Online Appendix to "An Axiomatic Model of Persuasion"

Alexander M. Jakobsen*

January 31, 2020

Abstract

This appendix proves Theorem 4 and Propositions 1–3 from Jakobsen (2020).

1 Proof of Theorem 4

Throughout this section, suppose $\succsim$ is representable by some unknown parameters (ν, μ, v, u). Elicitation of the parameters from $\succsim$ is carried out in several steps, broken down as a series of lemmas.

For an interior lottery p, the *agreement region* is the set of all lotteries q such that, for some neighborhood $N(q)$ of q, $\sigma^* \succsim^A \sigma$ for all $\sigma \in \mathcal{E}$ and pq'-bets A where $q' \in N(q)$. Given an interior lottery p, a pair of utility indices (w, w') is *consistent with the agreement region for* p if, for all q in the agreement region for p, either $[w(p) > w(q)$ and $w'(p) > w'(q)]$ or $[w(q) > w(p)$ and $w'(q) > w'(p)]$. Two pairs of indices (w, w') and $(\tilde{w}, \tilde{w}')$ are a *joint affine transformation* of one another if there exists $A, B \in \mathbb{R}$ such that $\tilde{w} = Aw + B$ and $\tilde{w}' = Aw' + B$ (note that this definition does not require $A > 0$).

Lemma 1. *There exists a unique (up to joint affine transformation) pair of utility indices* (w, w') *such that, for every interior lottery* p*, the agreement region for* p *is consistent with* (w, w')*.*

Proof. Fix an interior lottery p and let $q \in \Delta X$. If the (underlying) indices v and u give the same strict ranking of p and q, then σ^* gives the highest-possible payoff to Sender in any

*Department of Economics, University of Calgary.

pq-bet. Since the ranking is strict, q can be perturbed without reversing the strict ranking for either agent.

Conversely, suppose there is a neighborhood $N(q)$ such that σ^* is top-ranked by $\succsim^A$ for all pq'-bets where $q' \in N(q)$. Suppose toward a contradiction that v and u do not give the same strict ranking of p and q. Then there is a $q' \in N(q)$ where v and u strictly disagree on the ranking of p and q' (one index ranks p strictly better, the other ranks q' strictly better). Let $A' = \{pEq', pFq'\}$ and $E^* = (E \backslash F) \cup (F \backslash E)$. Receiver's choice of act from A' at states in $E \backslash F$ differs from that in states $F \backslash E$; his choices result in Sender's less-preferred lottery for all $\omega \in E^*$. Thus, σ^* is strictly less-preferred than an experiment σ formed by perturbing signals in σ^* corresponding to states in E^* to give positive likelihood to each state in E^*, because this gives a positive probability of Sender's more-preferred lottery for every $\omega \in E^*$. Hence, σ^* is not top-ranked by $\succsim^{A'}$, a contradiction.

Thus, the agreement region for p reveals the set of all q such that v and u agree on the ranking of p and q. The boundary of this region near p is determined by two (possibly identical) planes through p, corresponding to indifference curves for linear functions w and w'. □

For the remainder of the proof, we take the pair (w, w') elicited by Lemma 1 as given. A bet $A = \{pEq, pFq\}$ is *(w, w')-generic* if both $w(p) \neq w(q)$ and $w'(p) \neq w'(q)$. Signals s and t are *EF-equivalent* if there are neighborhoods $N(s)$ and $N(t)$ of s and t such that, for all $s' \in N(s)$, $t' \in N(t)$ and σ with $s', t' \in \sigma$, we have $\sigma \sim^A \sigma^{s'+t'}$ for all (w, w')-generic EF-bets A. For a full-support distribution $\tilde{\mu} \in \Delta\Omega$ and a signal $r \in S$, let $\tilde{\mu}^r$ denote the Bayesian posterior of $\tilde{\mu}$ at r. A full-support $\tilde{\mu}$ is *consistent with the EF-equivalence regions* if there exist $s, t \in S$ such that $\{r \in S : \tilde{\mu}^r(E) > \tilde{\mu}^r(F)\} = \{r \in S : r \text{ is } EF\text{-equivalent to } s\}$ and $\{r \in S : \tilde{\mu}^r(F) > \tilde{\mu}^r(E)\} = \{r \in S : r \text{ is } EF\text{-equivalent to } t\}$.

Lemma 2. *There is a unique full-support $\tilde{\mu} \in \Delta\Omega$ such that, for all E and F, $\tilde{\mu}$ is consistent with the EF-equivalence region.*

Proof. It will suffice to show that arbitrary signals s and t are EF-equivalent if and only if either $[\mu^s(E) > \mu^s(F)$ and $\mu^t(E) > \mu^t(F)]$ or $[\mu^s(F) > \mu^s(E)$ and $\mu^t(F) > \mu^t(E)]$ (that is, s and t *agree on the ranking of E and F*), where μ is the true underlying prior for Receiver. It is straightforward to show that if s and t agree on the ranking of E and F, then s and t are EF-equivalent (this follows from the fact that Receiver has a unique optimal choice in a (w, w')-generic EF-bet if his beliefs assign strictly higher probability to one of the events).

