Marginal Moral Cost in Proxy-based Competition

Working paper — pilot data

Sara Gorgzadeh[1] Paul Banse[1,2] Johannes Schultz[1,3] Oliver Braganza[1,3*]

[1]Center for Economics and Neuroscience, University of Bonn
[2]Institute for Socio-Economics, University of Duisburg-Essen
[3]Institute of Experimental Epileptology and Cognition Research, University Hospital Bonn

October 9, 2026

Abstract

Competitive institutions use proxies to promote goals they cannot measure directly: publications for good science, case numbers for health, profit for value. At the same time, many practitioners are morally motivated to serve the goal itself: the quality of their science, their patients, their customers. But where proxy and goal diverge, acting morally may cost them competitive rank. This raises a long-standing question: does competition erode morals? A growing experimental literature has begun to address this question, but with widely varying results and without formal theory to provide quantitative predictions of the expected degree of moral erosion. We elicit each participant's marginal moral valuation of a charitable donation in an individual decision block and build on contest theory to derive equilibrium erosion predictions. Importantly, the cost of competing is a moral one, which each participant weighs differently and which we measure beforehand, person by person. We compare two fully incentivized contest conditions modeling real-world proxy-based competition in the short and long run, formally structured as repeated all-pay and winner-pay auctions, respectively. A pilot with 68 participants in 17 groups validates our approach. Both group and individual level predictions from prior individual elicitations are confirmed by subsequent tournament behavior —with one revealing exception: in all-pay-like contests, participants over-dissipate, acting significantly less moral than the equilibrium prediction; the all-pay format erodes behavior less than the winner-pay format but selects more strongly against the most morally motivated. We also find that individual valuations measured after competing are lower than before, in this study and in an independent market experiment. All results are exploratory.

1 Introduction

Society often relies on competitive proxy measures to further abstract social goals. In the academic system, individuals are incentivized and selected on measures such as publication counts, impact factors or citations, each intended to approximate contributions to knowledge. In markets, firms are selected on profitability as a proxy for consumer value; in medicine, providers may be paid on case numbers while the goal is patient health. Many agents in such systems may nevertheless be intrinsically, morally motivated to serve the underlying goal. But proxy and goal often diverge along some dimensions (Goodhart, 1984; Kerr, 1975; John et al., 2024), e.g., questionable research practices in academia or consumer misdirection in markets. Along these dimensions, moral incentives begin to conflict with competitive proxy incentives (Holmström and Milgrom, 1991; Braganza, 2022), raising the perennial question of whether competition erodes moral behavior (Shleifer, 2004; Falk and Szech, 2013).

*Corresponding author: oliver.braganza@ukbonn.de.

Whether competition can affect moral behavior has been studied at scale. Huber et al. (2023) randomized 18,123 participants across 45 independently designed experiments. While they did find a small adverse meta-analytic effect, they emphasized design heterogeneity driving estimates about 1.6 times the average standard error, with some studies suggesting competition increases moral behavior. Studies on experimental markets yield similarly inconclusive results, sometimes supporting (Falk and Szech, 2013; Ziegler et al., 2024) sometimes contradicting a moral erosion effect (Bartling et al., 2023). What this literature lacks is a formal framework that allows quantitative predictions of the expected degree of moral erosion, as a function of study design features.

We combine individual marginal moral cost elicitations and contest theory to make such predictions and test them. Each participant first reveals their marginal moral valuation of a charitable donation in an individual decision-making (IDM) block. They then enter a contest phase in fixed group-sizes between 2 and 6. Specifically, we use a multi-prize contest where the cost of bidding is the moral value forgone. In traditional contest experiments the cost of competing is money or effort; here it is a moral cost, elicited from each participant before competition. The equilibrium of each group's contest is therefore an out-of-sample prediction from measured, heterogeneous, non-monetary moral costs.

We test two conditions designed to model real proxy-based competition in the short or the long run. A 'long run' (all-or-nothing, aon) condition assumes that proxy rank determines both monetary incentives and participation—only those with sufficient proxy performance survive competition long-term and get to morally contribute. Proxy rank thus gates both the personal payout and whether the donation is made. In addition we model a 'short run' (consolation, CONS) condition, where proxy rank is monetarily rewarded, but participants are assumed to remain in the system even with insufficient proxy performance—they get to generate donations regardless of proxy rank.

Three results emerge from a pilot of 68 participants across 17 groups. First, when the charitable contribution is gated on winning (AON), the contest is winner-pay and the equilibrium predicts that play collapses onto the proxy; it does, at the predicted level, with no parameter taken from tournament data. Second, the prior valuations predict individual and group-level play: they rank who resists competition, and so who survives it, and together with the contest's structure they predict that a player opts out less when their rivals are more morally motivated, which the data show. No aggregate or representative-agent model could make this prediction. Third, when the contribution counts regardless (CONS), the contest is all-pay and the equilibrium predicts that the most morally motivated drop out; instead participants over-bid the equilibrium by 57%, close to the median over-bidding rate of 72% across contest experiments (Sheremeta, 2013). Participants thus act markedly less morally than their elicited valuations imply, and this finding can be directly related to the established contest literature. Prospectively, our approach will allow to dissect whether overdissipation reflects a value of winning or a lower moral valuation during competition. Interestingly, valuations measured after competing are lower than before, in this study and in an independent market experiment, pointing towards the latter, but crucial controls are still missing from the present pilot data. Importantly, the participants themselves judge the choice as a moral one: 84% believe that most participants would rate full proxy orientation as morally inappropriate.

