Welfare from Pairwise Comparisons

What Can Any Alignment Rule Guarantee?

Aaron Roth

University of Pennsylvania

Fall 2026

Overview

- A maximal lottery is robust to every head-to-head challenger, but robustness is not yet a guarantee about population welfare.

Overview

- A maximal lottery is robust to every head-to-head challenger, but robustness is not yet a guarantee about population welfare.
- A heterogeneous Bradley–Terry model connects latent cardinal utility to the population comparison matrix.

Overview

- A maximal lottery is robust to every head-to-head challenger, but robustness is not yet a guarantee about population welfare.
- A heterogeneous Bradley–Terry model connects latent cardinal utility to the population comparison matrix.
- Maximal lotteries recover a quantifiable fraction of optimal welfare, with the loss determined by how nonlinear the sigmoid response model is.

Overview

- A maximal lottery is robust to every head-to-head challenger, but robustness is not yet a guarantee about population welfare.
- A heterogeneous Bradley–Terry model connects latent cardinal utility to the population comparison matrix.
- Maximal lotteries recover a quantifiable fraction of optimal welfare, with the loss determined by how nonlinear the sigmoid response model is.
- A matching lower bound shows that no rule seeing only aggregate pairwise win probabilities can guarantee more in general.

Pairwise Robustness and Welfare Ask Different Questions

Write p^{N} (N for Nash) for a maximal lottery, reserving the superscript $\star$ for welfare-optimal policies. Lecture 5 showed that

$$\mathbb{E}_{a \sim p^{\mathrm{N}},\, b \sim q}[P_{ab}] \geq \frac{1}{2} \qquad \text{for every } q \in \Delta(C).$$

Thus p^{N} defeats or ties every challenger in the population comparison game.

Pairwise Robustness and Welfare Ask Different Questions

Write p^{N} (N for Nash) for a maximal lottery, reserving the superscript $\star$ for welfare-optimal policies. Lecture 5 showed that

$$\mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q}[P_{ab}] \geq \frac{1}{2} \qquad \text{for every } q \in \Delta(C).$$

Thus p^{N} defeats or ties every challenger in the population comparison game.
But this statement concerns *choices*. It does not yet compare the welfare p^{N} obtains with the policy that maximizes users' average cardinal utility.

Pairwise Robustness and Welfare Ask Different Questions

Write p^{N} (N for Nash) for a maximal lottery, reserving the superscript $\star$ for welfare-optimal policies. Lecture 5 showed that

$$\mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q}[P_{ab}] \geq \frac{1}{2} \qquad \text{for every } q \in \Delta(C).$$

Thus p^{N} defeats or ties every challenger in the population comparison game.
But this statement concerns *choices*. It does not yet compare the welfare p^{N} obtains with the policy that maximizes users' average cardinal utility.

Today we ask how much welfare can be recovered from the comparison matrix P alone.

A Welfare Comparison Requires a Common Cardinal Scale

To average utility across people, the numbers must have the same meaning for every user. We therefore assume utilities are normalized to $[0, 1]$ and are interpersonally comparable: a change of 0.1 represents the same welfare change for every user. This is an additional welfare assumption; pairwise choices alone do not fix these units.

A Welfare Comparison Requires a Common Cardinal Scale

To average utility across people, the numbers must have the same meaning for every user. We therefore assume utilities are normalized to $[0, 1]$ and are interpersonally comparable: a change of 0.1 represents the same welfare change for every user. This is an additional welfare assumption; pairwise choices alone do not fix these units.

Each user has a utility vector

$$u = (u(a))_{a \in C} \in [0, 1]^C,$$

drawn from a population distribution $\mathcal{D}$.

A Welfare Comparison Requires a Common Cardinal Scale

To average utility across people, the numbers must have the same meaning for every user. We therefore assume utilities are normalized to $[0, 1]$ and are interpersonally comparable: a change of 0.1 represents the same welfare change for every user. This is an additional welfare assumption; pairwise choices alone do not fix these units.

Each user has a utility vector

$$u = (u(a))_{a \in C} \in [0, 1]^C,$$

drawn from a population distribution $\mathcal{D}$. The population welfare of a completion is

$$W(a) := \mathbb{E}_{u \sim \mathcal{D}}[u(a)].$$

Policy Welfare Averages Over Users and Outputs

The policy does not observe which user was drawn. Thus, for $p \in \Delta(C)$, draw its completion $A \sim p$ independently of $u \sim \mathcal{D}$. Its population welfare is

$$\begin{aligned} W(p) &:= \mathbb{E}_{u \sim \mathcal{D},\, A \sim p}[u(A)] \\ &= \sum_{a \in C} p_a \, \mathbb{E}_{u \sim \mathcal{D}}[u(a)] \\ &= \sum_{a \in C} p_a W(a). \end{aligned}$$

Policy Welfare Averages Over Users and Outputs

The policy does not observe which user was drawn. Thus, for $p \in \Delta(C)$, draw its completion $A \sim p$ independently of $u \sim \mathcal{D}$. Its population welfare is

$$\begin{aligned} W(p) &:= \mathbb{E}_{u \sim \mathcal{D},\, A \sim p}[u(A)] \\ &= \sum_{a \in C} p_a \, \mathbb{E}_{u \sim \mathcal{D}}[u(a)] \\ &= \sum_{a \in C} p_a W(a). \end{aligned}$$

Let

$$p^\star_W \in \arg\max_{p \in \Delta(C)} W(p).$$

Because W is linear, a welfare-optimal policy can put all its mass on a completion with maximum $W(a)$.