Conversely, suppose s and t are EF-equivalent. First, we prove that $\mu^s(E) > \mu^s(F) \Rightarrow \mu^t(E) \geq \mu^t(F)$; the proof is by contradiction.

Suppose $\mu^s(E) > \mu^s(F)$ and $\mu^t(F) > \mu^t(E)$. Let $A = \{pEq, pFq\}$ be a (w, w')-generic bet. This implies $u(p) \neq u(q)$ and $v(p) \neq v(q)$ for the true indices u and v. Suppose $u(p) > u(q)$ (the case $u(q) > u(p)$ is similar). We will show that for all $\varepsilon > 0$, there exist $s' \in N^\varepsilon(s)$, $t' \in N^\varepsilon(t)$ and an experiment σ with $s', t' \in \sigma$ such that $\sigma \not\sim^A \sigma^{s'+t'}$. In particular, take $\varepsilon > 0$ small enough so that $\mu^{s'}(E) > \mu^{s'}(F)$ and $\mu^{t'}(F) > \mu^{t'}(E)$ for all $s' \in N^\varepsilon(s)$, $t' \in N^\varepsilon(t)$.

For $s' \in N^\varepsilon(s)$ and $t' \in N^\varepsilon(t)$, let $\sigma = [r', s', t']$ (where $r' = e - s' - t'$), so that $\sigma^{s'+t'} = [r', s' + t']$. Assume that $r' \neq 0$ (if $r' = 0$, take $\sigma = [s', t']$ instead; the proof for this case is similar). By hypothesis, $V^A(\sigma) = V^A(\sigma^{s'+t'})$. Abusing notation slightly, we may write $V^A(\sigma) = V^A(r') + V^A(s') + V^A(t')$ where, for arbitrary signals $\hat{s}$,

$$V^A(\hat{s}) := \sum_{\omega \in \Omega} \hat{s}_\omega \nu_\omega v(f^{\hat{s}}_\omega(A))$$

where $f^{\hat{s}}(A) \in \Delta c^{\hat{s}}(A)$ is the act at $\hat{s}$ specified by the stable selection. Similarly, we may write $V^A(\sigma^{s'+t'}) = V^A(r') + V^A(s'+t')$. Thus, $\sigma \sim^A \sigma^{s'+t'}$ if and only if $V^A(s') + V^A(t') = V^A(s'+t')$ (the same condition would hold if $r' = 0$). Observe that

$$\begin{aligned} V^A(s') + V^A(t') &= \left[\sum_{\omega \in E} s'_\omega \nu_\omega v(p) + \sum_{\omega' \in E^c} s'_{\omega'} \nu_{\omega'} v(q)\right] + \left[\sum_{\omega \in F} t'_\omega \nu_\omega v(p) + \sum_{\omega' \in F^c} t_{\omega'} \nu_{\omega'} v(q)\right] \\ &= v(p) \left[\sum_{\omega \in E} s'_\omega \nu_\omega + \sum_{\omega \in F} t'_\omega \nu_\omega\right] + v(q) \left[\sum_{\omega' \in E^c} s'_{\omega'} \nu_{\omega'} + \sum_{\omega' \in F^c} t'_{\omega'} \nu_{\omega'}\right]. \end{aligned}$$

Now consider $V^A(s' + t')$. There are three cases:

(1) $c^{s'+t'}(A) = pEq$. This means $\mu^{s'+t'}(E) > \mu^{s'+t'}(F)$. Then

$$V^A(s' + t') = \sum_{\omega \in E} (s'_\omega + t'_\omega) \nu_\omega v(p) + \sum_{\omega' \in E^c} (s'_{\omega'} + t'_{\omega'}) \nu_{\omega'} v(q)$$

and $\sigma \sim^A \sigma^{s'+t'}$ if and only if

$$\begin{aligned} & v(p) \left[\sum_{\omega \in F} t'_\omega \nu_\omega - \sum_{\omega \in E} t'_\omega \nu_\omega\right] + v(q) \left[\sum_{\omega' \in F^c} t'_{\omega'} \nu_{\omega'} - \sum_{\omega' \in E^c} t'_{\omega'} \nu_{\omega'}\right] = 0 \\ \Leftrightarrow \quad & v(p) \sum_{\omega \in F} t'_\omega \nu_\omega + v(q) \sum_{\omega' \in F^c} t'_{\omega'} \nu_{\omega'} = v(p) \sum_{\omega \in E} t'_\omega \nu_\omega + v(q) \sum_{\omega' \in E^c} t'_{\omega'} \nu_{\omega'}. \end{aligned}$$

In other words, Sender must be indifferent between pEq and pFq at signal t'. Since $v(p) \neq v(q)$, this is equivalent to $\nu^{t'}(E) = \nu^{t'}(F)$; that is, $\sum_{\omega \in E} t'_\omega \nu_\omega - \sum_{\omega \in F} t'_\omega \nu_\omega = 0$.

