the law* [17] *$P_{V|\mathcal{C}}(\cdot, x)$, where y_x is the expectation of the random variable with law $P_{V|\mathcal{C}}(\cdot, x)$ and a_{y_x} is any subgradient of u at y_x.*

PROOF: The proof is based on the proof of the multivariate conditional Jensen's inequality, as in Theorem 10.2.7 of Dudley (2002). This theorem requires $|V|$ and $u \circ V$ to be integrable, which is true in our setting. We note that the theorem applies when u is defined in an open convex set, but because u is Lipschitz continuous, we can extend it to a function with domain an open set containing X. The multivariate conditional Jensen's inequality states that, almost surely, $\mathbb{E}[V|\mathcal{C}] \in \mathcal{C}$ and $\mathbb{E}[u(V)|\mathcal{C}] \geq u(\mathbb{E}[V|\mathcal{C}])$. The proof of Theorem 10.2.7 in Dudley (2002) furthermore shows that the following two equalities hold:

$$\mathbb{E}[V|\mathcal{C}](x) = \int_X y P_{V|\mathcal{C}}(dy, x); \qquad \mathbb{E}\big[u(V)|\mathcal{C}\big](x) = \int_X u(y) P_{V|\mathcal{C}}(dy, x).$$

Since $\mathbb{E}[u(V)|\mathcal{C}](x) = u(\mathbb{E}[V|\mathcal{C}])(x)$ for almost all x, we apply the unconditional Jensen inequality (Lemma 8) to the laws $P_{V|\mathcal{C}}(\cdot, x)$ to prove the lemma. *Q.E.D.*

We now present Lemma 10. This lemma states that for random variables A and B with $A \preceq_{\text{cvx}} B$, if it holds that $u(A) = u(B)$ for some convex function u, then there exists a coupling between A and B with several desirable properties, including that points are only matched if u shares a subgradient at these points.

LEMMA 10: *Let A and B be vector random variables with values in X, where $X = \prod_{i=1}^n [x_i^{\text{low}}, x_i^{\text{high}}]$, such that $A \preceq_{\text{cvx}} B$. Let $u : X \to \mathbb{R}$ be 1-Lipschitz with respect to the ℓ_1 norm, convex, and monotonically nondecreasing. Suppose that $\mathbb{E}[u(A)] = \mathbb{E}[u(B)]$ and that $g : X \to [0, 1]^n$ is a measurable function such that, for all $z \in X$, $g(z)$ is a subgradient of u at z.*

Then there exist random variables $\hat{A} =_{\text{st}} A$ and $\hat{B} =_{\text{st}} B$ such that, almost surely:

- $u(\hat{B}) = u(\hat{A}) + g(\hat{A}) \cdot (\hat{B} - \hat{A})$.
- *$g(\hat{A})$ is a subgradient of u at $\hat{B}$.*
- *$\mathbb{E}[\hat{B}|\hat{A}]$ is component-wise greater than or equal to $\hat{A}$.*
- $u(\mathbb{E}[\hat{B}|\hat{A}]) = u(\hat{A})$.

PROOF: By Lemma 6, there exist random variables $\hat{A} =_{\text{st}} A$ and $\hat{B} =_{\text{st}} B$ such that $\mathbb{E}[\hat{B}|\hat{A}]$ is component-wise greater than or equal to $\hat{A}$ almost surely. We have

$$0 = \mathbb{E}\big[u(\hat{B}) - u(\hat{A})\big] \geq \mathbb{E}\big[u(\hat{B}) - u\big(\mathbb{E}[\hat{B}|\hat{A}]\big)\big] = \mathbb{E}\big[\mathbb{E}\big[u(\hat{B})|\hat{A}\big] - u\big(\mathbb{E}[\hat{B}|\hat{A}]\big)\big] \geq 0$$

and therefore $\mathbb{E}[\mathbb{E}[u(\hat{B})|\hat{A}]] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])] = \mathbb{E}[u(\hat{B})] = \mathbb{E}[u(\hat{A})]$.

[17] The law $P_{V|\mathcal{C}}(\cdot, x)$ allows us to express the conditional distribution of V given $\mathcal{C}$.

Since u is monotonic, $u(\hat{A}) \leq u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely. Since $\mathbb{E}[u(\hat{A})] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])]$, it follows that $u(\hat{A}) = u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely.

Select any collection of random variables $\{\hat{B}|_{\hat{A}=x}\}$ corresponding to the laws $P_{\hat{B}|\hat{A}}(\cdot, x)$. For almost all values x of $\hat{A}$, $\mathbb{E}[\hat{B}|_{\hat{A}=x}]$ is component-wise greater than x and $u(x) = u(\mathbb{E}[\hat{B}|_{\hat{A}=x}])$. We claim now that any subgradient a_x of u at x is also a subgradient of u at $\mathbb{E}[\hat{B}|_{\hat{A}=x}]$. Indeed, choose such a subgradient a_x. We compute

$$u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) \geq u(x) + a_x \cdot \big(\mathbb{E}[\hat{B}|_{\hat{A}=x}] - x\big) = u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) + a_x \cdot \big(\mathbb{E}[\hat{B}|_{\hat{A}=x}] - x\big)$$

and therefore $a_x \cdot \mathbb{E}[\hat{B}|_{\hat{A}=x}] = a_x \cdot x$, by nonnegativity of the subgradient. Furthermore, for any point $z \in X$,

$$\begin{aligned} u(z) &\geq u(x) + a_x \cdot (z - x) \\ &= u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) + a_x \cdot (z - x) \\ &= u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) + a_x \cdot \big(z - \mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) \end{aligned}$$

and thus a_x is a subgradient of u at $\mathbb{E}[\hat{B}|_{\hat{A}=x}]$.

Since $\mathbb{E}[\mathbb{E}[u(\hat{B})|\hat{A}]] = \mathbb{E}[u(\mathbb{E}[\hat{B}|\hat{A}])]$, by Jensen's inequality it follows that $\mathbb{E}[u(\hat{B})|\hat{A}] = u(\mathbb{E}[\hat{B}|\hat{A}])$ almost surely. By Lemma 9, it therefore holds for almost all values x of $\hat{A}$ that the equality

$$\begin{aligned} u(y) &= a_x \cdot \big(y - \mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) + u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) \\ &= a_x \cdot (y - x) + u\big(\mathbb{E}[\hat{B}|_{\hat{A}=x}]\big) \\ &= a_x \cdot (y - x) + u(x) \end{aligned}$$

holds $\hat{B}|_{\hat{A}=x}$ almost surely.

Last, we will show that, almost surely, a_x is a subgradient of u at $\hat{B}|_{\hat{A}=x}$. Indeed, for any $p \in X$, and almost all values of x, we have

$$\begin{aligned} u(p) &\geq u(x) + a_x \cdot (p - x) \\ &= u(x) + a_x \cdot (\hat{B}|_{\hat{A}=x} - x) + a_x \cdot (p - \hat{B}|_{\hat{A}=x}) \\ &= u(\hat{B}|_{\hat{A}=x}) + a_x \cdot (p - \hat{B}|_{\hat{A}=x}). \end{aligned}$$

Q.E.D.

B.2. *Proof of the Optimal Menu Theorem (Theorem 3)*

To prove the equivalence, we prove both implications of the theorem separately.

B.2.1. *Sufficiency Conditions*

We will show that the Optimal Menu Conditions of Definition 9 imply that a mechanism $\mathcal{M}$ is optimal. To show the theorem, we construct a measure γ such that the conditions of Corollary 1 are satisfied. We will construct this measure separately for every region that corresponds to a menu choice of mechanism $\mathcal{M}$.

Consider a menu choice $(p, t) \in \text{Menu}_{\mathcal{M}}$, the corresponding region R and the corresponding vector $\vec{v}$ as in Definition 9. Let A and B be random vectors distributed according to the (normalized) measures $\mu_+|R$ and $\mu_-|R$. From the Optimal Menu Conditions, we have that $A|_R \preceq_{\text{cvx}(\vec{v})} B|_R$ (almost surely). By the extended version of Strassen's theorem (Lemma 7), it holds that there exist random vectors $\hat{A}, \hat{B}$ with $\hat{A} =_{\text{st}} A|_R$ and $\hat{B} =_{\text{st}} B|_R$, such that (almost surely):

- if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i$.
- if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i$.
- if $v_i = -1$, then $E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$.

Now define the random variable $\hat{C} = \min(E[\hat{B}|\hat{A}], \hat{A})$, where we take the coordinate-wise minimum. We now have that (almost surely):

- if $v_i = +1$, then $E[\hat{B}_i|\hat{A}] \geq \hat{A}_i = \hat{C}_i$.
- if $v_i = 0$, then $E[\hat{B}_i|\hat{A}] = \hat{A}_i = \hat{C}_i$.
- if $v_i = -1$, then $\hat{C}_i = E[\hat{B}_i|\hat{A}] \leq \hat{A}_i$.