We Observe Noisy Pairwise Choices

When a user with utility vector u compares a and b, suppose that a is chosen with probability

$$\sigma(\beta(u(a) - u(b))), \qquad \sigma(z) := \frac{1}{1 + e^{-z}}.$$

The parameter $\beta > 0$ controls choice noise: larger β makes the user more likely to choose the completion with higher utility.

We Observe Noisy Pairwise Choices

When a user with utility vector u compares a and b, suppose that a is chosen with probability

$$\sigma(\beta(u(a) - u(b)))\,, \qquad \sigma(z) := \frac{1}{1 + e^{-z}}.$$

The parameter $\beta > 0$ controls choice noise: larger β makes the user more likely to choose the completion with higher utility. Drawing a new user independently from $\mathcal{D}$ for each comparison—and retaining no user identity—gives the population probability

$$P_{ab} = \mathbb{E}_{u \sim \mathcal{D}}\left[\sigma(\beta(u(a) - u(b)))\right].$$

We Observe Noisy Pairwise Choices

When a user with utility vector u compares a and b, suppose that a is chosen with probability

$$\sigma(\beta(u(a) - u(b))), \qquad \sigma(z) := \frac{1}{1 + e^{-z}}.$$

The parameter $\beta > 0$ controls choice noise: larger β makes the user more likely to choose the completion with higher utility. Drawing a new user independently from $\mathcal{D}$ for each comparison—and retaining no user identity—gives the population probability

$$P_{ab} = \mathbb{E}_{u \sim \mathcal{D}} \left[\sigma(\beta(u(a) - u(b))) \right].$$

We assume in this lecture exact knowledge of the entire matrix P, ignoring sampling error so that we can isolate what P identifies about welfare.

Distortion Measures Welfare Lost by the Rule

For an output policy p with positive welfare, define

$$\mathrm{dist}_{\mathcal{D}}(p) := \frac{W(p^{\star}_{W})}{W(p)}.$$

If $W(p) = 0 < W(p^{\star}_{W})$, its distortion is infinite.

Distortion Measures Welfare Lost by the Rule

For an output policy p with positive welfare, define

$$\mathrm{dist}_{\mathcal{D}}(p) := \frac{W(p^\star_W)}{W(p)}.$$

If $W(p) = 0 < W(p^\star_W)$, its distortion is infinite.
A worst-case distortion guarantee has the form

$$W(p) \geq \frac{1}{\rho} W(p^\star_W) \qquad \text{for every population } \mathcal{D}$$

satisfying the heterogeneous Bradley–Terry model above.

Distortion Measures Welfare Lost by the Rule

For an output policy p with positive welfare, define

$$\mathrm{dist}_{\mathcal{D}}(p) := \frac{W(p^\star_W)}{W(p)}.$$

If $W(p) = 0 < W(p^\star_W)$, its distortion is infinite.
A worst-case distortion guarantee has the form

$$W(p) \geq \frac{1}{\rho} W(p^\star_W) \qquad \text{for every population } \mathcal{D}$$

satisfying the heterogeneous Bradley–Terry model above.

Smaller ρ is better; $\rho = 1$ means welfare optimality.

A Linear Rule Would Reveal Welfare Differences

Imagine that a user with utility vector u chose a over b with probability

$$\Pr(a \succ b \mid u) = \frac{1}{2} + c\big(u(a) - u(b)\big), \qquad c \in (0, 1/2].$$

Because utility differences lie in $[-1, 1]$, this range of c keeps the right-hand side between zero and one.

A Linear Rule Would Reveal Welfare Differences

Imagine that a user with utility vector u chose a over b with probability

$$\Pr(a \succ b \mid u) = \frac{1}{2} + c\big(u(a) - u(b)\big), \qquad c \in (0, 1/2].$$

Because utility differences lie in $[-1, 1]$, this range of c keeps the right-hand side between zero and one. Averaging this probability over users would give

$$\begin{aligned} P_{ab} - \frac{1}{2} &= c\,\mathbb{E}_u\big[u(a) - u(b)\big] \\ &= c\big(W(a) - W(b)\big). \end{aligned}$$

Thus P_{ab} would reveal both the sign and the scaled magnitude of the population welfare difference.

The Sigmoid Can Hide Welfare Differences

The Bradley–Terry model instead uses the nonlinear sigmoid. Consequently, in general

$$\begin{aligned} P_{ab} &= \mathbb{E}_u[\sigma(\beta(u(a) - u(b)))] \\ &\neq \sigma(\beta \mathbb{E}_u[u(a) - u(b)]) = \sigma(\beta(W(a) - W(b)))\,. \end{aligned}$$

This failure can be severe: P_{ab} need not reveal even the sign of $W(a) - W(b)$.