There is a hyperplane (in S) of such t'. Clearly, then, $N^\varepsilon(t)$ contains a signal t' that is not on that hyperplane. Hence, for small enough ε, $N^\varepsilon(t)$ contains a t' such that $\sigma \not\sim^A \sigma^{s'+t'}$.

(2) $c^{s'+t'}(A) = pFq$. This means $\mu^{s'+t'}(F) > \mu^{s'+t'}(E)$. Similar algebra to case (1) shows that $\sigma \sim^A \sigma^{s'+t'}$ if and only if $\nu^{s'}(E) = \nu^{s'}(F)$. Hence, for all $\varepsilon > 0$ sufficiently small, there exists $s' \in N^\varepsilon(s)$ such that $\sigma \not\sim^A \sigma^{s'+t'}$.

(3) $c^{s'+t'}(A) = \{pEq, pFq\}$. This means $\mu^{s'+t'}(E) = \mu^{s'+t'}(F)$. Clearly, small perturbations of s' and t' yield $\mu^{s'+t'}(E) \neq \mu^{s'+t'}(F)$, bringing us to either case (1) or case (2).

Thus, in all cases, we have $s' \in N^\varepsilon(s)$ and $t' \in N^\varepsilon(t)$ such that $\sigma \not\sim^A \sigma^{s'+t'}$ for ε sufficiently small, contradicting EF-equivalence.

We have shown that if s and t are EF-equivalent and $\mu^s(E) > \mu^s(F)$, then $\mu^t(E) \geq \mu^t(F)$. To establish that $\mu^t(E) > \mu^t(F)$, suppose toward a contradiction that $\mu^t(E) = \mu^t(F)$. Then every neighborhood $N^\varepsilon(t)$ contains a signal t' such that $\mu^{t'}(F) > \mu^{t'}(E)$. As shown above, this implies that $N^\varepsilon(t)$ contains a t' such that $\sigma \not\sim^A \sigma^{s+t'}$ for some EF-bet A and experiment σ with $s, t' \in \sigma$. This contradicts EF-equivalence of s and t. Thus, $\mu^t(E) > \mu^t(F)$. □

For the remainder of the proof, we take μ to be the unique $\tilde{\mu}$ elicited by Lemma 2. Signals s and t *disagree on the ranking of E and F* if either $[\mu^s(E) > \mu^s(F)$ and $\mu^t(F) > \mu^t(E)]$ or $[\mu^s(F) > \mu^s(E)$ and $\mu^t(E) > \mu^t(F)]$. Binary experiments $\sigma = [s, t]$ and $\sigma^\delta = [s+\delta, t-\delta]$ are an *EF-pair* if s is EF-equivalent to $s+\delta$, t is EF-equivalent to $t-\delta$, and s and t disagree on the ranking of E and F.

Lemma 3. *There is a unique full-support $\tilde{\nu} \in \Delta\Omega$ such that for all (w, w')-generic EF-bets A and EF-pairs σ and σ^δ, $\sum_{\omega\in E} \tilde{\nu}_\omega \delta_\omega = \sum_{\omega \in F} \tilde{\nu}_\omega \delta_\omega$ if $\sigma \sim^A \sigma^\delta$.*

Proof. First, note that for any (w, w')-generic bet $A = \{pEq, pFq\}$ and EF-pairs σ and σ^δ, we have $\sigma \sim^A \sigma^\delta$ if and only if

$$\sum_{\omega \in E} \nu_\omega \delta_\omega = \sum_{\omega\in F} \nu_\omega \delta_\omega \tag{1}$$

where ν is Sender's true prior. This follows from straightforward (but somewhat tedious) algebra using the representation V^A, the fact that (w, w')-genericity implies Sender is not indifferent between p and q, and the fact that signals in σ and σ^δ do not yield ties for Receiver (EF-equivalent signals lead to strict rankings of E and F).