2 Design

The choice. The experiment is framed as a research career (Smaldino and McElreath, 2016; Tiokhin et al., 2021), with the concrete, simplified tradeoff anchored in empirically grounded expectations of false-positive rates for different sample-sizes (Braganza, 2020). Each participant has a budget of 200 samples and chooses, on a slider, how many studies $k \in \{1, \ldots, 40\}$ to run with it. Many small studies maximize the publication rate (the proxy); few large ones maximize

informativeness (the goal). The payout is linear in the slider position

$$x \;=\; 1 - \frac{\ln k}{\ln 40}, \qquad \text{payout} = c_P(1-x), \qquad \text{donation} = c_D\, x, \tag{1}$$

where $x = 1$ is fully goal-oriented and $x = 0$ fully proxy-oriented, and the donation (goal value) is paid to UNICEF. Note that, in order to keep the geometric slider-to-value mapping linear, but use actual budget allocation options, available slider positions in x are log-spaced: Moving from 1 to 2 studies changes x, sample size, scientific value, and donations much more than moving from 39 to 40 studies.

Blocks. Participants complete an IDM block, two competitive blocks of 15 rounds each in fixed groups, and a second IDM block. The IDM blocks vary the stakes (c_P, c_D) along two orthogonal sweeps: a *price* sweep holds $c_P = €5$ and raises c_D through 5, 10 and 20; a *stake* sweep holds the price of giving at one and scales both legs through €1, 2.50, 5 and 10. The pair (5, 10) appears twice, and two comprehension checks set one leg to zero. From the fourth session on, two further pairs, (5, 2.5) and (2.5, 5), were added. The competitive blocks run at $(c_P, c_D) = (5, 10)$, so the individual (5, 10) rounds are a like-for-like baseline. The two competitive blocks reflect the two conditions, randomized in order (Appendix F).

Competition. Within a group of n, players are ranked each round by their payout $c_P(1-x)$, noiselessly, with ties broken at random; the top half survives (odd groups resolve the fractional cutoff by a coin flip each round). After every round each player sees all rivals' choices and the survivors. The two conditions, 15 rounds each, differ only in what survival gates:

- **Consolation (cons):** survivors receive their payout; the donation is paid for everyone.
- **All-or-nothing (aon):** non-survivors receive neither payout nor donation.

Survivors receive the same payout as in the IDM block. Multi-winner tournaments commonly pay winners a multiple of the individual piece rate (Gneezy et al., 2003), which confounds competition with higher incentives. We deliberately do not: since competition only lowers the expected payout, any increase in proxy-oriented play cannot be attributed to higher monetary incentives. For each participant, one round from all blocks is drawn for payment, on top of a €17 base payment.

Sample. Six sessions, 68 participants in 17 groups of 2 to 6 (Appendix A). Condition order is constant within a group and, in the pilot, partially confounded with session.

Norms and a second instrument. At the end, participants evaluate "what you believe the majority of participants here today would consider to be 'morally appropriate' or 'morally inappropriate'," where "by 'morally appropriate', we mean behavior that is considered correct or ethical." Each of several allocations is rated in each context on a four-point scale from "very morally inappropriate" to "very morally appropriate", without a neutral midpoint, with a bonus for matching the modal rating (Krupka and Weber, 2013). A separate sample of 52 participants completed a multiple price list over donations at five stake scales, before and after a market experiment adapted from Ziegler et al. (2024).

3 Model

3.1 Preferences

Agent i is risk-neutral over own money and derives moral value from their own realized donation ('warm glow' in Andreoni (1990), Appendix E discusses the alternative). Written in shares and

weighted by a stake term,

$$U_i \;=\; c_P(1-x_i) \;+\; \omega_D\, u_i(x_i), \qquad \omega_D \equiv c_D^{1-\theta}. \tag{2}$$

Assumption 1 (Moral value)**.** u_i is twice differentiable, strictly increasing and weakly concave, with $u_i(0) = 0$ and $u_i'(0^+) < \infty$.

The stake exponent θ is the elasticity of the effective price of giving, c_P/ω_D, with respect to the scale of the stakes: at $\theta = 0$ the chosen share is invariant to that scale, and for $\theta > 0$ it falls as the stakes grow. Money stays linear so that risk neutrality, and with it the competitive conditions as lotteries, remains coherent. The finite slope at zero makes the proxy corner reachable. We fit an exponential form to each participant's donation choices in the IDM block (Section 4),

$$u_i'(x) = a_i\, e^{-bx}, \qquad u_i(x) = \frac{a_i}{b}\big(1 - e^{-bx}\big), \qquad \alpha_i \equiv \ln a_i \sim \mathcal{N}(\mu, \tau^2), \tag{3}$$

so that a_i is each person's moral valuation at the proxy corner and b how fast it declines as the donation grows.