The Sigmoid Can Hide Welfare Differences

The Bradley–Terry model instead uses the nonlinear sigmoid. Consequently, in general

$$\begin{aligned} P_{ab} &= \mathbb{E}_u[\sigma(\beta(u(a) - u(b)))] \\ &\neq \sigma(\beta\mathbb{E}_u[u(a) - u(b)]) = \sigma(\beta(W(a) - W(b)))\,. \end{aligned}$$

This failure can be severe: P_{ab} need not reveal even the sign of $W(a) - W(b)$. The lower-bound construction takes advantage of this with instances on which: $P_{ab} = 1/2$ even though $W(a) > W(b)$.

The Sigmoid Can Hide Welfare Differences

The Bradley–Terry model instead uses the nonlinear sigmoid. Consequently, in general

$$\begin{aligned} P_{ab} &= \mathbb{E}_u[\sigma(\beta(u(a) - u(b)))] \\ &\neq \sigma(\beta \mathbb{E}_u[u(a) - u(b)]) = \sigma(\beta(W(a) - W(b))) \,. \end{aligned}$$

This failure can be severe: P_{ab} need not reveal even the sign of $W(a) - W(b)$. The lower-bound construction takes advantage of this with instances on which: $P_{ab} = 1/2$ even though $W(a) > W(b)$.

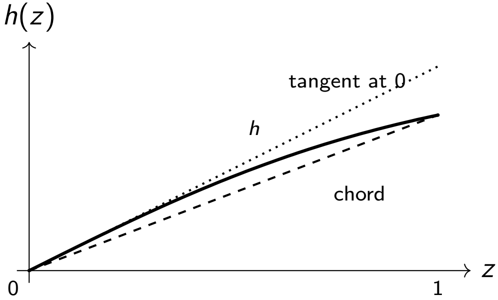
We need linearity to average across users; we'll use the tightest linear upper and lower bounds.

A Chord and a Tangent Bound the Sigmoid

For $z \in [0, 1]$, let $h(z) := \sigma(\beta z) - 1/2$. For $z \geq 0$,

$$\sigma''(z) = \sigma(z)(1 - \sigma(z))(1 - 2\sigma(z)) \leq 0.$$

Thus $h''(z) = \beta^2 \sigma''(\beta z) \leq 0$, so h is concave; also $h(0) = 0$. The figure uses $\beta = 2$.



The dashed chord joins $z = 0$ and $z = 1$—equivalently, sigmoid inputs $\beta z = 0$ and $\beta z = \beta$. Concavity puts h above this chord and below its tangent at 0.

The Chord and Tangent Bound Every Positive Gap

The figure uses the coordinate z, but the sigmoid input is $x = \beta z$. With respect to x, the tangent and chord slopes are

$$\bar{s} := \sigma'(0) = \sigma(0)(1 - \sigma(0)) = \frac{1}{4}, \qquad \underline{s}_\beta := \frac{\sigma(\beta) - 1/2}{\beta}.$$

Multiplying by $dx/dz = \beta$, the corresponding slopes in the figure are $\beta\bar{s}$ and $\beta\underline{s}_\beta$.

The Chord and Tangent Bound Every Positive Gap

The figure uses the coordinate z, but the sigmoid input is $x = \beta z$. With respect to x, the tangent and chord slopes are

$$\bar{s} := \sigma'(0) = \sigma(0)(1 - \sigma(0)) = \frac{1}{4}, \qquad \underline{s}_\beta := \frac{\sigma(\beta) - 1/2}{\beta}.$$

Multiplying by $dx/dz = \beta$, the corresponding slopes in the figure are $\beta\bar{s}$ and $\beta\underline{s}_\beta$. Concavity gives, for every $z \in [0, 1]$,

$$\boxed{\beta\underline{s}_\beta z \leq \sigma(\beta z) - \frac{1}{2} \leq \beta\bar{s}z.}$$

The Chord and Tangent Bound Every Positive Gap

The figure uses the coordinate z, but the sigmoid input is $x = \beta z$. With respect to x, the tangent and chord slopes are

$$\bar{s} := \sigma'(0) = \sigma(0)(1 - \sigma(0)) = \frac{1}{4}, \qquad \underline{s}_\beta := \frac{\sigma(\beta) - 1/2}{\beta}.$$

Multiplying by $dx/dz = \beta$, the corresponding slopes in the figure are $\beta\bar{s}$ and $\beta\underline{s}_\beta$. Concavity gives, for every $z \in [0, 1]$,

$$\boxed{\beta\underline{s}_\beta z \leq \sigma(\beta z) - \frac{1}{2} \leq \beta\bar{s}z.}$$

Equivalently, for every positive utility gap z,

$$\beta\underline{s}_\beta \leq \frac{\sigma(\beta z) - 1/2}{z} \leq \beta\bar{s}.$$

The fraction is the win-probability margin per unit utility gap.

The Conversion Rate Depends on the Gap Size

For a positive utility gap t, define

$$w_\beta(t) := \frac{\sigma(\beta t) - 1/2}{t} \quad (t > 0).$$

This is the win-probability margin above $1/2$ *per unit of utility gap*.