Consider the case $F = E^c$. Then every δ satisfying (1) belongs to the kernel of the linear transformation $L : \delta \mapsto \sum_{\omega\in E} \nu_\omega\delta_\omega - \sum_{\omega\in E^c} \nu_\omega \delta_\omega$. Hence, the set of all such δ form a hyperplane with normal vector ν. We can elicit only those δ for which there is an EF-pair

σ and σ^δ; in particular, this means we can elicit those δ for which $\|\delta\|$ is sufficiently small (those δ within a particular neighborhood of 0). This is a full-dimensional subset of the hyperplane and, hence, sufficient to reveal its normal vector ν. $\square$

For the remainder of the proof, take ν to be the unique distribution elicited by Lemma 3. Having identified ν and μ, as well as candidates (w, w') for the utility indices, the next step is to identify which of w or w' is an (not necessarily positive) affine transformation of Receiver's utility index u.

For an arbitrary utility index $\tilde{w}$ and acts $f, g \in F$, write $\tilde{w}(f) \neq \tilde{w}(g)$ to indicate that $\tilde{w}(f_\omega) \neq \tilde{w}(g_\omega)$ for all ω. A binary menu $A = \{f, g\}$ is *$(\tilde{w}, \tilde{w}')$-regular* if $\tilde{w}(f) \neq \tilde{w}(g)$, $\tilde{w}'(f) \neq \tilde{w}'(g)$, and neither act dominates the other according to $\tilde{w}$ (that is, there exist ω, ω' such that $\tilde{w}(f_\omega) > \tilde{w}(g_\omega)$ and $\tilde{w}(g_{\omega'}) > \tilde{w}(f_{\omega'})$). It is straightforward to construct binary menus that are (w, w')-regular, and straightforward to construct ones that are (w', w)-regular.

Let A be a $(\tilde{w}, \tilde{w}')$-regular menu. Signals s and t are *A-equivalent* if there are neighborhoods $N(s)$ and $N(t)$ such that, for all $s' \in N(s)$ and $t' \in N(t)$ and all σ with $s', t' \in \sigma$, we have $\sigma \sim^A \sigma^{s'+t'}$. Signals s and t are $\tilde{w}$-equivalent at A if the maps $\tilde{f} \mapsto \sum_{\omega \in \Omega} \tilde{w}(\tilde{f}_\omega)\mu^s_\omega$ and $\tilde{f} \mapsto \sum_{\omega \in \Omega} \tilde{w}(\tilde{f}_\omega)\mu^t_\omega$ have the same (unique) maximizer $\tilde{f} \in A$. Finally, $(\tilde{w}, \tilde{w}')$ is *consistent with the A-equivalence regions* if there exist $s, t \in S$ such that $\{r \in S : r \text{ is } A\text{-equivalent to } s\} = \{r \in S : r \text{ and } s \text{ are } \tilde{w}\text{-equivalent}\}$ and $\{r \in S : r \text{ is } A\text{-equivalent to } t\} = \{r \in S : r \text{ and } t \text{ are } \tilde{w}\text{-equivalent}\}$.

Lemma 4. *If $(\tilde{w}, \tilde{w}')$ is consistent with the A-equivalence regions for all $(\tilde{w}, \tilde{w}')$-regular menus A (where either $(\tilde{w}, \tilde{w}') = (w, w')$ or $(\tilde{w}, \tilde{w}') = (w', w)$), then u is a (not necessarily positive) affine transformation of $\tilde{w}$.*

Proof. Let $U^s : f \mapsto \sum_{\omega \in \Omega} u(f_\omega)\mu^s_\omega$ denote Receiver's expected utility function for acts under μ and the true index u. In a (u, v)-regular menu A (where v is the true index for Sender), $U^s(f) \geq U^s(g)$ if and only if

$$\sum_{\omega \in \Omega} [u(f_\omega) - u(g_\omega)] s_\omega \mu_\omega \geq 0. \tag{2}$$

Thus, the set of signals s making Receiver indifferent between f and g is a hyperplane in S with normal vector $(u(f_\omega) - u(g_\omega))_{\omega \in \Omega}$. It therefore follows (by a similar argument to that of Lemma 2) that signals s and t are A-equivalent if and only if $[U^s(f) > U^s(g)$ and $U^t(f) > U^t(g)]$ or $[U^s(g) > U^s(f)$ and $U^t(g) > U^t(f)]$ (that is, s and t are u-equivalent).

Notice that the set (hyperplane) of signals satisfying (2) with equality is unaffected by replacing u with $-u$, so at least one of w or w' is a (not necessarily positive) affine transformation of u. If w and w' are cardinally distinct (that is, there do not exist $A, B \in \mathbb{R}$

such that $w' = Aw + B$), then only one of w or w' is an affine transformation of u, and different hyperplanes of signals correspond to indifference between f and g (for appropriate choices of f and g). Thus, if w and w' are cardinally distinct, only one pair (w, w') or (w', w) is consistent with the associated A-equivalence regions for all A. □

By Lemma 4, we may suppose without loss of generality that u is a (not necessarily positive) affine transformation of w. Then, using the agreement region employed by Lemma 1, v is a (not necessarily positive) affine transformation of w'. Hence, either (w', w) or $(-w', -w)$ are the true indices (up to positive affine transformation) because if one index is reversed, then so must the other in order to be consistent with the agreement region. For convenience, I relabel the two possibilities as (v, u) and $(-v, -u)$.