3.2 The three conditions

In a group of n, the s players with the highest payout survive, with the survival threshold set at 50%: $s = n/2$, and in odd groups s is $\lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$ with equal probability each round. Writing $\mathbb{1}_i$ for survival and $\pi_0 \geq 0$ for a value of winning in itself,

$$\textbf{Individual:} \quad U_i^B = c_P(1-x_i) + \omega_D\, u_i(x_i), \tag{4}$$

$$\textbf{cons:} \quad U_i^C = \big[c_P(1-x_i) + \pi_0\big]\, \mathbb{1}_i + \omega_D\, u_i(x_i), \tag{5}$$

$$\textbf{aon:} \quad U_i^A = \big[\, c_P(1-x_i) + \omega_D\, u_i(x_i) + \pi_0 \,\big]\, \mathbb{1}_i. \tag{6}$$

In CONS the donation is realized unconditionally; in AON it requires survival, reflecting the assumption that agents have to survive competition in order to contribute to the goal (Braganza, 2022). Writing $W_i \equiv \Pr(\mathbb{1}_i = 1)$ for the probability of surviving, which depends on the rivals' choices, and $\gamma \in \{0, 1\}$ for a gate that records whether the donation counts only for survivors, one expression covers all three conditions:

$$\mathbb{E}[U_i] \;=\; W_i \left[c_P(1-x_i) + \pi_0 + \gamma\, \omega_D\, u_i(x_i)\right] \;+\; (1-\gamma)\, \omega_D\, u_i(x_i), \tag{7}$$

with CONS as $\gamma = 0$, AON as $\gamma = 1$, and the individual block as $W_i = 1$ and $\pi_0 = 0$. All predictions below set $\pi_0 = 0$.

3.3 The individual baseline

Proposition 1 (Baseline choice and identification)**.** *The individual optimum x_i^B satisfies*

$$u_i'(x_i^B) \;=\; \frac{c_P}{c_D^{1-\theta}} \tag{8}$$

whenever it is interior; otherwise $x_i^B = 0$ if $u_i'(0^+) \leq c_P/\omega_D$ and $x_i^B = 1$ if $u_i'(1) \geq c_P/\omega_D$.

Proof. U_i^B is concave in x_i; its derivative is $-c_P + \omega_D u_i'(x_i)$. □

Each individual round therefore fixes an effective price and locates one point on the participant's marginal moral value curve; the price sweep traces its curvature, the stake sweep identifies θ. Under (3) the latent optimum is linear in α_i, $x_i^* = \big(\alpha_i - \ln c_P + (1-\theta)\ln c_D\big)/b$.

3.4 The contests

Ranking on the payout makes a player's proxy score $c_P(1 - x_i)$ their *bid*. Its cost is the moral value forgone. The two conditions then correspond to two canonical contest formats.

cons is an all-pay contest. The moral cost of a bid is incurred whether or not the player survives, as a bid is paid in an all-pay auction, and the prizes are the s payouts. With heterogeneous costs, all-pay contests pin down equilibrium payoffs and participation, and players with high costs drop out (Barut and Kovenock, 1998; Siegel, 2009). Here dropping out means choosing the moral end of the slider.

aon is a winner-pay contest. A loser receives nothing whatever they bid, so the cost of a bid matters only to winners, as in a pay-as-bid auction. The player's own moral value is part of the prize. With noiseless ranking, undercutting is always profitable:

Proposition 2 (Collapse under AON)**.** *With deterministic scores and ties broken uniformly, the unique symmetric equilibrium of* AON *is* $x^* = 0$ *for every type.*

Proof. Conditional on surviving, the payoff $c_P(1 - x) + \omega_D u(x) + \pi_0$ is positive; a non-survivor receives zero. From any symmetric profile $x^* > 0$, undercutting to $x^* - \epsilon$ raises survival from s/n to one at a cost of order ϵ. At $x^* = 0$ any deviation to $x > 0$ loses for sure. □

This contrast between the two formats drives the model's predictions. Under AON caring about the goal is an argument *for* competing, since the contribution counts only if one survives. Under CONS it is an argument for dropping out. The same contrast between all-pay and winner-pay formats is central in the theory of charitable fundraising (Goeree et al., 2005).

Remark 1 (Moral opt-out). Under CONS the objective can have two local maxima, an interior one and the moral end $x = 1$: a sufficiently moral type prefers to abandon the race rather than contest it. The two must be compared by value, not by first-order conditions.

4 Measuring individual moral valuations

We estimate (3) by censored maximum likelihood on individual choices with a participant random effect. Two features of the data make this necessary. About two-fifths of individual choices sit at a corner, where a choice is an inequality, not a point; and fitting design-point means instead of choices reproduces every mean while implying a grossly wrong choice distribution. The level μ is weakly identified, so we maximize the likelihood subject to the model-implied pooled mean matching the observed one (a restriction, not an extra parameter; it costs 0.08 log-likelihood units). Each participant's α_i is the posterior mean. Details are in Appendix B.

Estimates and validity. On both IDM blocks: $\mu = 1.80$, $\tau = 2.59$, $b = 5.89$, $\theta = 0.36$, $\sigma = 0.27$. Four checks support the measurement (Figure 1), and the individual curves explain $R^2 = 0.76$ of the variance across participant × design-point means, and 0.68 out of sample at a design point left out of the fit (Figure 1D). The fitted surface matches the design points without being fitted to them (largest residual 0.02). It reproduces the corner shares (implied 34% and 9% at the two ends, observed 32% and 9%). The stake elasticity agrees with an independent instrument: the ratio θ/b, which both instruments identify, is 0.061 on the slider and 0.067 on the price list, in different samples. And the two choices at the repeated (5, 10) pair correlate at 0.78 within a block. The comprehension checks are passed in 87% of check choices.