Let γ_R be the measure according to which the vector $(\hat{A}, \hat{C})$ is distributed. By construction, $\gamma_{R1} = \mu_+|_R$ and $\gamma_{R2} \preceq_{\text{cvx}} \mu_-|_R$, and thus $\gamma_{R1} - \gamma_{R2} \succeq_{\text{cvx}} \mu|_R$. Moreover, the conditions of Corollary 1 are satisfied:

- $u(x) - u(y) = \|x - y\|_1$ is satisfied $\gamma_R(x, y)$-almost surely since $\hat{A}$ is larger than $\hat{C}$ only in coordinates for which $v_i = -1$ and thus $p_i = 1$.
- $\int u\, d(\gamma_{R1} - \gamma_{R2}) = \int u\, d(\mu_+|_R - \mu_-|_R)$ is satisfied: By definition, we have that $\int u\, d\gamma_{R1} = \int u\, d\mu_+|_R$. Moreover, we can also show that $\int u\, d\gamma_{R2} = \int u\, d\mu_-|_R$ by noting that $\int u\, d\mu_-|_R = \mu_-(R)E[u(\hat{B})] = \mu_-(R)E[p \cdot \hat{B} - t] = \mu_-(R)E[p \cdot E[\hat{B}|\hat{A}] - t]$ and that $\mu_-(R)E[p \cdot E[\hat{B}|\hat{A}] - t]$ is equal to $\mu_-(R)E[p \cdot \hat{C} - t] = \int u\, d\gamma_{R2}$ since $\hat{C}_i \neq E[\hat{B}_i|\hat{A}]$ only when $E[\hat{B}_i|\hat{A}]$ is strictly larger than $\hat{A}_i$, which only happens in coordinates i where $v_i = +1$ and thus $p_i = 0$.

This completes the proof that the Optimal Menu Conditions imply optimality of the mechanism since we can construct a feasible measure γ satisfying the conditions of Corollary 1 by considering the sum of the constructed measures for each region.

B.2.2. *Optimality Implies Stochastic Conditions*

We will now prove the other direction of the result. Consider an optimal mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ with a finite menu size over type space $X = \prod_{i=1}^{n} [x_i^{\text{low}}, x_i^{\text{high}}]$. Since $\mathcal{M}$ is given in essential form, in the menu of $\mathcal{M}$ there is no dominated option. So for all options on the menu, there is a set of buyer types that strictly prefer it from any other option, and that set of types occurs with positive probability.

Now, define the set $Z = \{x \in X : p \cdot x - t = \mathcal{P}(x) \cdot x - \mathcal{T}(x)$ for $(p, t) \in \text{Menu}_{\mathcal{M}}$ with $(p, t) \neq (\mathcal{P}(x), \mathcal{T}(x))\}$. This is the set of types where there is no single option that is the best and it is where the utility function of the mechanism is not differentiable. We show the following lemma.

LEMMA 11: $\mu_-(Z) = 0$.

PROOF: Note that, by its construction, μ_- assigns zero mass to any k-dimensional surface for $k \leq n - 2$. Moreover, it only assigns mass to $(n - 1)$-dimensional surfaces which lie along the boundary of X.

Every pair of distinct choices $(p, t), (p', t') \in \text{Menu}_{\mathcal{M}}$ defines a hyperplane $p \cdot x - t = p' \cdot x - t'$ containing the types who derive the same utility from these two choices. As the menu is finite, there exist a finite number of such pairs, hence a finite number of hyperplanes. The set Z contains a subset of types in the finite union of these hyperplanes, so μ_- assigns no mass to the subset of Z which lies on the interior of X.

Regarding the μ_--measure of Z on the boundaries, notice that the intersection of each of the aforementioned hyperplanes $p \cdot x - t = p' \cdot x - t'$ with each boundary $x_i = x_i^{\text{low}}$ is $(n-2)$-dimensional, unless the hyperplane coincides with $x_i = x_i^{\text{low}}$. If it is $(n-2)$-dimensional, then its measure under μ_- is 0. Otherwise, it must be that $p_j = p'_j$, for all $j \neq i$, and $p_i \neq p'_i$; say $p_i > p'_i$ without loss of generality. This implies that (p, t) must dominate (p', t'), for all types $x \in X$. This contradicts our assumption that no menu choices are dominated. *Q.E.D.*

Let u be the utility function of the optimal mechanism $\mathcal{M} = (\mathcal{P}, \mathcal{T})$ and γ be the optimal measure of Theorem 2. Then, γ satisfies the properties of Corollary 1. In particular, it holds that:

1.

$$\int u\, d(\gamma_1 + \mu_-) = \int u\, d(\mu_+ + \gamma_2). \tag{6}$$

2. $u(x) - u(y) = \|x - y\|_1$, $\gamma(x, y)$ almost surely. Since this can happen only if x is coordinate-wise greater than y, it holds (almost surely with respect to γ) that $\|x - y\|_1 = \sum_i x_i - \sum_i y_i$, which implies that (almost surely) $u(x) - \sum_i x_i = u(y) - \sum_i y_i$ and thus

$$\int \left(u(x) - \sum_i x_i \right) d\gamma_1 = \int \left(u(y) - \sum_i y_i \right) d\gamma_2. \tag{7}$$

Moreover, again since x is coordinate-wise greater than y almost surely with respect to γ, it follows that $\gamma_2 \succeq_{\text{cvx}(-\vec{1})} \gamma_1$.

We are now ready to use Lemma 10, which follows from Jensen's inequality. We will apply it in two different steps, which we will then combine to show that $\mu_+|_R \preceq_{\text{cvx}(\vec{v})} \mu_-|_R$.

Step (ia): We will first apply Lemma 10 to random variables A, B distributed according to the measures $\gamma_2 + \mu_+$ and $\gamma_1 + \mu_-$, respectively. Since $\mu_+ - \mu_- \preceq_{\text{cvx}} \gamma_1 - \gamma_2$, by the feasibility of γ, we have that $A \preceq_{\text{cvx}} B$. Moreover, $\mathbb{E}[u(A)] = \mathbb{E}[u(B)]$, from Equation (6) above, and u is convex and nondecreasing, from the feasibility of u.

To apply Lemma 10, we choose the function $g(x)$, which is a subgradient function of u, as follows:

- For all $x \in X \setminus Z$, the best choice from the menu of $\mathcal{M}$ is unique, hence the subgradient of u is uniquely defined. For all such x, we set $g(x) = \mathcal{P}(x)$.
- For all other x, u has a continuum of different subgradients at x. In particular, any vector in the convex hull of $\{p : p \cdot x - t = u(x), (p, t) \in \text{Menu}_{\mathcal{M}}\}$ is a valid subgradient. Thus, we can always choose $g(x)$ to equal a vector of probabilities that does not appear as an allocation of any choice in menu $\mathcal{M}$.

Step (ib): It follows from Lemma 10 that there exist random variables $\hat{A} =_{\text{st}} A$ and $\hat{B} =_{\text{st}} B$ such that, almost surely, $g(\hat{A})$ is a subgradient of u at $\hat{B}$. Fixing some $(p, t) \in \text{Menu}_{\mathcal{M}}$ and its corresponding region $R = \{x : p = \mathcal{P}(x)\}$, we denote by $\text{cl}(R) = R \cup \partial R$ the closure of R and by $\text{int}(R) = \text{cl}(R) \setminus Z$ the set of types which strictly prefer (p, t) to any other option in the menu. Note in particular that $\text{int}(R)$ may contain points on the boundary of X. With this notation, we have that, almost surely,

$$\hat{B} \in \text{int}(R) \implies \hat{A} \in \text{int}(R); \tag{8}$$

$$\hat{A} \in \text{int}(R) \implies \hat{B} \in \text{cl}(R). \tag{9}$$

This is because, from Lemma 10, we know that $g(\hat{A})$ is a subgradient of u at $\hat{B}$ almost surely, and we know by definition of $\text{int}(R)$ that the subgradient is unique whenever $\hat{B} \in \text{int}(R)$. Thus, it holds almost surely that whenever $\hat{B} \in \text{int}(R)$, we have $g(\hat{A}) = g(\hat{B})$. Since g is chosen to have differing values on $\text{int}(R)$ and on Z, it follows that whenever $\hat{B} \in \text{int}(R)$, $\hat{A} \in \text{int}(R)$ almost surely. The implication $\hat{A} \in \text{int}(R) \implies \hat{B} \in \text{cl}(R)$ follows from the fact that the subgradient at any point $x \in \text{int}(R)$ can only serve as a subgradient for points $y \in \text{cl}(R)$.

From Lemma 10, we also have that $u(\mathbb{E}[\hat{B}|\hat{A}]) = u(\hat{A})$ almost surely. It follows that, almost surely,

$$u\big(\mathbb{E}[\hat{B}|\hat{A}]\big) \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} = u(\hat{A}) \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)}.$$

Given (9) and since u is linear restricted to $\text{cl}(R)$, it follows that the left-hand side equals

$$\mathbb{E}\big[u(\hat{B})|\hat{A}\big] \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)}.$$

We also have from Lemma 10 that, almost surely, it holds component-wise

$$\mathbb{E}[\hat{B}|\hat{A}] \geq \hat{A}. \tag{10}$$

The above imply that, almost surely,

$$p_i > 0 \quad \Longrightarrow \quad \mathbb{E}[\hat{B}_i|\hat{A}] \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} = \hat{A}_i \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)}, \tag{11}$$

as otherwise we cannot have $\mathbb{E}[u(\hat{B})|\hat{A}] \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} = u(\hat{A}) \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)}$, given that u is linear and nondecreasing in $\text{cl}(R)$.