The Conversion Rate Depends on the Gap Size

For a positive utility gap t, define

$$w_\beta(t) := \frac{\sigma(\beta t) - 1/2}{t} \quad (t > 0).$$

This is the win-probability margin above $1/2$ *per unit of utility gap*. The previous slide then gives

$$\beta \underline{s}_\beta \le w_\beta(t) \le \beta \overline{s} \qquad (0 \le t \le 1).$$

Thus the chord and tangent bound the rate at which utility gaps become comparison margins.

Nonlinearity Reweights Users Before We Average

For a fixed pair a, b, write $\Delta_u := u(a) - u(b)$. For user u, the comparison margin favoring a over b is

$$\sigma(\beta\Delta_u) - \frac{1}{2} = w_\beta(|\Delta_u|)\Delta_u.$$

For $\Delta_u < 0$, the equality uses $\sigma(-z) - 1/2 = -(\sigma(z) - 1/2)$.

Nonlinearity Reweights Users Before We Average

For a fixed pair a, b, write $\Delta_u := u(a) - u(b)$. For user u, the comparison margin favoring a over b is

$$\sigma(\beta\Delta_u) - \frac{1}{2} = w_\beta(|\Delta_u|)\Delta_u.$$

For $\Delta_u < 0$, the equality uses $\sigma(-z) - 1/2 = -(\sigma(z) - 1/2)$. Averaging these individual margins over users gives

$$P_{ab} - \frac{1}{2} = \mathbb{E}_u[w_\beta(|\Delta_u|)\Delta_u], \qquad W(a) - W(b) = \mathbb{E}_u[\Delta_u].$$

Nonlinearity Reweights Users Before We Average

For a fixed pair a, b, write $\Delta_u := u(a) - u(b)$. For user u, the comparison margin favoring a over b is

$$\sigma(\beta\Delta_u) - \frac{1}{2} = w_\beta(|\Delta_u|)\Delta_u.$$

For $\Delta_u < 0$, the equality uses $\sigma(-z) - 1/2 = -(\sigma(z) - 1/2)$. Averaging these individual margins over users gives

$$P_{ab} - \frac{1}{2} = \mathbb{E}_u[w_\beta(|\Delta_u|)\Delta_u], \qquad W(a) - W(b) = \mathbb{E}_u[\Delta_u].$$

Thus comparison data average *reweighted* utility gaps, rather than the utility gaps themselves.

Maximality Gives an Inequality We Want to Translate

A maximal lottery p^{N} defeats or ties every challenger q, so

$$\mathbb{E}_{a \sim p^{\mathrm{N}},\, b \sim q}\left[P_{ab} - \frac{1}{2}\right] \geq 0 \qquad \text{for every } q \in \Delta(C).$$

Maximality Gives an Inequality We Want to Translate

A maximal lottery p^{N} defeats or ties every challenger q, so

$$\mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q}\left[P_{ab} - \frac{1}{2}\right] \geq 0 \qquad \text{for every } q \in \Delta(C).$$

Recall that $h(z) := \sigma(\beta z) - 1/2$, so $P_{ab} - 1/2$ averages the individual comparison margin $h(u(a) - u(b))$. Hence maximality says

$$0 \leq \mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q,\, u\sim\mathcal{D}}\left[h\big(u(a) - u(b)\big)\right].$$

Maximality Gives an Inequality We Want to Translate

A maximal lottery p^{N} defeats or ties every challenger q, so

$$\mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q}\left[P_{ab}-\frac{1}{2}\right]\geq 0 \qquad \text{for every } q\in\Delta(C).$$

Recall that $h(z) := \sigma(\beta z) - 1/2$, so $P_{ab} - 1/2$ averages the individual comparison margin $h(u(a)-u(b))$. Hence maximality says

$$0 \leq \mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim q,\, u\sim\mathcal{D}}\left[h\big(u(a)-u(b)\big)\right].$$

To deduce a welfare inequality, we need an *upper* bound whose expectation separates into a term for a and a term for b. We will prove that, for $x, y \in [0,1]$,

$$\boxed{h(x-y)\leq \beta\big(\overline{s}x-\underline{s}_{\beta}y\big).}$$

After averaging, $x = u(a)$ will become $W(p^{\mathrm{N}})$ and $y = u(b)$ will become $W(q)$.

First Case: Completion a Has Higher Utility

Suppose $x \geq y$, so the utility gap $x - y$ is nonnegative. The tangent side of the chord–tangent bound gives

$$h(x - y) \leq \beta\bar{s}(x - y).$$

First Case: Completion a Has Higher Utility

Suppose $x \geq y$, so the utility gap $x - y$ is nonnegative. The tangent side of the chord–tangent bound gives

$$h(x - y) \leq \beta\bar{s}(x - y).$$

Now separate the two utility levels:

$$\begin{aligned} h(x - y) &\leq \beta\big(\bar{s}x - \bar{s}y\big) \\ &\leq \beta\big(\bar{s}x - \underline{s}_\beta y\big). \end{aligned}$$

The last step uses $\bar{s} \geq \underline{s}_\beta$ and $y \geq 0$: replacing $-\bar{s}y$ by $-\underline{s}_\beta y$ makes the upper bound larger.