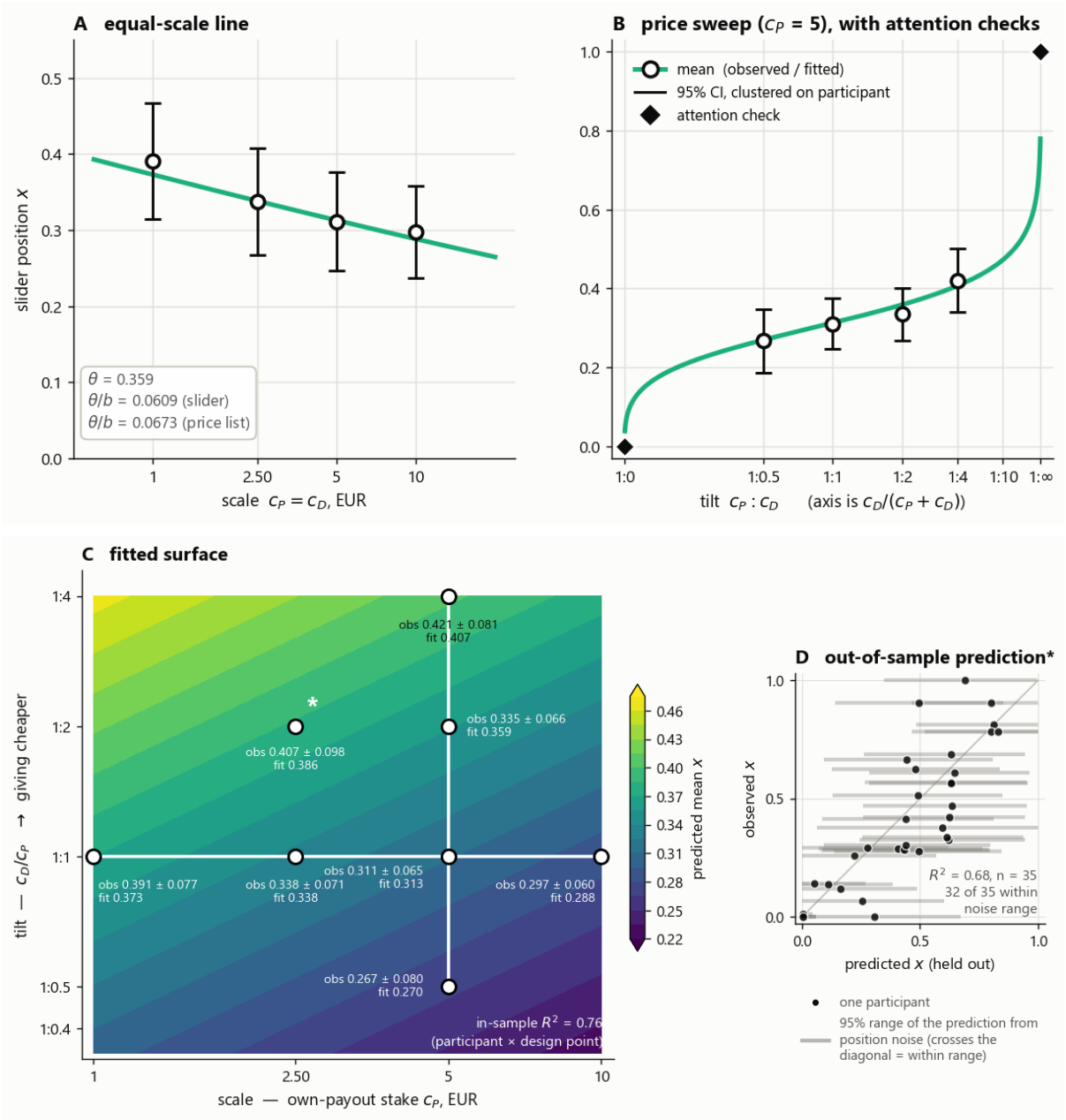

Figure 1: Measuring moral valuations. (A, B) The IDM block's two orthogonal sweeps with the fitted population mean. (A) Scaling both stakes at a constant price of giving; the fall identifies the stake elasticity θ. (B) Raising the donation leg at a fixed payout of €5, with the two comprehension checks at the ends; the rise traces the curvature of u'. Circles are observed means with 95% intervals clustered on participant. (C) The implied surface of mean slider positions over stake scale and price of giving; circles are design points, labeled with observed mean, interval and fit; the individual curves explain $R^2 = 0.76$ of the variance across participant × design-point means. (D) Out-of-sample prediction at the one design point off both sweeps, (2.5, 5), marked * in C and added from the fourth session on. A fit that leaves out all choices at this point predicts each participant's position there from their remaining choices, plotted against their observed mean (35 participants, $R^2 = 0.68$). Horizontal bars: the 95% range of each prediction implied by the fitted position noise; a bar crossing the diagonal means the observation lies within it, as for 32 of 35 participants. All 68 participants, 1,090 choices.