Lemma 5. *There is a menu A and experiments σ, σ' such that $\sigma \sim^A \sigma'$ if $\succsim$ is represented by (ν, μ, v, u) and $\sigma \not\sim^A \sigma'$ if $\succsim$ is represented by $(\nu, \mu, -v, -u)$.*

Proof. Pick $p^1, p^2, p^3 \in \Delta X$ such that $u(p^3) > u(p^2) > u(p^1)$ and $u(p^2) - u(p^1) > u(p^3) - u(p^2)$. We also require $v(p^3) \neq v(p^1)$; this is easily achieved by perturbing p^1 and p^3 along their respective indifference planes for u. Pick any $E = [\omega, \omega']$, $q \in \Delta X$, and $h' \in F$, and let $A = \{f, g, h\}$ where

$$\begin{aligned} f &= (\mu_{\omega'} p^1 + (1 - \mu_{\omega'})q, \mu_\omega p^3 + (1 - \mu_\omega)q)Eh' \\ g &= (\mu_{\omega'} p^2 + (1 - \mu_{\omega'})q, \mu_\omega p^2 + (1 - \mu_\omega)q)Eh' \\ h &= (\mu_{\omega'} p^3 + (1 - \mu_{\omega'})q, \mu_\omega p^1 + (1 - \mu_\omega)q)Eh' \end{aligned}$$

It is straightforward to verify that Receiver's choices from A (under μ and u) induce a symmetric division of S. In particular, $c^s(A) = A$ if $s_\omega = 0 = s_{\omega'}$; $c^s(A) = f$ if $\frac{s_{\omega'}}{s_\omega} > \frac{u(p^2)-u(p^1)}{u(p^3)-u(p^2)}$; $c^s(A) = h$ if $\frac{s_{\omega'}}{s_\omega} < \frac{u(p^3)-u(p^2)}{u(p^2)-u(p^1)}$; and $c^s(A) = g$ if $\frac{u(p^3)-u(p^2)}{u(p^2)-u(p^1)} < \frac{s_{\omega'}}{s_\omega} < \frac{u(p^2)-u(p^1)}{u(p^3)-u(p^2)}$ (these inequalities assume $s_\omega > 0$; at signal $e_{\omega'}$, Receiver chooses f). Since $u(p^2) - u(p^1) > u(p^3) - u(p^2)$, there is an interval of values for $\frac{s_{\omega'}}{s_\omega}$ where one act in A is strictly optimal. In particular, $c^e(A) = g$.

Consider an experiment $\sigma = [s, t]$ where $c^s(A) = c^t(A) = g$, $s_{\hat\omega} = t_{\hat\omega} = 1/2$ for all $\hat\omega \notin E$, and $s \neq t$. This can be done by taking $s = \frac{1}{2}e + \delta$, $t = \frac{1}{2}e - \delta$ for some $\delta \in \mathbb{R}^\Omega$ where $\delta_{\hat\omega} = 0$ for all $\hat\omega \notin E$. As long as $\|\delta\|$ is sufficiently small, we maintain $c^s(A) = c^t(A) = g$. Then $c^\sigma(A) = g$. Applying another such perturbation yields an experiment $\sigma' = [s', t']$ where $s' = s + \delta'$, $t' = t - \delta'$; for small $\|\delta'\|$, we once again have $c^{\sigma'}(A) = g$, so that $\sigma \sim^A \sigma'$.

Now consider behavior under $(-v, -u)$. Under utility index $-u$, Receiver satisfies $f \succ^s g$ if and only if $s_\omega \mu_\omega(-\mu_{\omega'} u(p^1)) + s_{\omega'} \mu_{\omega'}(-\mu_\omega u(p^3)) > s_\omega \mu_\omega(-\mu_{\omega'} u(p^2)) + s_{\omega'} \mu_{\omega'}(-\mu_\omega u(p^2))$; equivalently, $s_\omega(u(p^2) - u(p^1)) > s_{\omega'}(u(p^3) - u(p^2))$. Similarly, $g \succsim^s h$ if and only if $s_\omega(u(p^3)-$