Equations (10), (11) and Lemma 7 imply that

$$\hat{A} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \hat{B} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} \tag{12}$$

for the $\vec{v}$ defined in Definition 9 for the menu choice (p, t). Note that

$$\begin{aligned}\hat{B} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)} &= \hat{B} \cdot \mathbb{I}_{\hat{A},\hat{B}\in\text{int}(R)} + \hat{B} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)\wedge\hat{B}\notin\text{int}(R)} \\ &= \hat{B} \cdot \mathbb{I}_{\hat{B}\in\text{int}(R)} + \hat{B} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)\wedge\hat{B}\notin\text{int}(R)},\end{aligned}$$

where, for the second equality, we used (8). Hence, (12) implies

$$\gamma_2|_{\text{int}(R)} + \mu_+|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \mu_-|_{\text{int}(R)} + \gamma_1|_{\text{int}(R)} + \xi_R, \tag{13}$$

where ξ_R is the nonnegative measure corresponding to $\hat{B} \cdot \mathbb{I}_{\hat{A}\in\text{int}(R)\wedge\hat{B}\notin\text{int}(R)}$ (scaled back appropriately by $\mu_+(X) = \mu_-(X)$).

Step (iia): We will now apply a flipped version of Lemma 10, for convex non-increasing functions,[18] to the convex function $u(x) - \sum_i x_i$.[19] We set random variables A', B' distributed according to the measures γ_1 and γ_2. Since $\gamma_2 \succeq_{\text{cvx}(-\vec{1})} \gamma_1$, we have that $B' \succeq_{\text{cvx}(-\vec{1})} A'$. Moreover, $\mathbb{E}[u(A') - \sum_i A'_i] = \mathbb{E}[u(B') - \sum_i B'_i]$ from Equation (7) shown above.

[18]It is easy to verify that the guarantees of the lemma remain the same except the third guarantee changes to “component-wise smaller than.”

[19]Notice that the partial derivatives are non-positive.

We choose the function $g(x) - \vec{1}$ as the subgradient of $u(x) - \sum_i x_i$.

Step (iib): Fixing any region R and the corresponding $\text{int}(R)$, $\text{cl}(R)$, and $\vec{v}$ as above, we mirror the arguments of Step (i). Now, the version of Lemma 10 for non-increasing functions implies that there exist random variables $\hat{A}' =_{\text{st}} A'$ and $\hat{B}' =_{\text{st}} B'$ such that, almost surely,

$$\mathbb{E}[\hat{B}'|\hat{A}'] \leq \hat{A}'; \tag{14}$$

$$p_i < 1 \quad \Longrightarrow \quad \mathbb{E}[\hat{B}'_i|\hat{A}'] \cdot \mathbb{I}_{\hat{A}'\in\text{int}(R)} = \hat{A}'_i \cdot \mathbb{I}_{\hat{A}'\in\text{int}(R)}. \tag{15}$$

Equations (14), (15) and Lemma 7 imply that

$$\hat{A}' \cdot \mathbb{I}_{\hat{A}'\in\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \hat{B}' \cdot \mathbb{I}_{\hat{A}'\in\text{int}(R)} \tag{16}$$

and, hence,

$$\gamma_1|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \gamma_2|_{\text{int}(R)} + \xi'_R, \tag{17}$$

where, similarly to our derivation above, ξ'_R is the nonnegative measure corresponding to $\hat{B}' \cdot \mathbb{I}_{\hat{A}'\in\text{int}(R)\wedge\hat{B}'\notin\text{int}(R)}$.

We now combine the results of Steps (i) and (ii) to finish the proof. Combining (13) and (17), we get that

$$\mu_+|_{\text{int}(R)} \preceq_{\text{cvx}(\vec{v})} \mu_-|_{\text{int}(R)} + \xi_R + \xi'_R. \tag{18}$$

From Proposition 1, it must hold that

$$\mu_+|_{\text{int}(R)}(X) = \mu_-|_{\text{int}(R)}(X) + \xi_R(X) + \xi'_R(X).$$

Summing over all regions and noticing that $\sum_R \mu_-|_{\text{int}(R)}(X) = \mu_-(X)$, from Lemma 11, we get that

$$\mu_+(X) - \mu_+(Z) = \mu_-(X) + \sum_R \big(\xi_R(X) + \xi'_R(X)\big).$$

But $\mu_+(X) = \mu_-(X)$, hence $\mu_+(Z) = \sum_R(\xi_R(X) + \xi'_R(X)) = 0$, as all of μ_+, ξ_R, and ξ'_R are nonnegative. Therefore, we can rewrite the property (18) as

$$\mu_+|_R \preceq_{\text{cvx}(\vec{v})} \mu_-|_R.$$

APPENDIX C: MISSING PROOFS OF SECTION 6—THEOREM 5

In this appendix, we complete the proof of Theorem 5.

PROOF: We define the mapping $\varphi : A \to B$ by $\varphi(x) = y$, where

$$y_1 = \big[1 - \rho\big(1 - (1 - x_n)^{n-1}\big)\big]^{1/(n-1)}; \qquad y_i = \frac{x_i - x_n}{1 - x_n} \cdot y_1 \quad \text{for } i > 1.$$

We first claim that φ is a bijection. As x_n ranges from 0 to $1 - (\frac{\rho-1}{\rho})^{1/(n-1)}$, we see that y_1 ranges from 1 to 0, and thus there is a bijection between valid y_1 values and valid x_n

values. Furthermore, for any fixed y_1 and x_n, there is a bijection between x_i and y_i for $i = 2, \dots, n-1$. (By varying x_i between x_n and 1, we can achieve all values of y_i between 0 and y_1.) Furthermore, for any fixed y_1 and x_n, the mapping from x_i to y_i is an increasing function of x_i, and therefore, for all $x \in A$, we have $y_1 \in [0, 1]$ and $y_1 \geq y_2 \geq \cdots \geq y_n = 0$. Thus, φ is a bijection between A and B. Next, we claim that for any $x \in A$, it holds that x is component-wise at least as large as $\varphi(x)$. Since $x_1 = 1$, it trivially holds that $x_1 \geq \varphi_1(x)$. Fix a value of x_n (and hence of y_1), and consider the bijection $g : [x_n, 1] \to [0, y_1]$ given by $g(z) = y_1(z - x_n)/(1 - x_n)$. We must show that $z - g(z) \geq 0$ for all $z \in [x_n, 1]$. This follows from noticing that $z - g(z)$ is a linear function of z and both $x_n - g(x_n) = x_n$ and $1 - g(1) = 1 - y_1$ are nonnegative.

We now show that φ scales surface measure of every measurable $S \subset A$ by a factor of $1/\rho$. Instead of directly analyzing surface measures, it suffices to prove that the function $\varphi' : W \to W$ scales volumes by ρ, where $W \subset \mathbb{R}^{n-1}$ is the set $\{w : 1 \geq w_1 \geq \cdots \geq w_{n-1} \geq 0\}$ and $\varphi'(w)$ drops the last (constant) coordinate of $\varphi(1, w_1, \dots, w_{n-1})$ and then (for notational convenience) permutes the first coordinate to the end. That is,

$$\varphi'(w_1, \dots, w_{n-1}) = \left(\frac{w_1 - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), \dots, \frac{w_{n-2} - w_{n-1}}{1 - w_{n-1}} z(w_{n-1}), z(w_{n-1}) \right),$$

where $z(w_{n-1}) = [1 - \rho(1 - (1 - w_{n-1})^{n-1})]^{1/(n-1)}$.

We now analyze the determinant of the Jacobian matrix J of φ'. We notice that the only nonzero entries of J are the diagonals and the rightmost column. In particular, J is upper triangular, and therefore its determinant is the product of its diagonal entries. We therefore compute

$$\begin{aligned}
\det(J) &= \left(\frac{z(w_{n-1})}{1 - w_{n-1}} \right)^{n-2} \cdot \frac{\partial}{\partial w_{n-1}} \bigl[1 - \rho\bigl(1 - (1 - w_{n-1})^{n-1}\bigr) \bigr]^{1/(n-1)} \\
&= \left(\frac{z(w_{n-1})}{1 - w_{n-1}} \right)^{n-2} \cdot \frac{-1}{n-1} \bigl(z(w_{n-1})^{-(n-2)} \cdot \rho \cdot (n-1)(1 - w_{n-1})^{n-2} \bigr) = -\rho,
\end{aligned}$$

as desired.

Last, suppose $y_1 \leq \varepsilon$. Then $[1 - \rho(1 - (1 - x_n)^{n-1})]^{1/(n-1)} \leq \varepsilon$ and thus $x_n \geq 1 - (\frac{\varepsilon^{n-1} + \rho - 1}{\rho})^{1/(n-1)}$. *Q.E.D.*

PROOF: We now complete the proof of Theorem 5. Fix the dimension n. For any value of c, the transformed measure on the hypercube $(c, c+1)^n$ we obtain is as follows:

- a point mass of $+1$ at $(c, c, \dots, c)$;
- mass of $-(n+1)$ uniformly distributed throughout the interior;
- mass of $-c$ distributed on each surface $x_i = c$ of the hypercube;
- mass of $c+1$ distributed on each surface $x_i = c+1$ of the hypercube.