All contest predictions are based on pre-contest individual elicitations. Only the first IDM block is measured before the tournament, so every prediction uses curves estimated on it alone. On half the data the curvature is less well pinned ($b = 7.9$, and $\theta/b = 0.043$ against

the price list's 0.067); this is the pilot's weakest measurement point.

5 The benchmark: equilibrium from measured valuations

For each group we compute the equilibrium of its actual game from its members' curves, estimated on the first IDM block. Two features of the measurement enter.

Position noise. Choices in the IDM block scatter around the latent optimum, with a pile-up at the slider's ends; the estimated model captures this as censored noise on the position, $x = \text{clip}(x^* + \varepsilon)$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. In the contest a player chooses an *intended* position, realized with the same noise, discretized onto the 40 options.

Mixed intentions. We solve for a Nash equilibrium in mixed intentions in which every player anticipates the rivals' realized choices. Survival probabilities are computed exactly on the noiseless ranking by a dynamic programme over the number of rivals ranked above and tied. The equilibrium is found as the high-precision end of the principal branch of the logit quantal response equilibrium (McKelvey and Palfrey, 1995); the precision parameter is a device for selecting an equilibrium and is neither estimated nor interpreted. The benchmark is thus the Nash limit of the logit path, not a logit equilibrium with an estimated precision, as used to explain over-bidding in all-pay auctions (Anderson et al., 1998; Gneezy and Smorodinsky, 2006); behavioral noise enters only as the position noise measured in the IDM block.

No parameter of the benchmark comes from tournament data. It uses each participant's curve, the population parameters (b, θ, σ) and the rules of the game. On the pilot every group's equilibrium turns out to be pure: players sort onto the two ends of the slider, the all-pay split into competing and dropping out.

Predictions. The benchmark predicts the following.

H1. Level under aon. Play collapses toward the proxy corner. Without noise the collapse is complete for every type (Proposition 2); with the measured position noise it is partial, and the benchmark predicts how far it goes.

H2. Level under cons. The most morally motivated drop out (Remark 1), which puts the equilibrium mean bid *below* the bid without competition (in all 17 pilot groups). The benchmark predicts each group's level; observed bids relative to it measure over- or under-dissipation.

H3. Sorting. Individual valuations rank play: strongly under CONS, where high valuations lead to opting out, and weakly under AON, where the collapse overrides them.

H4. Rivals. Under CONS, holding a player's own valuation fixed, more morally motivated rivals make the player opt out *less*, because rivals who drop out make winning cheap. Under AON there is no such effect. The benchmark predicts the sign and size of this effect.

6 Results

6.1 Behavior by condition

Play separates sharply by condition (Figure 2). Under AON, 70% of choices sit at the proxy corner ($x \leq 0.05$), against 51% under CONS. Under AON play converges toward the corner within about five rounds; under CONS it stays spread across the slider throughout, with a varying group of participants at or near the goal end. Condition order affects only the first rounds (Appendix F). Since the equilibrium describes settled play, we report both rounds 2–15, which drop only the round before any feedback, and the settled rounds 11–15; over-bidding is the same in both (Section 6.2).