$u(p^2)) \geq s_{\omega'}(u(p^2)-u(p^1))$. Since $u(p^2)-u(p^1) > u(p^3)-u(p^2)$, we get that $g \succsim^s h \Rightarrow f \succ^s g$. Thus, for all $s \in S$, $g \notin c^s(A)$. Similar algebra establishes that $f \succsim^s h$ iff $s_\omega \geq s_{\omega'}$ and $h \succsim^s f$ iff $s_{\omega'} \geq s_\omega$. Thus, under $-u$, only acts f and h are chosen by Receiver whenever $s_{\hat{\omega}} = 0$ for all $\hat{\omega} \notin E$ (if s has support E^c, then $c^s(A) = A$). With $\sigma = [s,t]$ and $\sigma' = [s',t'] = [s+\delta', t-\delta']$ as defined above, we may therefore assume that $c^s(A) = c^{s'}(A) = h$ and $c^t(A) = c^{t'}(A) = f$. The induced act for σ' is given by

$$c_{\hat{\omega}}^{\sigma'}(\sigma')(A) = \begin{cases} \mu_{\omega'}[(s_\omega + \delta'_\omega)p^1 + (t_\omega - \delta'_\omega)p^3] + (1-\mu_{\omega'})q & \text{if } \hat{\omega} = \omega \\ \mu_\omega[(s_{\omega'} + \delta'_{\omega'})p^3 + (t_{\omega'} - \delta'_{\omega'})p^1] + (1-\mu_\omega)q & \text{if } \hat{\omega} = \omega' \\ h'_{\hat{\omega}} & \text{if } \hat{\omega} \notin E \end{cases}$$

For $c^\sigma(A)$, set $\delta' = 0$. Straightforward algebra establishes that under $-v$, we have $\sigma \sim^A \sigma'$ if and only if

$$\nu_{\omega'}\mu_\omega\delta'_{\omega'}[v(p^1) - v(p^3)] = \nu_\omega\mu_{\omega'}\delta'_\omega[v(p^1) - v(p^3)]$$

Since $v(p^1) \neq v(p^3)$, this reduces to

$$\nu_{\omega'}\mu_\omega\delta'_{\omega'} = \nu_\omega\mu_{\omega'}\delta'_\omega$$

We are free to perturb δ'_ω and $\delta'_{\omega'}$ as needed because the only constraint on δ' is that $\|\delta'\|$ is small. Thus, there exists σ' such that $\sigma \not\sim^A \sigma'$ under $(-v,-u)$.

Since we have shown that $\sigma \sim^A \sigma'$ under (v,u) for all $\|\delta'\|$ small but $\sigma \not\sim^A \sigma'$ for some such δ' under $(-v,-u)$, we conclude that only one pair—(v,u) or $(-v,-u)$—can be consistent with $\succsim$. □

2 Proof of Proposition 1

Let $A = \{pEq, pFq\}$ be a generic bet. It follows immediately that Receiver is not indifferent between p and q: either $u(p) > u(q)$ or $u(q) > u(p)$. Suppose without loss of generality that $u(p) > u(q)$. If $v(p) \geq v(q)$, then σ^* maximizes V^A because Receiver's choices result in lottery p in every state. Thus, $\sigma^* \succsim^A \sigma$ for all σ. If $v(p) < v(q)$, an argument similar to that in the proof of Lemma 1 establishes that σ^* is strictly less-preferred than an experiment σ formed by perturbing signals in σ^* corresponding to states in $(E\backslash F) \cup (F\backslash E)$. Thus, σ^* is not top-ranked by $\succsim^A$.

This argument shows that in generic pq-bets A, σ^* is top-ranked by $\succsim^A$ if and only if Sender (at least weakly) agrees with Receiver's strict ranking of p and q. Thus, as the set of such pq-bets expands, Sender and Receiver agree on the ranking of a larger set of lotteries, proving (i). For (ii), observe that $v \approx u$ if and only if there are no lotteries p, q such that Sender disagrees with Receiver's strict ranking of p and q. Since v and u are non-constant, it follows that $v \approx u$ if and only if σ^* is top-ranked by $\succsim^A$ for all generic bets A.

3 Proof of Proposition 2

First, note that if A is a non-degenerate (p, q)-bet, then both $v(p) \neq v(q)$ and $u(p) \neq u(q)$.

A signal s belongs to the *EF-agreement region* if, for all non-degenerate EF-bets A, the sets

$$\left\{ f \in A : \sum_\omega v(f_\omega)\nu^s_\omega \geq \sum_\omega v(g_\omega)\nu^s_\omega \ \forall g \in A \right\}$$

and

$$\left\{ f \in A : \sum_\omega u(f_\omega)\mu^s_\omega \geq \sum_\omega u(g_\omega)\mu^s_\omega \ \forall g \in A \right\}$$

are singletons and have nonempty intersection (that is, there exists a unique act $f \in A$ that maximizes both $V^s : \tilde{f} \mapsto \sum_{\omega \in \Omega} v(\tilde{f}_\omega)\nu^s_\omega$ and $U^s : \tilde{f} \mapsto \sum_{\omega \in \Omega} u(\tilde{f}_\omega)\mu^s_\omega$). Otherwise, s is in the *EF-disagreement region.*