For notational convenience when checking the stochastic dominance properties of Theorem 3, we will shift the hypercube to the origin. That is, we will consider instead the measure μ^c on $[0, 1]^n$ which has mass $+1$ at the origin, mass of $-c$ on each surface $x_i = 0$, et cetera. It is important to notice that the mass that μ assigns to the interior of $[0, 1]^n$ and to the origin does not depend on c, while the mass on each surface is a function of c.

For any $h \in (0, 1)$, define the region $Z(h) = \{x \in [0, 1]^n : \|x\|_1 \leq h\}$. For any fixed c_0, it holds that $\mu_+^{c_0}(Z(h)) = 1$ for all $h \in (0, 1)$ and there exists a small enough $h' > 0$ such that $\mu_-^{c_0}(Z(h')) < 1$. Since, for this fixed h', it holds that $\mu_-^c(Z(h'))$ increases with c

(and becomes arbitrarily large as c becomes large), there must exist a $c' > c_0$ such that $\mu_-^{c'}(Z(h')) = 1$, and thus $\mu^{c'}(Z(h')) = 0$. We can therefore pick a decreasing function $p^* : \mathbb{R}_{\geq 0} \to (0, 1)$ such that, for all sufficiently large c, $\mu^c(Z(p^*(c))) = 0$.[20] As argued above, for any small enough $h' > 0$, there exists a c' such that $\mu_-^{c'}(Z(h')) = 1$ and thus $p^*(c') = h'$. It follows that $p^*(c) \to 0$ as $c \to \infty$.

For all c, define the following subsets of $[0, 1]^n$:

$$Z_c = \{x : \|x\|_1 \leq p^*(c)\}; \qquad W_c = \{x : \|x\|_1 \geq p^*(c)\}.$$

We notice that $\mu_+^c(Z_c \cap W_c) = \mu_-^c(Z_c \cap W_c) = 0$. By construction, for large enough c, we have $\mu^c(Z_c) = 0$. In addition, the only positive mass in Z_c is at the origin, and thus $\mu_-^c|_{Z_c} \succeq_{\mathrm{cvx}} \mu_+^c|_{Z_c}$.

To apply Theorem 3, it remains to show that, for sufficiently large c, $\mu_+^c|_{W_c} \preceq_{\mathrm{cvx}(-\vec{1})} \mu_-^c|_{W_c}$. To prove this, we partition W_c into $2(n! + 1)$ disjoint[21] regions, $P_0, P_{\sigma_1}, \dots, P_{\sigma_{n!}}$ and $N_0, N_{\sigma_1}, \dots, N_{\sigma_{n!}}$, where σ_j is a permutation of $1, \dots, n$. This partition will be such that $\bigcup_j P_j$ contains the entire support of $\mu_+^c|_{W_c}$ and $\bigcup_j N_j$ contains the entire support of $\mu_-^c|_{W_c}$. We will show that $\mu_+^c|_{P_j} \preceq_{\mathrm{cvx}(-\vec{1})} \mu_-^c|_{N_j}$ for all j, thereby proving $\mu_+^c|_{W_c} \preceq_{\mathrm{cvx}(-\vec{1})} \mu_-^c|_{W_c}$.

For every permutation σ of $1, \dots, n$, define

$$P'_\sigma = \left\{x : 1 = x_{\sigma(1)} \geq x_{\sigma(2)} \geq \dots \geq x_{\sigma(n)} \geq 0 \text{ and } x_{\sigma(n)} \leq 1 - \left(\frac{1}{c+1}\right)^{1/(n-1)}\right\},$$

$$N'_\sigma = \{y : 1 \geq y_{\sigma(1)} \geq \dots \geq y_{\sigma(n-1)} \geq y_{\sigma(n)} = 0\}.$$

Denote by $\rho \triangleq (c+1)/c$ the ratio between the surface densities of μ_+^c and μ_-^c on P'_σ and N'_σ, respectively, and let $\varphi_\sigma : P'_\sigma \to N'_\sigma$ be the bijection given by Lemma 2. By construction, $\mu_+^c(S) = \mu_-^c(\varphi_\sigma(S))$ for all measurable $S \subseteq P'_\sigma$.

Denote $N_\sigma \triangleq N'_\sigma \setminus Z_c$ and $P_\sigma \triangleq \varphi^{-1}(N_\sigma)$. By construction, φ is a bijection between P_σ and N_σ, preserving the respective measures μ_+^c and μ_-^c, such that, for all $x \in P_\sigma$, x is component-wise at least as large as $\varphi(x)$. Therefore, by Strassen's theorem, $\mu_+^c|_{P_\sigma} \preceq_{\mathrm{cvx}(-\vec{1})} \mu_-^c|_{N_\sigma}$. Last, we define

$$P_0 = \{x \in [0, 1]^n : x_i = 1 \text{ for some } i\} \setminus \left(\bigcup_\sigma P_\sigma\right); \qquad N_0 = (0, 1)^n \setminus Z_c.$$

P_0 consists of all points on the outer surface of the hypercube which have not yet been matched to any N_σ, and N_0 consists of all points on which μ_-^c is nontrivial which have not yet been matched.[22] It therefore remains only to show that $\mu_+^c|_{P_0} \preceq_{\mathrm{cvx}(-\vec{1})} \mu_-^c|_{N_0}$.

We claim that, for large enough c, P_0 only contains points with all coordinates greater than $3/4$. Indeed:

- Every x with $x_i = 1$ but some $x_j < 1 - (\frac{1}{c+1})^{1/(n-1)}$ is in some P'_σ.
- For large c, every x with $x_i = 1$ but some $x_j \leq 3/4$ is in some P'_σ.

[20] Our intention is to argue that for c large enough, the optimal mechanism will be grand bundling for a price of $p^*(c) + c$, where the additive $+c$ term comes from our shift of the hypercube to the origin.

[21] For notational simplicity, our regions overlap slightly, although the overlap always has zero mass under both μ_+^c and μ_-^c.

[22] All other points on which μ_-^c is nontrivial have been matched either to the origin (if the point lies in Z_c), or to some point in P_σ (if the point lies in $N'_\sigma \setminus Z_c$).

- We claim that for large c, every $x \in P'_\sigma \setminus P_\sigma$ has all coordinates at least 3/4. Indeed, for every $x \in P'_\sigma \setminus P_\sigma$, it must be that $\varphi(x) \in Z_c$, and thus $\|\varphi(x)\|_1 \leq p^*(c)$. By Lemma 2, we have $x_{\sigma(n)} \geq 1 - (\frac{p^*(c)^{n-1}+\rho-1}{\rho})^{1/(n-1)}$. As c gets large, $\rho \to 1$ and $p^*(c) \to 0$. Thus, for sufficiently large c, we have $x \in P'_\sigma \setminus P_\sigma$ implies $x_{\sigma(n)} \geq 3/4$. Since $x_{\sigma(n)}$ is the smallest coordinate of x, it follows that all coordinates of any $x \in P'_\sigma \setminus P_\sigma$ are greater than 3/4.
- Thus, for sufficiently large c, every x with $x_i = 1$ but some $x_j < 3/4$ lies in some P_σ, and hence does not lie in P_0.

By construction, $\mu^c_-|_{N_0}$ and $\mu^c_+|_{P_0}$ have the same total mass. Consider independent random variables X and Y corresponding to $\mu^c_-|_{N_0}$ and $\mu^c_+|_{P_0}$, respectively, where we scale both measures so that they are probability distributions. By Lemma 6, it suffices to show that for sufficiently large c, $Y \geq \mathbb{E}[X]$ almost surely.[23] Since $\mu^c_+|_{P_0}$ is supported on P_0, we need only show that all coordinates of $\mathbb{E}[X]$ are less than 3/4. We recall that μ^c_- assigns a total mass of $n+1$, distributed uniformly, to the interior of the hypercube. As c gets large, $p^*(c)$ approaches 0, and thus

$$\frac{\mu^c_-\big(Z_c \cap (0,1)^n\big)}{\mu^c_-\big((0,1)^n\big)} \to 0.$$

For large c, therefore, $\mathbb{E}[X]$ becomes arbitrarily close to the center of the hypercube, which is the point with all coordinates equal to 1/2. Therefore, we have

$$\mu^c_+|_{P_0} \preceq_{\text{cvx}(-\vec{1})} \mu^c_-|_{N_0}. \qquad Q.E.D.$$

APPENDIX D: SUPPLEMENTARY MATERIAL FOR SECTION 7

PROOF: It is obvious that u_Z is nonnegative. To show that u_Z is nondecreasing, it suffices to prove that $u_Z(x) \geq u_Z(y)$ for $x, y \in X \setminus Z$ with x component-wise greater than or equal to y. Let $z_x \in Z$ be the closest point to x. Denote by z_y the point with each coordinate being the component-wise minimum of z_x and y. Since Z is decreasing, $z_y \in Z$. We now compute

$$u_Z(x) = \|z_x - x\|_1 = \sum_i \big|(z_x)_i - x_i\big| \geq \sum_i \big|\min\{(z_x)_i, y_i\} - y_i\big| = \|z_y - y\|_1 \geq u_Z(y)$$

and thus u_Z is nondecreasing.