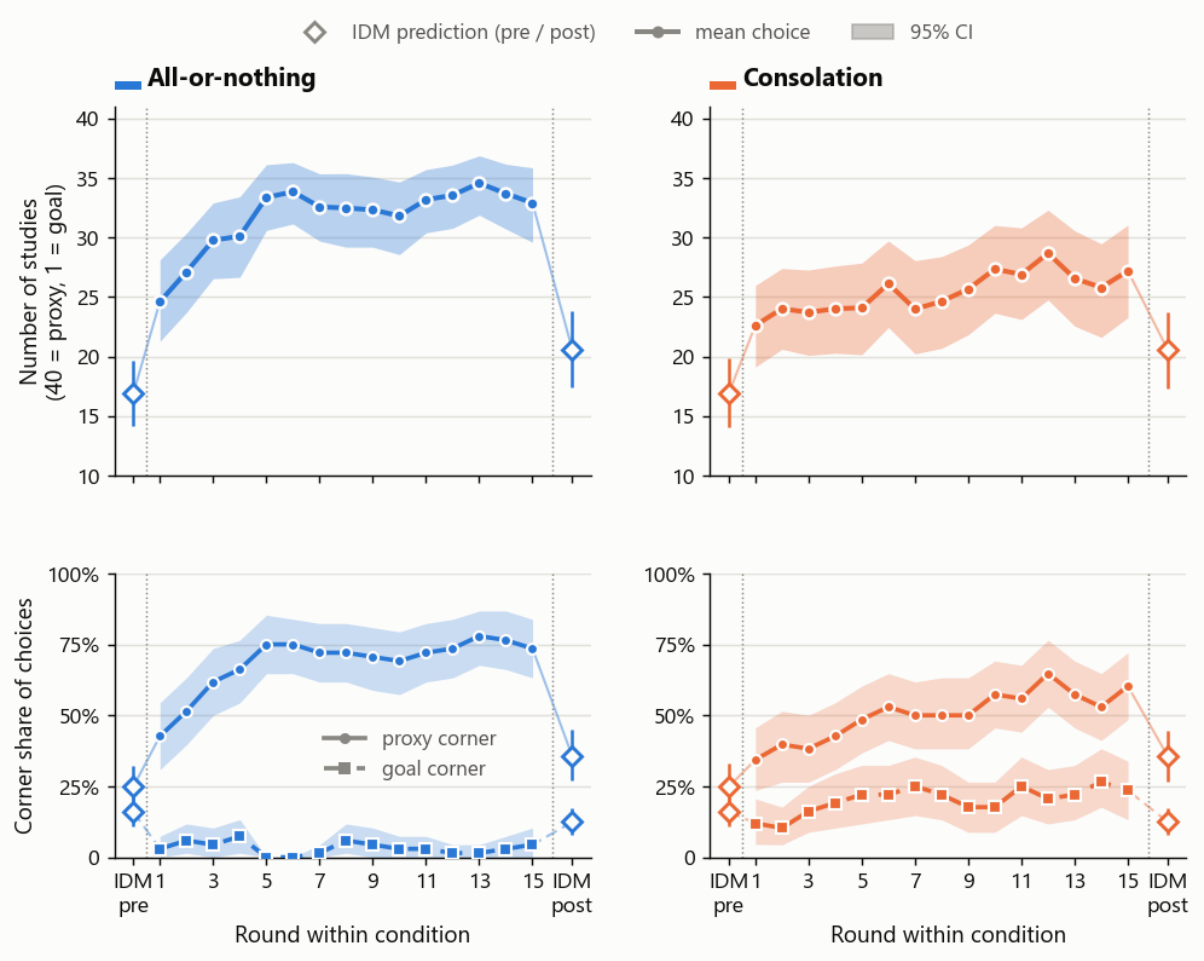

Figure 2: Data across conditions. Rounds are aligned on each participant's own order of conditions. *Top:* mean number of studies per round (40 = fully proxy-oriented, 1 = fully goal-oriented) with 95% bootstrap intervals over participants. *Bottom:* share of choices at the proxy corner ($x \leq 0.05$, solid) and at the goal corner ($x \geq 0.8$, dashed). Diamonds: the IDM prediction at the contest stake, each participant's expected choice from their curve estimated on the IDM block before ("IDM pre", the prediction L0) and after ("IDM post") the contests, with 95% intervals; the faint connectors only link prediction and data. 68 participants in 17 groups.

6.2 Level of play: the equilibrium against observed bids

Figure 3 confronts each group's equilibrium with its play (H1, H2), in the units of the contest literature.

aon: the winner-pay contest collapses as predicted. The equilibrium moves every group from a no-contest bid of €3.11 to a Nash bid of €4.43; observed bids are €4.32. Relative to Nash, over-bidding is −2% (95% interval −9% to 3%; +2% in rounds 11–15). The level of play is predicted with no parameter from tournament data: the mean prediction error per participant is zero, and the error is as small as that of a benchmark that simply uses other sessions' observed AON choices (Appendix C). The level depends on the measured position noise and the rules of the game, not on individual differences in valuation: giving every participant the sample mean valuation predicts it equally well (bias −0.01, mean bid €4.53). The corner caps bids at €5, so any over-dissipation can show only as extra mass at the corner, which is what we see (70% of choices at $x \leq 0.05$ observed against 54% predicted, rounds 2–15).

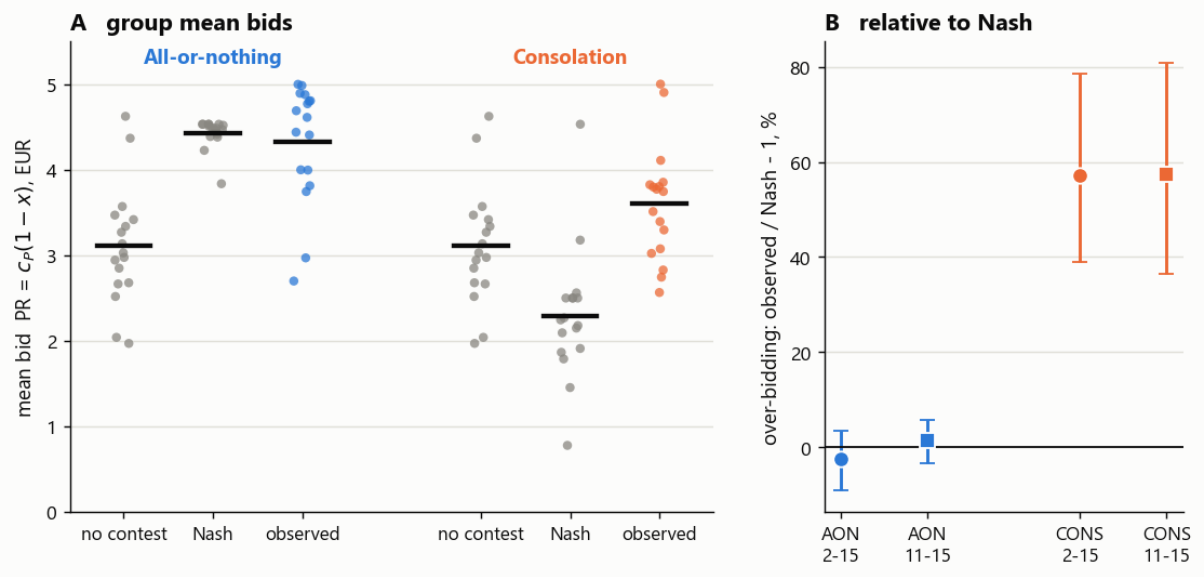

Figure 3: Bids against the parameter-free equilibrium. A player's bid is the proxy score $c_P(1-x)$, €0–5. (A) Each point is one group's mean bid (rounds 2–15): the no-contest bid implied by its members' curves, the Nash bid of its contest, and the observed bid; bars are means over groups. (B) Over-bidding, observed mean bid over Nash mean bid minus one, in rounds 2–15 and 11–15, with 95% bootstrap intervals over groups. Curves from the first IDM block, no parameter from tournament data.