Lemma 6. *Suppose σ is EF-informative. If every $s \in \sigma$ belongs to the EF-agreement region, then σ is EF-extreme. Conversely, if σ is EF-extreme, then every $s \in \sigma$ belongs to the EF-agreement region.*

Proof. Suppose every $s \in \sigma$ belongs to the EF-agreement region. This means every $s \in \sigma$ makes ν^s and μ^s (strictly) agree on the ranking of E and F. Thus, there exists $\varepsilon > 0$ such that if $s \in \sigma$ and $s' \in N^\varepsilon(s)$, then s' belongs to the EF-agreement region (there is room to perturb each signal of σ while maintaining the strict ranking of E and F). Let $A = \{pEq, pFq\}$ be non-degenerate. Suppose Sender and Receiver agree on the ranking of p and q. Then Receiver's choices from A are single-valued for all $s' \in \sigma' \in N^\varepsilon(\sigma)$ and coincide with Sender's preferred act at such s' (Receiver's choice maximizes V^s). Thus, V^A coincides with the value of information for a standard Bayesian with prior ν and utility index v on $N^\varepsilon(\sigma)$, so that $\succsim^A$ is Blackwell monotone on $N^\varepsilon(\sigma)$. If instead Sender and Receiver disagree on the ranking of p, then V^A coincides with the value of information for a standard Bayesian with prior ν and utility index $-v$, so that $\succsim^A$ reverses the Blackwell order on $N^\varepsilon(\sigma)$.

For the converse, suppose σ is EF-extreme. Suppose toward a contradiction that there exists $s \in \sigma$ such that s is in the EF-disagreement region. Let $A = \{pEq, pFq\}$ be a non-degenerate EF-bet. Will consider the case where Sender and Receiver agree on the ranking of p and q (the case where they disagree is similar), so that σ^* is top-ranked by $\succsim^A$. If every

$s \in \sigma$ belongs to the disagreement region, then $\succsim^A$ (locally) reverses the Blackwell ordering on A (it is not difficult to construct a strict reversal), contradicting extremeness of σ.

Now suppose some (but not all) signals of σ belong to the disagreement region. We will find signals $s, t \in \sigma$ such that s belongs to the disagreement region, t belongs to the agreement region, and $c^s(A) \neq c^t(A)$. Pick some $t' \in \sigma$ that belongs to the agreement region. Since σ is EF-informative, there is at least one $s' \in \sigma$ such that $c^{s'}(A) \neq c^{t'}(A)$. If every such s' is in the disagreement region, take $t = t'$ and $s = s'$ for such an s'. Otherwise, every $s' \in \sigma$ such that $c^{s'}(A) \neq c^{t'}(A)$ is in the agreement region. Hence, the set of all $s'' \in \sigma$ such that $c^{s''}(A) = c^{t'}(A)$ intersects the disagreement region (recall that σ contains at least one signal in the disagreement region). So, there exists $s \in \sigma$ such that $c^s(A) = c^{t'}(A)$ and s belongs to the disagreement region. Then take t to be any $s' \in \sigma$ such that $c^{s'}(A) \neq c^{t'}(A)$; our choice of s and t satisfies all requirements.

We will now show that $\succsim^A$ violates the Blackwell information ordering. Without loss of generality, assume $c^s(A) = pEq$ and $c^t(A) = pFq$. Since s is in the disagreement region, this implies $\sum_{\omega \in F} \nu_\omega s_\omega - \sum_{\omega \in E} \nu_\omega s_\omega > 0$ (that is, Sender strictly prefers pFq at signal s). We may write $\sigma = [r^1, \ldots, r^K, t, s]$. Consider the garbling matrix M given by

$$M = \begin{bmatrix} I_K & 0 & \\ 0 & 1 & 0 \\ & 1-\alpha & \alpha \end{bmatrix}$$

where I_K denotes the $K \times K$ identity matrix. Then $\sigma' := \sigma M = [r_1, \ldots, r^K, t', s']$ where $t' = t + (1-\alpha)s$ and $s' = \alpha s$. Clearly, $c^{s'}(A) = c^s(A)$ and, for large enough $\alpha \in (0,1)$, $c^{t'}(A) = c^t(A)$. Thus, for $\alpha \in (0,1)$ sufficiently large,

$$\begin{aligned} \frac{V^A(\sigma') - V^A(\sigma)}{1-\alpha} &= -\left[\sum_{\omega \in E} \nu_\omega s_\omega v(p) + \sum_{\omega \in E^c} \nu_\omega s_\omega v(q)\right] + \left[\sum_{\omega \in F} \nu_\omega s_\omega v(p) + \sum_{\omega \in F^c} \nu_\omega s_\omega v(q)\right] \\ &= v(p)\left[\sum_{\omega \in F} \nu_\omega s_\omega - \sum_{\omega \in E} \nu_\omega s_\omega\right] - v(q)\left[\sum_{\omega \in E^c} \nu_\omega s_\omega - \sum_{\omega \in F^c} \nu_\omega s_\omega\right]. \end{aligned}$$