We will now show that u_Z is convex. Pick arbitrary $x, y \in X$. Denote by z_x and z_y points in Z such that $u_Z(x) = \|x - z_x\|_1$ and $u_Z(y) = \|y - z_y\|_1$. Since Z is convex, the point $(z_x + z_y)/2$ is in Z. Thus

$$u_Z\left(\frac{x+y}{2}\right) \leq \left\|\frac{x+y}{2} - \frac{z_x+z_y}{2}\right\|_1 \leq \frac{\|x-z_x\|_1 + \|y-z_y\|_1}{2} = \frac{u_Z(x)+u_Z(y)}{2}$$

and therefore u_Z is convex.

Last, we verify that u_Z has Lipschitz constant at most 1. Indeed,

$$u_Z(x) - u_Z(y) \leq \|x - z_y\|_1 - u_Z(y) = \|x - z_y\|_1 - \|y - z_y\|_1 \leq \|x - y\|_1. \quad Q.E.D.$$

[23]In general, to prove second-order dominance, we might need to nontrivially couple X and Y. In this case, however, choosing independent random variables suffices.

APPENDIX E: SUPPLEMENTARY MATERIAL FOR SECTIONS 7 AND 8

E.1. *Verifying Stochastic Dominance—Proof of Lemma 3*

We begin with the standard result that a sufficient condition for first-order stochastic dominance is that one measure assigns more mass than the other to all increasing sets.

CLAIM 5: *Let α, β be positive finite Radon measures on $\mathbb{R}^n_{\geq 0}$ with $\alpha(\mathbb{R}^n_{\geq 0}) = \beta(\mathbb{R}^n_{\geq 0})$. A necessary and sufficient condition for $\alpha \succeq_1 \beta$ is that, for all increasing*[24] *measurable sets A, $\alpha(A) \geq \beta(A)$.*

PROOF: Without loss of generality, assume that $\alpha(\mathbb{R}^n_{\geq 0}) = \beta(\mathbb{R}^n_{\geq 0}) = 1$.

It is obvious that the condition is necessary by considering the indicator function of any increasing set A. To prove sufficiency, suppose that the condition holds and that on the contrary, α does not stochastically dominate β. Then there exists an increasing, bounded, measurable function f such that

$$\int f\,d\beta - \int f\,d\alpha > 2^{-k+1}$$

for some positive integer k. Without loss of generality, we may assume that f is nonnegative, by adding the constant of $-f(0)$ to all values. We now define the function $\tilde{f}$ by point-wise rounding f upwards to the nearest multiple of 2^{-k}. Clearly, $\tilde{f}$ is increasing, measurable, and bounded. Furthermore, we have

$$\int \tilde{f}\,d\beta - \int \tilde{f}\,d\alpha \geq \int f\,d\beta - \int f\,d\alpha - 2^{-k} > 2^{-k+1} - 2^{-k} > 0.$$

We notice, however, that $\tilde{f}$ can be decomposed into the weighted sum of indicator functions of increasing sets. Indeed, let $\{r_1, \ldots, r_m\}$ be the set of all values taken by $\tilde{f}$, where $r_1 > r_2 > \cdots > r_m$. We notice that, for any $s \in \{1, \ldots, m\}$, the set $A_s = \{z : \tilde{f}(z) \geq r_s\}$ is increasing and measurable. Therefore, we may write

$$\tilde{f} = \sum_{s=1}^{m} (r_s - r_{s-1}) I_s,$$

where I_s is the indicator function for A_s and where we set $r_0 = 0$. We now compute

$$\int \tilde{f}\,d\beta = \sum_{s=1}^{m} (r_s - r_{s-1})\beta(A_s) \leq \sum_{s=1}^{m} (r_s - r_{s-1})\alpha(A_s) = \int \tilde{f}\,d\alpha,$$

contradicting the fact that $\int \tilde{f}\,d\beta > \int \tilde{f}\,d\alpha$. *Q.E.D.*

Due to Claim 5, to verify that a measure α stochastically dominates β in the first order, we must ensure that $\alpha(A) \geq \beta(A)$ for all increasing measurable sets A. This verification might still be difficult, since an increasing set can have fairly unconstrained structure. In Lemma 13, we simplify this task by showing that we need not verify the inequality for all increasing A, but rather only for a special class of increasing subsets.

[24] An increasing set $A \subset \mathbb{R}^n_{\geq 0}$ satisfies the property that, for all $a, b \in \mathbb{R}^n_{\geq 0}$ such that a is component-wise greater than or equal to b, if $b \in A$, then $a \in A$ as well.

DEFINITION 16: For any $z \in \mathbb{R}^n_{\geq 0}$, we define the *base rooted at* z to be

$$B_z \triangleq \{z' : z \preceq z'\},$$

the minimal increasing set containing z, where the notation $z \preceq z'$ denotes that every component of z is at most the corresponding component of z'.

We denote Q_k to be the set of points in $\mathbb{R}^n_{\geq 0}$ with all coordinates multiples of 2^{-k}.

DEFINITION 17: An increasing set S is *k-discretized* if $S = \bigcup_{z \in S \cap Q_k} B_z$. A *corner* c of a k-discretized set S is a point $c \in S \cap Q_k$ such that there does not exist $z \in S \setminus \{c\}$ with $z \preceq c$.

LEMMA 12: *Every k-discretized set S has only finitely many corners. Furthermore*, $S = \bigcup_{c \in \mathcal{C}} B_c$, *where $\mathcal{C}$ is the collection of corners of S.*

PROOF: We prove that there are finitely many corners by induction on the dimension, n. In the case $n = 1$, the result is obvious, since if S is nonempty it has exactly one corner. Now suppose S has dimension n. Pick some corner $\hat{c} = (c_1, \dots, c_n) \in S$. We know that any other corner must be strictly less than $\hat{c}$ in some coordinate. Therefore,

$$|\mathcal{C}| \leq 1 + \sum_{i=1}^{n} \bigl|\{c \in \mathcal{C} \text{ s.t. } c_i < \hat{c}_i\}\bigr| = 1 + \sum_{i=1}^{n} \sum_{j=1}^{2^k \hat{c}_i} \bigl|c \in \mathcal{C} \text{ s.t. } c_i = \hat{c}_i - 2^{-k} j\bigr|.$$

By the inductive hypothesis, we know that each set $\{c \in \mathcal{C} \text{ s.t. } c_i = \hat{c}_i - 2^{-k} j\}$ is finite, since it is contained in the set of corners of the $(n-1)$-dimensional subset of S whose points have ith coordinate $\hat{c}_i - 2^{-k} j$. Therefore, $|\mathcal{C}|$ is finite.

To show that $S = \bigcup_{c \in \mathcal{C}} B_c$, pick any $z \in S$. Since S is k-discretized, there exists a $b \in S \cap Q_k$ such that $z \in B_b$. If b is a corner, then z is clearly contained in $\bigcup_{c \in \mathcal{C}} B_c$. If b is not a corner, then there is some other point $b' \in S \cap Q_k$ with $b' \preceq b$. If b' is a corner, we are done. Otherwise, we repeat this process at most $2^k \sum_j b_j$ times, after which time we will have reached a corner c of S. By construction, we have $z \in B_c$, as desired. *Q.E.D.*

We now show that, to verify that one measure dominates another on all increasing sets, it suffices to verify that this holds for all sets that are the union of finitely many bases.

LEMMA 13: *Let $g, h : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ be bounded integrable functions such that $\int_{\mathbb{R}^n_{\geq 0}} g(x)\,dx$ and $\int_{\mathbb{R}^n_{\geq 0}} h(x)\,dx$ are finite. Suppose that, for all finite collections Z of points in $\mathbb{R}^n_{\geq 0}$, we have*

$$\int_{\bigcup_{z \in Z} B_z} g(x)\,dx \geq \int_{\bigcup_{z \in Z} B_z} h(x)\,dx.$$

Then for all increasing sets $A \subseteq \mathbb{R}^n_{\geq 0}$,

$$\int_A g(x)\,dx \geq \int_A h(x)\,dx.$$

PROOF: Let A be an increasing set. We clearly have $A = \bigcup_{z \in A} B_z$. For any point $z \in \mathbb{R}^n_{\geq 0}$, denote by $z^{n,k}$ the point in $\mathbb{R}^n_{\geq 0}$ such that, for each component i, the ith component of $z^{n,k}$ is the maximum of 0 and $z_i - 2^{-k}$.

We define the following two sets, which we think of as approximations of A:

$$A_k^l \triangleq \bigcup_{z\in A\cap Q_k} B_z; \qquad A_k^u \triangleq \bigcup_{z\in A\cap Q_k} B_{z^{n,k}}.$$

It is clear that both A_k^l and A_k^u are k-discretized. Furthermore, for any $z \in A$, there exists a $z' \in A \cap Q_k$ such that each component of z' is at most 2^{-k} more than the corresponding component of z. Therefore, $A_k^l \subseteq A \subseteq A_k^u$.