cons: the all-pay contest over-dissipates. Here the equilibrium predicts that moral types drop out, which *lowers* the mean bid from the no-contest level of €3.11 to €2.29, in every group. Observed bids go the other way, to €3.60, above even the no-contest level in 15 of 17 groups: over-bidding of +57% (39% to 79%), the same in rounds 11–15. Because survivors are paid no more than without competition, this cannot reflect higher monetary incentives. This is the pattern contest experiments find: over-bidding relative to Nash is their most robust regularity, strongest in all-pay auctions and lottery contests and weak in noisy rank-order tournaments (Dechenaux et al., 2015), with a median rate of 72% across studies (Sheremeta, 2013). The pilot's CONS result sits within that range, now measured against elicited moral costs.

The size of the deviation. Because the costs are measured, the deviation under CONS has a size in the model's own terms: participants give up more of the donation than their elicited valuations imply, an erosion of moral behavior beyond equilibrium. Adding one parameter to the benchmark fits the CONS level out of sample, and two readings of that parameter fit equally well. Each chosen on five sessions and scored on the sixth, a value of winning of about $\pi_0 = €40$ per round and a common downward shift of every participant's log moral valuation by δ between -2 and -3 during competition both match the held-out level (bias +0.02 and +0.04; Appendix D). They are nearly the same model under CONS: dividing (5) by e^{δ} scales the prize by $e^{-\delta}$, and $(5+40)/5 = 9 = e^{2.2}$. What the pilot identifies is the prize-to-cost ratio: relative to their elicited moral costs, participants act as if winning were worth about nine times its monetary prize, or, equivalently, as if their moral valuation during competition were about a ninth of the one elicited individually. Whether competition adds value to winning or lowers moral valuations is an additional question the pilot cannot answer. Varying the money prize would separate the two, since play under a large additive π_0 is insensitive to the prize and play under a scaled cost is not.

6.3 Who: individuals and their groups

Valuations rank individual play. Under CONS a participant's predicted valuation strongly predicts their opt-out share (slope 0.66, $p < 0.001$, $n = 68$; Figure 4B), and their mean position (rank correlation 0.74, and 0.68 within groups, so group composition does not drive it). Under AON, where almost everyone ends at the corner, the relation is weak but positive (slope 0.09, $p = 0.028$). This confirms the individual-level prediction (H3), and it is not self-evident: moral behavior is highly context-dependent (Huber et al., 2023), yet a ranking measured without competition survives strong competitive incentives.

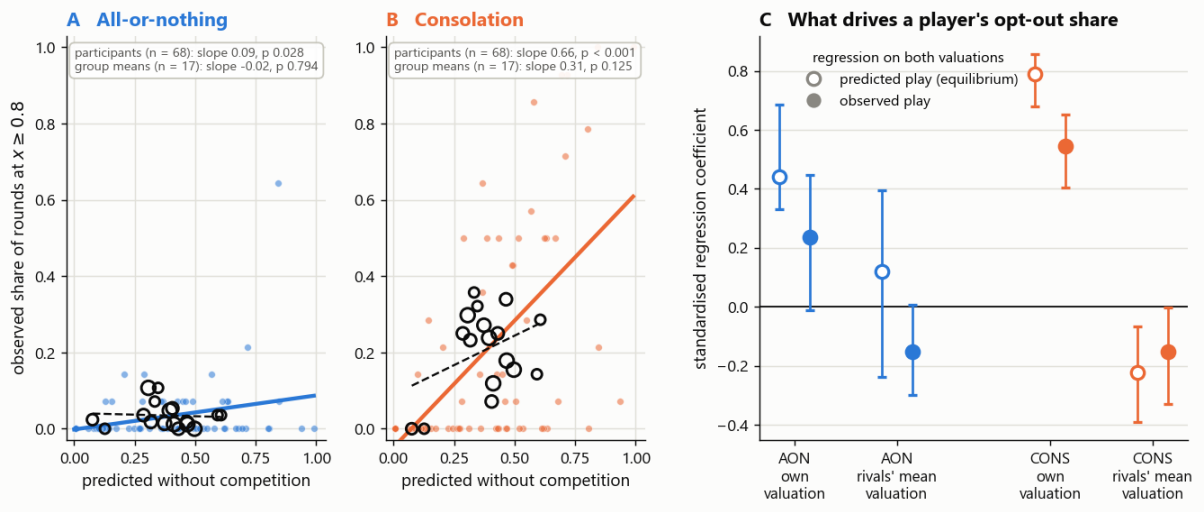

Figure 4: Individual valuations and contest theory together. (A, B) Each participant's slider position predicted without competition, from the first IDM block, against their share of rounds at the opt-out end ($x \geq 0.8$). Coloured line: regression over participants; black circles: group means, dashed line through them; p against a zero slope. (C) One regression per condition of a player's opt-out share on their own valuation α_i and their rivals' mean valuation, all standardized; hollow: the equilibrium's predicted play; filled: observed play; 95% bootstrap intervals over groups. The opt-out outcome was chosen after inspecting the mean-position version (Figure 10).