Second Case: Completion b Has Higher Utility

Now suppose $x < y$, and write $d := y - x > 0$. Oddness of the sigmoid margin gives

$$h(x - y) = h(-d) = -h(d).$$

Second Case: Completion b Has Higher Utility

Now suppose $x < y$, and write $d := y - x > 0$. Oddness of the sigmoid margin gives

$$h(x - y) = h(-d) = -h(d).$$

The chord bound says $h(d) \geq \beta \underline{s}_\beta d$. Negating it therefore gives

$$h(x - y) = -h(d) \leq -\beta \underline{s}_\beta d = \beta \underline{s}_\beta (x - y).$$

Second Case: Completion b Has Higher Utility

Now suppose $x < y$, and write $d := y - x > 0$. Oddness of the sigmoid margin gives

$$h(x - y) = h(-d) = -h(d).$$

The chord bound says $h(d) \geq \beta \underline{s}_\beta d$. Negating it therefore gives

$$h(x - y) = -h(d) \leq -\beta \underline{s}_\beta d = \beta \underline{s}_\beta (x - y).$$

Finally,

$$\underline{s}_\beta (x - y) = \underline{s}_\beta x - \underline{s}_\beta y \leq \bar{s} x - \underline{s}_\beta y,$$

because $\bar{s} \geq \underline{s}_\beta$ and $x \geq 0$. This proves the needed upper bound in the second case.

Symmetry Gives the Lower Bound

We have proved, for every $x, y \in [0,1]$,

$$h(x-y) \leq \beta\big(\bar{s}x - \underline{s}_\beta y\big).$$

Symmetry Gives the Lower Bound

We have proved, for every $x, y \in [0,1]$,

$$h(x - y) \leq \beta\big(\bar{s}x - \underline{s}_\beta y\big).$$

Apply this upper bound with x and y interchanged:

$$h(y - x) \leq \beta\big(\bar{s}y - \underline{s}_\beta x\big).$$

Since $h(y - x) = -h(x - y)$, negating both sides gives

$$h(x - y) \geq \beta\big(\underline{s}_\beta x - \bar{s}y\big).$$

Symmetry Gives the Lower Bound

We have proved, for every $x, y \in [0, 1]$,

$$h(x - y) \le \beta\big(\bar{s}x - \underline{s}_\beta y\big).$$

Apply this upper bound with x and y interchanged:

$$h(y - x) \le \beta\big(\bar{s}y - \underline{s}_\beta x\big).$$

Since $h(y - x) = -h(x - y)$, negating both sides gives

$$h(x - y) \ge \beta\big(\underline{s}_\beta x - \bar{s}y\big).$$

Together, the two inequalities give

$$\boxed{\beta(\underline{s}_\beta x - \bar{s}y) \le h(x - y) \le \beta(\bar{s}x - \underline{s}_\beta y).}$$

Averaging Over Users Gives Completion Welfare

For fixed completions a, b, substitute $x = u(a)$ and $y = u(b)$ into the two-sided bound and average over $u \sim \mathcal{D}$.

Averaging Over Users Gives Completion Welfare

For fixed completions a, b, substitute $x = u(a)$ and $y = u(b)$ into the two-sided bound and average over $u \sim \mathcal{D}$. The middle term becomes

$$\mathbb{E}_u\left[h\big(u(a) - u(b)\big)\right] = P_{ab} - \frac{1}{2},$$

while linearity of expectation gives

$$\mathbb{E}_u[u(a)] = W(a), \qquad \mathbb{E}_u[u(b)] = W(b).$$

Therefore

$$\boxed{\begin{aligned} \beta\big(\underline{s}_\beta W(a) - \overline{s} W(b)\big) &\le P_{ab} - \frac{1}{2} \\ &\le \beta\big(\overline{s} W(a) - \underline{s}_\beta W(b)\big). \end{aligned}}$$

Averaging Gives Policy Welfare

Now draw $a \sim p$ and $b \sim q$ independently and average the completion-level inequality. Because

$$\mathbb{E}_{a \sim p}[W(a)] = W(p), \qquad \mathbb{E}_{b \sim q}[W(b)] = W(q),$$

we obtain

Averaging Gives Policy Welfare

Now draw $a \sim p$ and $b \sim q$ independently and average the completion-level inequality. Because

$$\mathbb{E}_{a\sim p}[W(a)] = W(p), \qquad \mathbb{E}_{b\sim q}[W(b)] = W(q),$$

we obtain

$$\boxed{\begin{aligned} \beta\bigl(\underline{s}_\beta W(p) - \overline{s} W(q)\bigr) &\le \mathbb{E}_{a\sim p,\, b\sim q}\left[P_{ab} - \frac{1}{2}\right] \\ &\le \beta\bigl(\overline{s} W(p) - \underline{s}_\beta W(q)\bigr). \end{aligned}}$$

Averaging Gives Policy Welfare

Now draw $a \sim p$ and $b \sim q$ independently and average the completion-level inequality. Because

$$\mathbb{E}_{a\sim p}[W(a)] = W(p), \qquad \mathbb{E}_{b\sim q}[W(b)] = W(q),$$

we obtain

$$\boxed{\begin{aligned} \beta\big(\underline{s}_\beta W(p) - \overline{s} W(q)\big) &\le \mathbb{E}_{a\sim p,\, b\sim q}\left[P_{ab} - \frac{1}{2}\right] \\ &\le \beta\big(\overline{s} W(p) - \underline{s}_\beta W(q)\big). \end{aligned}}$$

The nonlinear comparison expression has now become a pair of linear welfare bounds for policies p and q.