We now will bound

$$\int_{A_k^u} g(x)\,dx - \int_{A_k^l} g(x)\,dx.$$

Let

$$W_k = \{z \in \mathbb{R}^n_{\geq 0} : z_i > k \text{ for some } i\}; \qquad W_k^c = \{z \in \mathbb{R}^n_{\geq 0} : z_i \leq k \text{ for all } i\}.$$

The set W_k^c contains all points which lie inside a box of side length k rooted at the origin, and W_k contains all points outside of this box. We have the immediate (loose) bound that

$$\int_{A_k^u\cap W_k} g\,dx - \int_{A_k^l\cap W_k} g\,dx \leq \int_{W_k} g\,dx.$$

Furthermore, since $\lim_{k\to\infty}\int_{W_k^c} g\,dx = \int_{\mathbb{R}^n_{\geq 0}} g\,dx$, we know that $\lim_{k\to\infty}\int_{W_k} g\,dx = 0$. Therefore,

$$\lim_{k\to\infty}\left(\int_{A_k^u\cap W_k} g\,dx - \int_{A_k^l\cap W_k} g\,dx\right) = 0.$$

Next, we bound

$$\int_{A_k^u\cap W_k^c} g\,dx - \int_{A_k^l\cap W_k^c} g\,dx \leq |g|_{\sup}\big(V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c)\big),$$

where $|g|_{\sup} < \infty$ is the supremum of g, and $V(\cdot)$ denotes the Lebesgue measure.

For each $m \in \{1, \dots, n+1\}$ and $z \in \mathbb{R}^n_{\geq 0}$, we define the point $z^{m,k}$ by

$$z_i^{m,k} = \begin{cases} \max\{0, z_i - 2^{-k}\}, & \text{if } i < m,\\ z_i, & \text{otherwise,}\end{cases}$$

and set

$$A_k^m \triangleq \bigcup_{z\in A\cap Q_k} B_{z^{m,k}}.$$

We have, by construction, $A_k^l = A_k^1$ and $A_k^u = A_k^{n+1}$. Therefore,

$$V(A_k^u \cap W_k^c) - V(A_k^l \cap W_k^c) = \sum_{m=1}^{n}\big(V(A_k^{m+1} \cap W_k^c) - V(A_k^m \cap W_k^c)\big).$$

We notice that, for any point $(z_1, z_2, \dots, z_{m-1}, z_{m+1}, \dots, z_n) \in [0, k]^{n-1}$, there is an interval I of length at most 2^{-k} such that

$$(z_1, z_2, \dots, z_{m-1}, w, z_{m-2}, \dots, z_n) \in (A_k^{m+1} \setminus A_k^m) \cap W_k^c$$

if and only if $w \in I$. Therefore,

$$V\big(A_k^{m+1} \cap W_k^c\big) - V\big(A_k^m \cap W_k^c\big) \\ \leq \int_0^k \cdots \int_0^k \int_0^k \cdots \int_0^k 2^{-k}\, dz_1 \cdots dz_{m-1}\, dz_{m+1} \cdots dz_n = 2^{-k} k^{n-1}.$$

We thus have the bound

$$|g|_{\sup}\big(V\big(A_k^u \cap W_k^c\big) - V\big(A_k^l \cap W_k^c\big)\big) \leq |g|_{\sup} \sum_{m=1}^{n} 2^{-k} k^{n-1} = n|g|_{\sup} 2^{-k} k^{n-1}$$

and therefore

$$\int_{A_k^u} g\, dx - \int_{A_k^l} g\, dx = \int_{A_k^u \cap W_k} g\, dx - \int_{A_k^l \cap W_k} g\, dx + \int_{A_k^u \cap W_k^c} g\, dx - \int_{A_k^l \cap W_k^c} g\, dx \\ \leq \left(\int_{A_k^u \cap W_k} g\, dx - \int_{A_k^l \cap W_k} g\, dx\right) + n|g|_{\sup} 2^{-k} k^{n-1}.$$

In particular, we have

$$\lim_{k \to \infty} \left(\int_{A_k^u} g\, dx - \int_{A_k^l} g\, dx\right) = 0.$$

Since $\int_{A_k^u} g\, dx \geq \int_A g\, dx \geq \int_{A_k^l} g\, dx$, we have

$$\lim_{k \to \infty} \int_{A_k^u} g\, dx = \int_A g\, dx = \lim_{k \to \infty} \int_{A_k^l} g\, dx.$$

Similarly, we have

$$\int_A h\, dx = \lim_{k \to \infty} \int_{A_k^l} h\, dx$$

and thus

$$\int_A (g - h)\, dx = \lim_{k \to \infty} \left(\int_{A_k^l} g\, dx - \int_{A_k^l} h\, dx\right).$$

Since A_k^l is k-discretized, it has finitely many corners. Letting Z_k denote the corners of A_k^l, we have $A_k^l = \bigcup_{z \in Z_k} B_z$, and thus, by our assumption, $\int_{A_k^l} g\, dx - \int_{A_k^l} h\, dx \geq 0$ for all k. Therefore, $\int_A (g - h)\, dx \geq 0$, as desired. *Q.E.D.*

We are now ready to prove Lemma 3.

PROOF: We begin by defining, for any a and b with $p_1 \leq a \leq b \leq q_1$, the function $\zeta_a^b : [p_2, q_2] \to \mathbb{R}$ by

$$\zeta_a^b(w_2) \triangleq \int_a^b \big(g(z_1, w_2) - h(z_1, w_2)\big)\, dz_1.$$

This function $\zeta_a^b(w_2)$ represents the integral of $g - h$ along the vertical line from (a, w_2) to (b, w_2).

CLAIM 6: *If* $(a, w_2) \in R$, *then* $\zeta_a^b(w_2) \leq 0$.

PROOF: The inequality trivially holds unless there exists a $z_1 \in [a, b]$ such that $g(z_1, w_2) > h(z_1, w_2)$, so suppose such a z_1 exists. It must be that $(z_1, w_2) \notin R$, since both g and h are 0 in R. Indeed, because R is a decreasing set, it is also true that $(\tilde{z}_1, w_2) \notin R$ for all $\tilde{z}_1 \geq z_1$. This implies by our assumption that

$$g(\tilde{z}_1, w_2) - h(\tilde{z}_1, w_2) = \alpha(\tilde{z}_1) \cdot \beta(w_2) \cdot \eta(\tilde{z}_1, w_2),$$

for all $\tilde{z}_1 \geq z_1$. Given that $g(z_1, w_2) > h(z_1, w_2)$ and that $\eta(\cdot, w_2)$ is an increasing function, we know that $g(\tilde{z}_1, w_2) \geq h(\tilde{z}_1, w_2)$ for all $\tilde{z}_1 \geq z_1$. Therefore, we have

$$\zeta_a^{z_1}(w_2) \leq \zeta_a^b(w_2) \leq \zeta_a^{q_1}(w_2).$$

We notice, however, that $\zeta_a^{q_1}(w_2) \leq 0$ by assumption, and thus the claim is proven. *Q.E.D.*

We now claim the following:

CLAIM 7: *Suppose that* $\zeta_a^b(w_2^*) > 0$ *for some* $w_2^* \in [c_2, q_2)$. *Then* $\zeta_a^b(w_2) \geq 0$ *for all* $w_2 \in [w_2^*, q_2)$.

PROOF: Given that $\zeta_a^b(w_2^*) > 0$, our previous claim implies that $(a, w_2^*) \notin R$. Furthermore, since R is a decreasing set and $w_2 \geq w_2^*$, it follows that $(a, w_2) \notin R$, and furthermore that $(c, w_2) \notin R$ for any $c \geq a$ in $[c_1, q_1)$. Therefore, we may write

$$\zeta_a^b(w_2) = \int_a^b \big(g(z_1, w_2) - h(z_1, w_2)\big)\, dz_1 = \int_a^b \big(\alpha(z_1) \cdot \beta(w_2) \cdot \eta(z_1, w_2)\big)\, dz_1.$$

Similarly, $(c, w_2^*) \notin R$ for any $c \geq a$, so

$$\zeta_a^b\big(w_2^*\big) = \int_a^b \big(\alpha(z_1) \cdot \beta\big(w_2^*\big) \cdot \eta\big(z_1, w_2^*\big)\big)\, dz_1.$$

Note that, since $\zeta_a^b(w_2^*) > 0$, we have $\beta(w_2^*) > 0$. Thus, since η is increasing,

$$\zeta_a^b(w_2) \geq \int_a^b \big(\alpha(z_1) \cdot \beta(w_2) \cdot \eta\big(z_1, w_2^*\big)\big)\, dz_1 = \frac{\beta(w_2)}{\beta\big(w_2^*\big)} \zeta_a^b\big(w_2^*\big) \geq 0,$$

as desired. *Q.E.D.*

We extend g and h to all of $\mathbb{R}^2_{\geq 0}$ by setting them to be 0 outside of $\mathcal{C}$. By Claim 13, to prove that $g \succeq_1 h$, it suffices to prove that $\int_A g\, dx\, dy \geq \int_A h\, dx\, dy$ for all sets A which are the union of finitely many bases. Since g and h are 0 outside of $\mathcal{C}$, it suffices to consider only bases $B_{z'}$ where $z' \in \mathcal{C}$, since otherwise we can either remove the base (if it is disjoint from $\mathcal{C}$) or can increase the coordinates of z' moving it to $\mathcal{C}$ without affecting the value of either integral.