Dividing both sides by $\sum_{\omega \in \Omega} \nu_\omega s_\omega$, it follows that $V^A(\sigma') - V^A(\sigma) > 0$ if and only if

$$v(p)\left[P(F|s) - P(E|s)\right] - v(q)\left[(1 - P(E|s)) - (1 - P(F|s))\right] > 0$$

where $P(E|s)$ and $P(F|s)$ are conditional probabilities of events E and F, respectively, given prior ν and signal s. Thus, $V^A(\sigma') - V^A(\sigma) > 0$ if and only if

$$(v(p) - v(q))\,[P(F|s) - P(E|s)] > 0.$$

Since $v(p) > v(q)$, this is equivalent to $\sum_{\omega \in F} \nu_\omega s_\omega - \sum_{\omega \in E} \nu_\omega s_\omega > 0$. As demonstrated above, this condition is satisfied because s is in the disagreement region and $c^s(A) = pEq$. Thus, $V^A(\sigma') - V^A(\sigma) > 0$ and therefore $\sigma' \succ^A \sigma$. This violates the Blackwell information ordering because σ' is a garbling of σ. □

By Lemma 6, the EF-agreement region is characterized by EF-extreme experiments. Clearly, the EF-agreement region consists of signals s that make ν^s and μ^s strictly agree on the ranking of E and F. Thus, ν and μ are more aligned if and only if EF-agreement region expands (for all EF), which is equivalent to a larger set of EF-extreme experiments. This proves part (i) of the proposition.

For part (ii), suppose $\nu = \mu$ and let $A = \{pEq, pFq\}$ be a generic bet. By genericity, $u(p) \neq u(q)$. Suppose without loss of generality that $u(p) > u(q)$. If $v(p) \geq v(q)$, then Sender agrees with Receiver's choice at every signal s: Receiver finds pEq optimal if $\mu^s(E) \geq \mu^s(F)$, and Sender finds pEq optimal if $\nu^s(E) \geq \nu^s(F)$. Thus, regardless of Sender's stable selection at signals s such that $\nu^s(E) = \mu^s(E) = \mu^s(F) = \nu^s(F)$, $V^A(\sigma)$ coincides with that of a standard Bayesian with prior ν and utility index v. Thus, $\succsim^A$ satisfies the Blackwell ordering on all of $\mathcal{E}$. If instead $v(q) > v(p)$, then Receiver's (strict) choices disagree with Sender's, but (again) this has no effect at signals s that make posterior probabilities of E and F equal. Thus, V^A is the expected utility for a standard Bayesian with prior ν and utility index $-v$, so that $\succsim^A$ reverses the Blackwell ordering on all of $\mathcal{E}$.

For the converse, suppose $\nu \neq \mu$. Let $A = \{pEq, pFq\}$ be a bet where $u(p) > u(q)$. Hence, A is generic. Consider the case $v(p) > v(q)$ (the case $v(q) > v(p)$ is similar; by linearity of u and v, lotteries can be found that satisfy one of these two cases). Since Sender and Receiver agree on the ranking of p and q, σ^* is top-ranked by $\succsim^A$, and it is not difficult to find $\hat{\sigma}$ such that $\sigma^* \succ^A \hat{\sigma}$ (consistent with the Blackwell ordering). Since $\nu \neq \mu$, a similar construction to that in the proof of Lemma 6 yields experiments $\sigma \sqsupseteq \sigma'$ such that $\sigma' \succ^A \sigma$. Thus, $\succsim^A$ neither satisfies nor reverses the Blackwell ordering on all of $\mathcal{E}$.

4 Proof of Proposition 3

If $\nu = \mu$ and $v \approx u$, then Sender agrees with Receiver's choice(s) at every s in every menu A. Since Sender is indifferent between two or more acts whenever Receiver is (that is,

$V^s(f) = U^s(f)$ for all $s \in S$ and $f \in F$), Sender's expected payoff at (σ, A) does not depend on the choice of stable selection.

Conversely, suppose $\succsim^A$ satisfies the Blackwell ordering for all bets A. Then σ^* is top-ranked by $\succsim^A$ for all bets A, so that $v \approx u$ by Proposition 1. Similarly, Proposition 2 implies $\nu = \mu$ since $\succsim^A$ satisfies the Blackwell ordering in all bets A.

References

Jakobsen, A. (2020). An axiomatic model of persuasion. *Working Paper*.