Maximal Lotteries Guarantee a Γ_β Welfare Approximation

The chord and tangent slopes enter the final guarantee through their ratio:

$$\Gamma_\beta := \frac{\bar{s}}{\underline{s}_\beta} = \frac{\beta}{4\big(\sigma(\beta) - 1/2\big)}.$$

Maximal Lotteries Guarantee a Γ_β Welfare Approximation

The chord and tangent slopes enter the final guarantee through their ratio:

$$\Gamma_\beta := \frac{\bar{s}}{\underline{s}_\beta} = \frac{\beta}{4\big(\sigma(\beta) - 1/2\big)}.$$

Theorem

Under the heterogeneous Bradley–Terry model,

$$W(p^{\mathrm{N}}) \geq \frac{1}{\Gamma_\beta} W(p^\star_W).$$

Equivalently, $\mathrm{dist}_{\mathcal{D}}(p^{\mathrm{N}}) \leq \Gamma_\beta$.

Maximal Lotteries Guarantee a Γ_β Welfare Approximation

The chord and tangent slopes enter the final guarantee through their ratio:

$$\Gamma_\beta := \frac{\bar{s}}{\underline{s}_\beta} = \frac{\beta}{4\big(\sigma(\beta) - 1/2\big)}.$$

Theorem

Under the heterogeneous Bradley–Terry model,

$$W(p^{\mathrm{N}}) \geq \frac{1}{\Gamma_\beta} W(p^\star_W).$$

Equivalently, $\mathrm{dist}_{\mathcal{D}}(p^{\mathrm{N}}) \leq \Gamma_\beta$.

The proof has one step: compare p^{N} directly with $p^\star_W$ and convert that one comparison into welfare.

One Comparison Gives the Entire Welfare Guarantee

Maximality holds against every challenger. In particular, use the welfare-optimal policy $q = p^\star_W$. The policy-level upper bound from the previous slides then gives

$$
\begin{aligned}
0 &\le \mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim p^\star_W}\left[P_{ab} - \frac{1}{2}\right] \\
&\le \beta\left(\bar{s}W(p^{\mathrm{N}}) - \underline{s}_\beta W(p^\star_W)\right).
\end{aligned}
$$

One Comparison Gives the Entire Welfare Guarantee

Maximality holds against every challenger. In particular, use the welfare-optimal policy $q = p^\star_W$. The policy-level upper bound from the previous slides then gives

$$0 \le \mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim p^\star_W}\left[P_{ab} - \frac{1}{2}\right] \le \beta\left(\bar{s}W(p^{\mathrm{N}}) - \underline{s}_\beta W(p^\star_W)\right).$$

The first inequality is the maximal-lottery guarantee; the second is exactly the conversion from comparisons to policy welfare that we just proved.

One Comparison Gives the Entire Welfare Guarantee

Maximality holds against every challenger. In particular, use the welfare-optimal policy $q = p_W^\star$. The policy-level upper bound from the previous slides then gives

$$0 \leq \mathbb{E}_{a\sim p^{\mathrm{N}},\, b\sim p_W^\star}\left[P_{ab} - \frac{1}{2}\right] \leq \beta\left(\bar{s}W(p^{\mathrm{N}}) - \underline{s}_\beta W(p_W^\star)\right).$$

The first inequality is the maximal-lottery guarantee; the second is exactly the conversion from comparisons to policy welfare that we just proved. Therefore

$$W(p^{\mathrm{N}}) \geq \frac{\underline{s}_\beta}{\bar{s}} W(p_W^\star) = \frac{1}{\Gamma_\beta} W(p_W^\star). \qquad \square$$

The use of $p_W^\star$ is only a proof comparison: the alignment rule need not know $p_W^\star$ or observe users' cardinal utilities.

One Expected Comparison Suffices

maximal-lottery inequality p^{N} does not lose to $p^{\star}_{W}$ $\longrightarrow$ upper bound on $P_{ab} - 1/2$ $\longrightarrow$ welfare guarantee $W(p^{\mathrm{N}}) \geq W(p^{\star}_{W})/\Gamma_{\beta}$

One Expected Comparison Suffices

maximal-lottery inequality p^{N} does not lose to $p^\star_W$ → upper bound on $P_{ab} - 1/2$ → welfare guarantee $W(p^{\mathrm{N}}) \geq W(p^\star_W)/\Gamma_\beta$

The factor Γ_β is the price of translating the expected comparison margin into a statement about unobserved cardinal welfare.

The Welfare Loss Grows as the Sigmoid Saturates

Recall that a utility gap of magnitude t receives weight $w_\beta(t)$ before being averaged, where

$$\beta \underline{s}_\beta \leq w_\beta(t) \leq \beta \overline{s}.$$

Thus $\Gamma_\beta = \overline{s}/\underline{s}_\beta$ is the largest possible ratio between the weights placed on two users' utility gaps.