We now complete the proof of Lemma 3 by induction on the number of bases in the union.

Base Case. We aim to show $\int_{B_r}(g-h)\,dx\,dy \geq 0$ for any $r=(r_1,r_2)\in\mathcal{C}$. We have

$$\int_{B_r}(g-h)\,dx\,dy = \int_{r_2}^{q_2}\int_{r_1}^{q_1}(g-h)\,dz_1\,dz_2 = \int_{r_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)\,dz_2.$$

By Claim 7, we know that either $\zeta_{r_1}^{q_1}(z_2)\geq 0$ for all $z_2\geq r_2$, or $\zeta_{r_1}^{q_1}(z_2)\leq 0$ for all z_2 between p_2 and r_2. In the first case, the integral is clearly nonnegative, so we may assume that we are in the second case. We then have

$$\begin{aligned}\int_{r_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)\,dz_2 &\geq \int_{p_2}^{q_2}\zeta_{r_1}^{q_1}(z_2)\,dz_2\\ &= \int_{p_2}^{q_2}\int_{r_1}^{q_1}(g-h)\,dz_1\,dz_2\\ &= \int_{r_1}^{q_1}\int_{p_2}^{q_2}(g-h)\,dz_2\,dz_1.\end{aligned}$$

By an analogous argument to that above, we know that either $\int_{p_2}^{q_2}(g-h)(z_1,z_2)\,dz_2$ is nonnegative for all $z_1\geq r_1$ (in which case the desired inequality holds trivially) or is nonpositive for all z_1 between p_1 and r_1. We assume therefore that we are in the second case, and thus

$$\begin{aligned}\int_{r_1}^{q_1}\int_{p_2}^{q_2}(g-h)\,dz_2\,dz_1 &\geq \int_{p_1}^{q_1}\int_{p_2}^{q_2}(g-h)\,dz_2\,dz_1\\ &= \int_{\mathcal{C}}(g-h)\,dx\,dy,\end{aligned}$$

which is nonnegative by assumption.

Inductive Step. Suppose that we have proven the result for all sets which are finite unions of at most k bases. Consider now a set

$$A = \bigcup_{i=1}^{k+1} B_{z^{(i)}}.$$

We may assume that all $z^{(i)}$ are distinct and that there do not exist distinct $z^{(i)}$, $z^{(j)}$ with $z^{(i)}$ component-wise less than $z^{(j)}$, since otherwise we could remove one such $B_{z^{(i)}}$ from the union without affecting the set A and the desired inequality would follow from the inductive hypothesis.

We may therefore order the $z^{(i)}$ such that

$$\begin{aligned}p_1 &\leq z_1^{(k+1)} < z_1^{(k)} < z_1^{(k-1)} < \cdots < z_1^{(1)},\\ p_2 &\leq z_2^{(1)} < z_2^{(2)} < z_2^{(3)} < \cdots < z_2^{(k+1)}.\end{aligned}$$

We can now use Claim 7 to transform the set A that consists of a union of $k+1$ bases into a set consisting of k bases. See Figure 9 for an illustration. By Claim 7, we know that one of the two following cases must hold:

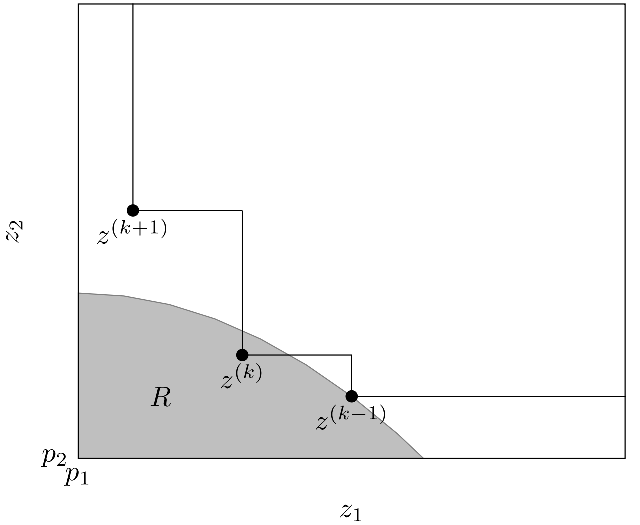

FIGURE 9.—We show that either decreasing $z_2^{(k+1)}$ to $z_2^{(k)}$ or removing $z^{(k+1)}$ entirely decreases the value of $\int_A(f-g)$. In either case, we can apply our inductive hypothesis.

CASE 1: $\zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w_2) \le 0$ *for all* $p_2 \le w_2 \le z_2^{(k+1)}$.

In this case, we see that

$$\int_{z_2^{(k)}}^{z_2^{(k+1)}} \int_{z_1^{(k+1)}}^{z_1^{(k)}} (f-g)\,dz_1\,dz_2 = \int_{z_2^{(k)}}^{z_2^{(k+1)}} \zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w)\,dw \le 0.$$

For notational purposes, we denote here by $(f-g)(S)$ *the integral* $\int_S(f-g)\,dz_1\,dz_2$ *for any set* S. *We compute*

$$\begin{aligned}(f-g)(A) &\ge (f-g)(A) \\ &\quad + (f-g)\big(\big\{z : z_1^{(k+1)} \le z_1 \le z_1^{(k)} \text{ and } z_2^{(k)} \le z_2 \le z_2^{(k+1)}\big\}\big) \\ &= (f-g)\left(\bigcup_{i=1}^{k} B_{z^{(i)}} \cup B_{(z_1^{(k+1)}, z_2^{(k)})}\right) \\ &= (f-g)\left(\bigcup_{i=1}^{k-1} B_{z^{(i)}} \cup B_{(z_1^{(k+1)}, z_2^{(k)})}\right),\end{aligned}$$

where the last equality follows from $(z_1^{(k)}, z_2^{(k)})$ *being component-wise greater than or equal to* $(z_1^{(k+1)}, z_2^{(k)})$. *The inductive hypothesis implies that the quantity in the last line of the above derivation is* ≥ 0.

CASE 2: $\zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w_2) \ge 0$ *for all* $w_2 \ge z_2^{(k+1)}$.

In this case, we have

$$\int_{z_2^{(k+1)}}^{q_2} \int_{z_1^{(k+1)}}^{z_1^{(k)}} (f-g)\, dz_1\, dz_2 = \int_{z_2^{(k+1)}}^{q_2} \zeta_{z_1^{(k+1)}}^{z_1^{(k)}}(w)\, dw \geq 0.$$

Therefore, it follows that

$$\begin{aligned}(f-g)(A) &= (f-g)\left(\bigcup_{i=1}^{k} B_{z^{(i)}}\right) \\ &\quad + (f-g)\left(\left\{z : z_1^{(k+1)} \leq z_1 \leq z_1^{(k)} \text{ and } z_2^{(k+1)} \leq z_2\right\}\right) \\ &\geq (f-g)\left(\bigcup_{i=1}^{k} B_{z^{(i)}}\right) \geq 0,\end{aligned}$$

where the final inequality follows from the inductive hypothesis. Q.E.D.

E.2. *Verifying Stochastic Dominance in Example 3*

We sketch the application of Lemma 3 for verifying that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ in Example 3. We set $\mathcal{C} = [x_{\text{crit}}, 1] \times [y_{\text{crit}}, 1]$ and $\mathcal{R} = Z \cap \mathcal{C}$, so that $\mathcal{W} = \mathcal{C} \setminus R$. We let g and h be the positive and negative parts of the density function of $\mu|_{\mathcal{W}}$, respectively, so that the density of $\mu|_{\mathcal{W}}$ is given by $g - h$. Since Z lies below *both* curves S_{top} and S_{right}, we know that integrating the density of μ along any horizontal or vertical line outwards starting anywhere on the boundary of Z yields a non-positive quantity, verifying the second condition of Lemma 3. In addition, on $\mathcal{W} = \mathcal{C} \setminus R$, we have

$$g(z_1, z_2) - h(z_1, z_2) = f_1(z_1) f_2(z_2)\left(\frac{1}{1-z_1} + \frac{1}{1-z_2} - 5\right),$$

which satisfies the third condition of Lemma 3, as $1/(1-z_1) + 1/(1-z_2) - 5$ is increasing. Finally, we verify the first condition of Lemma 3 by integrating $g - h$ over $\mathcal{C}$. This integral is equal to $\mu(\mathcal{W}) = 0$ and thus all conditions of Lemma 3 are satisfied.

E.3. *Uniqueness of Mechanism in Example 3*

To argue that the utility $u(x)$ is shared by all optimal mechanisms, we start by constructing an optimal solution γ^* to the RHS of (5). γ^* needs to satisfy the complementary slackness conditions of Corollary 1 against any optimal solution u^* to the LHS of (5). We will choose our solution γ^* so that the complementary slackness conditions will imply $u^* = u$. Let us proceed with the choice of γ^*. Recall the canonical partition $Z \cup \mathcal{A} \cup \mathcal{B} \cup \mathcal{W}$ of the type space, identified above, and illustrated in Figure 6. We define a solution γ^* to the RHS of (5) that separates into the four regions as follows (the optimality of this γ^* follows easily by checking that it satisfies the complementary slackness conditions of Corollary 1 against u):

Region Z

Recall that, in region Z, we have $\mu|_Z \preceq_{\text{cvx}} 0$. Our solution γ^* matches the $+1$ unit of mass sitting at the origin to the negative mass spread throughout region Z, by moving positive mass to coordinate-wise larger points and performing mean-preserving spreads. By the complementary slackness conditions of Corollary 1 (see Remark 1 for intuition), it follows that $u^*(x) = 0$, for any optimal solution u^* to the LHS of (5).

Regions $\mathcal{A}$ and $\mathcal{B}$

In regions $\mathcal{A}$ and $\mathcal{B}$, our solution γ^* transports mass vertically and, respectively, horizontally. The complementary slackness conditions imply then that any optimal solution u^* to the LHS of (5) u^* must change linearly in the second coordinate in region $\mathcal{A}$ and linearly in the first coordinate in region $\mathcal{B}$.

Region $\mathcal{W}$

Finally, in region $\mathcal{W}$, we want to show that any optimal u satisfies $|u(\vec{x}) - u(\vec{y})| = \|\vec{x} - \vec{y}\|_1$ if $\vec{x} \geq \vec{y}$ coordinate-wise. This is not as straightforward as the previous two cases, as we do not have an explicit description of the optimal dual solution. However, we can use Lemma 3 to show that there exists a measure γ^* which is optimal for the dual and matches types on the top right corner (with values $\approx (1,1)$) to types close to the bundling line (with values $x_1 + x_2 \approx p^*$), which implies that any optimal function u must be linear in $\mathcal{W}$.

By continuity, any optimal u must be equal to $z_1 + z_2 - p^* = 0$ when $z_1 + z_2 = p^*$. Moreover, it holds that $u(z) \leq z_1 + z_2 - p^*$, because u is 1-Lipschitz. We will now show the reverse inequality by showing that $u(1,1) = 2 - p^*$. Recall that the density of measure μ in region $\mathcal{W}$ is equal to

$$\mu(z_1, z_2) = f_1(z_1)f_2(z_2)\left(\frac{1}{1-z_1} + \frac{1}{1-z_2} - 5\right),$$

where $f_1(x) = f_2(x) = (1 - x)$. Lemma 3 implied that $\mu_+|_{\mathcal{W}} \succeq_1 \mu_-|_{\mathcal{W}}$ but did not give a transport map γ constructively. To partially specify a transport map γ that is optimal for the dual, we define, for sufficiently small $\varepsilon > 0$, the measure μ' which has density

$$\mu'(z_1, z_2) = f_1(z_1)f_2(z_2)\left(\frac{1}{\varepsilon} + \max\left(\frac{1}{1-z_2}, \frac{1}{1-z_1}\right) - 5\right)$$

when $(z_1, z_2) \in [1-\varepsilon, 1]^2$ and $\mu'(z_1, z_2) = \mu(z_1, z_2)$ otherwise. In particular, μ' is obtained by removing some positive mass from μ in $[1-\varepsilon, 1]^2$ and thus $\mu'(\mathcal{W}) < \mu(\mathcal{W}) = 0$. Moreover, notice that we defined μ' so that $\frac{\mu'(z_1,z_2)}{f_1(z_1)f_2(z_2)}$ is still an increasing function. Now, let R' be the region enclosed within the curves $s_1(x)$, $s_2(y)$, $x + y = p^*$ and $x + y = p'$ for $p' > p^*$ so that $\mu'(\mathcal{W} \setminus R') = 0$. This defines a decomposition of measure $\mu|_{\mathcal{W}}$ into two measures $\mu'|_{\mathcal{W}\setminus R'}$ and $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W}\setminus R'}$ of zero total mass (Figure 10).

We apply Lemma 3 for μ' in region $\mathcal{W} \setminus R'$ to get that $\mu'|_{\mathcal{W}\setminus R'} \succeq_1 0$. We also have that $(\mu - \mu')|_{\mathcal{W}} \succeq_1 \mu|_{R'}$ since $(\mu - \mu')|_{\mathcal{W}}$ contains only positive mass supported on $[1-\varepsilon, 1]^2$ and every point in the support point-wise dominates every point in the support of $\mu|_{R'}$. Thus, there exists an optimal transport map γ^* in region $\mathcal{W}$ such that $\gamma^* = \gamma^{(i)} + \gamma^{(ii)}$ and $\gamma^{(i)}$ transports the mass $\mu'|_{\mathcal{W}\setminus R'}$ while $\gamma^{(ii)}$ transports mass arbitrarily from $(\mu - \mu')|_{\mathcal{W}}$ to $\mu|_{R'}$. Given such an optimal γ^*, the complementary slackness conditions of Corollary 1 imply that any feasible u must satisfy $|u(\vec{z}) - u(\vec{z}')| = \|\vec{z} - \vec{z}'\|_1$ whenever mass is transfered from $\vec{z}$ to $\vec{z}'$. This can only happen if $u(1,1) = 2 - p^*$ and implies that $u(\vec{z}) = z_1 + z_2 - p^*$ everywhere on $\mathcal{W}$.

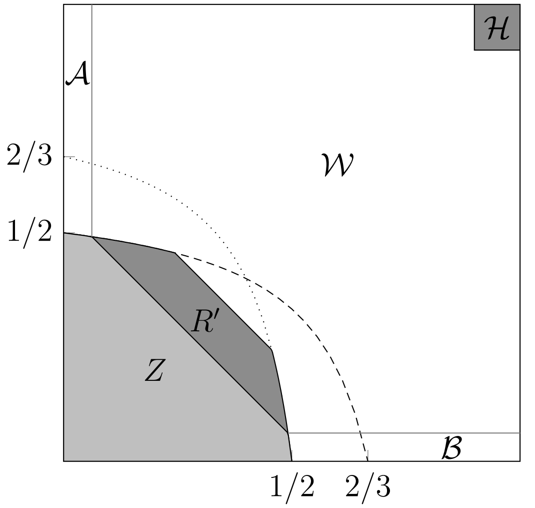

FIGURE 10.—Decomposition of measure $\mu|_{\mathcal{W}}$ into measures $\mu'|_{\mathcal{W}\setminus R'}$ and $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W}\setminus R'}$. The dark shaded regions R' and $\mathcal{H} = [1-\varepsilon, 1]^2$ show the support of $\mu|_{\mathcal{W}} - \mu'|_{\mathcal{W}\setminus R'}$.

APPENDIX F: EXTENDING TO UNBOUNDED DISTRIBUTIONS

Several results of this paper extend to unbounded type spaces, although such extensions impose additional technical difficulties. Here we briefly discuss how some of our results generalize.

We can often obtain a "transformed measure" (analogous to Theorem 1 even when type spaces are unbounded) using integration by parts. We wish to ensure, however, that the density function f decays sufficiently quickly so that there is no "surface term at infinity." For example, we may require that $\lim_{z_i \to \infty} f_i(z_i) z_i^2 \to 0$, as in Daskalakis, Deckelbaum, and Tzamos (2013). We note that without some conditions on the decay rate of f, it is possible that the supremum revenue achievable is infinite and thus no optimal mechanism exists.

Similar issues arise when integrating with respect to an unbounded measure μ. It is helpful, therefore, to consider only measures μ such that $\int \|x\|_1 \, d|\mu| < \infty$, to ensure that $\int u \, d\mu$ is finite for any utility function u. The measures in our examples satisfy this property. We can (informally speaking) attempt to extend this definition to unbounded measures (with regularity conditions such as $\int \|x\|_1 \, d|\mu| < \infty$) by ensuring that whenever the "smaller" side has infinite value, so does the larger side.

Importantly, the calculations of Lemma 1 (weak duality) hold for unbounded μ, provided $\int \|x\|_1 \, d|\mu| < \infty$. Thus, tight certificates still certify optimality, even in the unbounded case. However, our strong duality proof relies on technical tools which require compact spaces, and thus these proofs do not immediately apply when μ is unbounded.

To summarize our discussion so far, we can often transform measures and obtain an analogue of Theorem 1 for unbounded distributions (provided the distributions decay sufficiently quickly), and can easily obtain a weak duality result for such unbounded measures, but additional work is required to prove whether strong duality holds.

REFERENCES

DASKALAKIS, C., A. DECKELBAUM, AND C. TZAMOS (2013): "Mechanism Design via Optimal Transport," in *14th ACM Conference on Electronic Commerce (EC)*. [27]

DUDLEY, R. M. (2002): *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press. [9]

SHAKED, M., AND J. G. SHANTHIKUMAR (2010): *Stochastic Orders*. Springer Series in Statistics. New York: Springer. [8]

VILLANI, C. (2008): *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. 338. Berlin: Springer. [1]

EECS, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.; costis@csail.mit.edu,
Math, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.; deckel@alum.mit.edu,
and
EECS, MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, U.S.A.; tzamos@mit.edu.

Co-editor Dirk Bergemann handled this manuscript.

Manuscript received 2 July, 2014; final version accepted 20 October, 2016; available online 15 November, 2016.