The Welfare Loss Grows as the Sigmoid Saturates

Recall that a utility gap of magnitude t receives weight $w_\beta(t)$ before being averaged, where

$$\beta \underline{s}_\beta \le w_\beta(t) \le \beta \overline{s}.$$

Thus $\Gamma_\beta = \overline{s}/\underline{s}_\beta$ is the largest possible ratio between the weights placed on two users' utility gaps.
Numerically,

β	0.5	1	2	5	10
Γ_β	1.02	1.08	1.31	2.53	5.00

The Welfare Loss Grows as the Sigmoid Saturates

Recall that a utility gap of magnitude t receives weight $w_\beta(t)$ before being averaged, where

$$\beta \underline{s}_\beta \leq w_\beta(t) \leq \beta \overline{s}.$$

Thus $\Gamma_\beta = \overline{s}/\underline{s}_\beta$ is the largest possible ratio between the weights placed on two users' utility gaps.
Numerically,

β	0.5	1	2	5	10
Γ_β	1.02	1.08	1.31	2.53	5.00

For small β, the sigmoid is nearly linear and $\Gamma_\beta \to 1$. For large β, $\Gamma_\beta \sim \beta/2$.

Is Γ_β Merely a Weakness of the Proof?

Perhaps a different rule could extract more welfare from the same aggregate comparison data P.

Is Γ_β Merely a Weakness of the Proof?

Perhaps a different rule could extract more welfare from the same aggregate comparison data P.
To rule this out, we will construct many different welfare populations that all produce exactly the same observable matrix:

$$P_{ij} = \frac{1}{2} \qquad \text{for every pair } i, j.$$

Is Γ_β Merely a Weakness of the Proof?

Perhaps a different rule could extract more welfare from the same aggregate comparison data P.

To rule this out, we will construct many different welfare populations that all produce exactly the same observable matrix:

$$P_{ij} = \frac{1}{2} \qquad \text{for every pair } i, j.$$

An alignment rule that sees only P cannot tell which completion has high welfare.

Two User Types Hide One Special Completion

Let

$$C = \{a, b_1, \ldots, b_{m-1}\}.$$

For a small $\varepsilon > 0$, consider two user types:

	$u(a)$	$u(b_j)$ for every j
Type A	1	0
Type B	0	ε

Let Type A have probability α and Type B probability $1 - \alpha$.

Two User Types Hide One Special Completion

Let

$$C = \{a, b_1, \ldots, b_{m-1}\}.$$

For a small $\varepsilon > 0$, consider two user types:

	$u(a)$	$u(b_j)$ for every j
Type A	1	0
Type B	0	ε

Let Type A have probability α and Type B probability $1 - \alpha$. The special completion a has welfare α; every b_j has welfare $\varepsilon(1 - \alpha)$.

Choose the Mixture to Make Every Comparison Tied

Write

$$s_1 := \sigma(\beta) - \frac{1}{2}, \qquad s_\varepsilon := \sigma(\beta\varepsilon) - \frac{1}{2}.$$

Type A favors a over b_j by a utility gap of 1, whereas Type B favors b_j over a by a gap of ε. Their comparison margins are therefore s_1 and $-s_\varepsilon$, respectively. Hence, for every j,

$$P_{ab_j} - \frac{1}{2} = \alpha s_1 - (1-\alpha)s_\varepsilon.$$

Choose the Mixture to Make Every Comparison Tied

Write

$$s_1 := \sigma(\beta) - \frac{1}{2}, \qquad s_\varepsilon := \sigma(\beta\varepsilon) - \frac{1}{2}.$$

Type A favors a over b_j by a utility gap of 1, whereas Type B favors b_j over a by a gap of ε. Their comparison margins are therefore s_1 and $-s_\varepsilon$, respectively. Hence, for every j,

$$P_{ab_j} - \frac{1}{2} = \alpha s_1 - (1-\alpha)s_\varepsilon.$$

Choose α so that this comparison margin is zero:

$$\alpha = \frac{s_\varepsilon}{s_1 + s_\varepsilon} = \frac{\sigma(\beta\varepsilon) - 1/2}{\sigma(\beta) + \sigma(\beta\varepsilon) - 1}.$$

Choose the Mixture to Make Every Comparison Tied

Write

$$s_1 := \sigma(\beta) - \frac{1}{2}, \qquad s_\varepsilon := \sigma(\beta\varepsilon) - \frac{1}{2}.$$

Type A favors a over b_j by a utility gap of 1, whereas Type B favors b_j over a by a gap of ε. Their comparison margins are therefore s_1 and $-s_\varepsilon$, respectively. Hence, for every j,

$$P_{ab_j} - \frac{1}{2} = \alpha s_1 - (1-\alpha)s_\varepsilon.$$

Choose α so that this comparison margin is zero:

$$\alpha = \frac{s_\varepsilon}{s_1 + s_\varepsilon} = \frac{\sigma(\beta\varepsilon) - 1/2}{\sigma(\beta) + \sigma(\beta\varepsilon) - 1}.$$

All users are indifferent among the b_j's, so $P_{b_j b_k} = 1/2$ as well. Thus the entire observable matrix is the fair-coin matrix.

The Same Matrix Can Hide Any Alternative

Consider any rule that receives only the aggregate matrix P and returns a policy p. On the fair-coin matrix, because $\sum_i p_i = 1$, some alternative receives probability at most $1/m$.

The Same Matrix Can Hide Any Alternative

Consider any rule that receives only the aggregate matrix P and returns a policy p. On the fair-coin matrix, because $\sum_i p_i = 1$, some alternative receives probability at most $1/m$.
Every relabeling of the two-type construction produces the same matrix P. We can therefore make an alternative with

$$p_a \leq \frac{1}{m}$$

the high-welfare special completion a without changing what the rule observes (and hence without changing the chosen policy).

The Hidden Welfare Ratio Approaches Γ_β

The ratio between the special completion and each b_j is

$$r_\varepsilon := \frac{W(a)}{W(b_j)} = \frac{\alpha}{\varepsilon(1-\alpha)} = \frac{\sigma(\beta\varepsilon) - 1/2}{\varepsilon\big(\sigma(\beta) - 1/2\big)}.$$

The Hidden Welfare Ratio Approaches Γ_β

The ratio between the special completion and each b_j is

$$r_\varepsilon := \frac{W(a)}{W(b_j)} = \frac{\alpha}{\varepsilon(1-\alpha)} = \frac{\sigma(\beta\varepsilon) - 1/2}{\varepsilon\big(\sigma(\beta) - 1/2\big)}.$$

By the definition of the derivative,

$$\lim_{\varepsilon \downarrow 0} \frac{\sigma(\beta\varepsilon) - 1/2}{\varepsilon} = \beta\sigma'(0) = \frac{\beta}{4}.$$

Therefore

$$\lim_{\varepsilon \downarrow 0} r_\varepsilon = \frac{\beta/4}{\sigma(\beta) - 1/2} = \Gamma_\beta.$$

Because $\Gamma_\beta > 1$ for $\beta > 0$, sufficiently small ε has $r_\varepsilon > 1$, so a is welfare optimal.

Maximal Lotteries Are Comparison-Optimal

Theorem

In the heterogeneous Bradley–Terry model:

1. *a maximal lottery has distortion at most* Γ_β*;*
2. *every rule mapping the aggregate matrix* P *to a policy has worst-case distortion at least* Γ_β*.*

Maximal Lotteries Are Comparison-Optimal

Theorem

In the heterogeneous Bradley–Terry model:

1. *a maximal lottery has distortion at most* Γ_β*;*
2. *every rule mapping the aggregate matrix* P *to a policy has worst-case distortion at least* Γ_β*.*

In the lower-bound family every pair ties: $P_{ij} = 1/2$. Hence every policy defeats or ties every challenger and is a maximal lottery. Choosing the special completion among those assigned minimum probability therefore gives a maximal lottery whose distortion approaches Γ_β.

Summary

- Distortion compares an alignment rule's population welfare with the best welfare achievable on a chosen common cardinal scale.

Summary

- Distortion compares an alignment rule's population welfare with the best welfare achievable on a chosen common cardinal scale.
- The sigmoid reweights users' utility gaps before they are averaged; chord and tangent bounds control how much those weights can differ.

Summary

- Distortion compares an alignment rule's population welfare with the best welfare achievable on a chosen common cardinal scale.
- The sigmoid reweights users' utility gaps before they are averaged; chord and tangent bounds control how much those weights can differ.
- Comparing a maximal lottery directly with the welfare-optimal policy and applying the upper welfare bound once gives distortion at most Γ_β.

Summary

- Distortion compares an alignment rule's population welfare with the best welfare achievable on a chosen common cardinal scale.
- The sigmoid reweights users' utility gaps before they are averaged; chord and tangent bounds control how much those weights can differ.
- Comparing a maximal lottery directly with the welfare-optimal policy and applying the upper welfare bound once gives distortion at most Γ_β.
- A two-type construction makes every pair look like a fair coin while hiding a welfare ratio approaching Γ_β.

Summary

- Distortion compares an alignment rule's population welfare with the best welfare achievable on a chosen common cardinal scale.
- The sigmoid reweights users' utility gaps before they are averaged; chord and tangent bounds control how much those weights can differ.
- Comparing a maximal lottery directly with the welfare-optimal policy and applying the upper welfare bound once gives distortion at most Γ_β.
- A two-type construction makes every pair look like a fair coin while hiding a welfare ratio approaching Γ_β.
- Maximal lotteries are comparison-optimal.

Thanks!

See you next class!

References

- Paul Gölz, Nika Haghtalab, and Kunhe Yang. "Distortion of AI Alignment: Does Preference Optimization Optimize for Preferences?" arXiv:2505.23749, 2025.
- Peter C. Fishburn. "Probabilistic Social Choice Based on Simple Voting Comparisons." *Review of Economic Studies*, 1984.
- Ariel D. Procaccia and Jeffrey S. Rosenschein. "The Distortion of Cardinal Preferences in Voting." *Cooperative Information Agents X*, 2006.
- Daniel Halpern et al. "AI Alignment From Social Choice Perspectives." *ACM SIGecom Exchanges*, 2026.