Group means depart from the individual relation, as the equilibrium predicts. Averaged over groups, the same relation almost disappears under CONS (slope 0.31, $p = 0.13$, $n = 17$). A group's average valuation is also the average valuation of each member's rivals, and the equilibrium says that rivals matter: holding own valuation fixed, more moral rivals make a player opt out *less*, because rivals who drop out make winning cheap. That rivals effect is -0.22 in the equilibrium and -0.15 (-0.31 to -0.01) in the data (Figure 4C). Under AON the equilibrium predicts no such effect. The flattening and the rivals coefficient are two views of the same data, and together they confirm the group-level prediction (H4) in sign and approximate size. The effect is modest, but it is a proof of principle: an aggregate or representative-agent model cannot produce it, and it follows only from combining the individual measurements with the contest's strategic structure.

Who survives. Survival is the share of a participant's 15 rounds in the top s (ties at random); we relate it to α_i from the first IDM block, centered within groups, and permute α within groups for inference. More moral participants survived fewer rounds: within-group $r = -0.35$ (-0.57 to -0.15) under AON and -0.67 (-0.77 to -0.51) under CONS, slopes -0.025 (-0.048 to -0.009) and -0.073 (-0.094 to -0.054), $p = 0.002$ and $p < 0.001$. The equilibrium gives each player a survival probability W_i; simulated over the same 15 rounds, it expects $r = -0.32$ under AON, as observed, but -0.85 under CONS: there the most moral eroded beyond the equilibrium and

so survived more often, the CONS excess again, and δ, chosen out of sample on the level, closes the gap (-0.70). With a noiseless ranking, survival follows from play relative to rivals, so this largely mirrors H3. It adds how strongly each gate selects on the moral type: AON erodes behavior most (bids 39% above the no-contest level) but sorts weakly; CONS erodes less (16%) but selects against the most moral. It is exploratory, not one of H1–H4 (Appendix I).

6.4 Valuations after competing

The second IDM block, after the tournament, is lower than the first (Figure 5): mean x falls from 0.376 to 0.305 (paired difference -0.071, 95% interval -0.108 to -0.036; $d = -0.46$), 40 participants lower and 16 higher. The drop appears at every design point and is reproduced by a single common shift in valuation, which improves the likelihood by about 18 units for one parameter. The ranking of participants is preserved ($\rho = 0.83$).

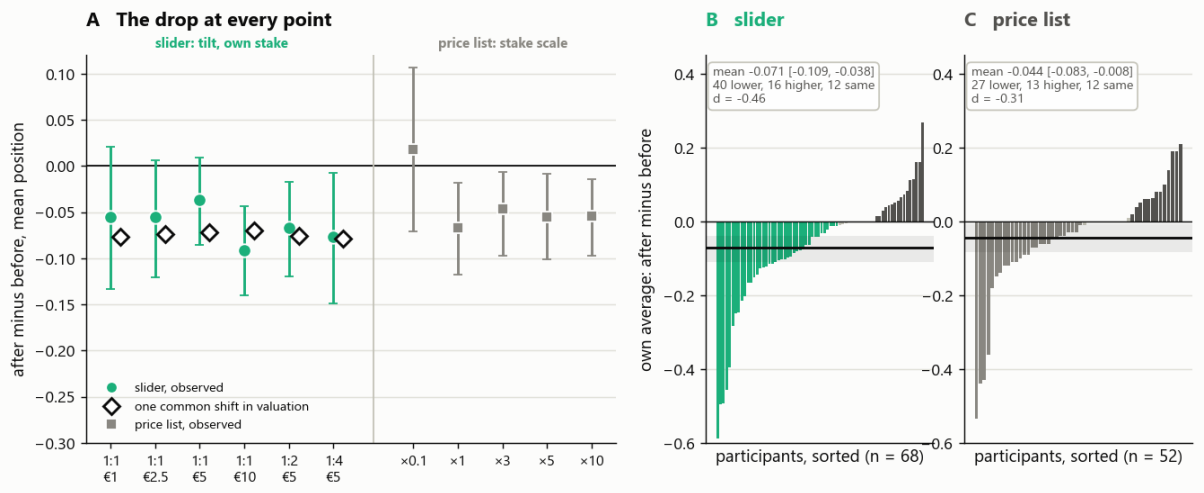

Figure 5: Moral valuations are lower after competing, in two instruments. (A) After minus before, averaged over participants, at each of the six slider design points every session faced (tilt and own stake) and at each stake of an independent multiple price list run around a market experiment; 95% bootstrap intervals over participants. Diamonds: the change predicted at each slider point by one common shift $\delta = -0.75$ in every participant's log valuation, fitted jointly with the curves. (B, C) Each participant's average difference, sorted. Before/after is confounded with order in both designs.

The same pattern appears in an independent sample and instrument. In the price list run before and after a market experiment, mean positions fall from 0.340 to 0.300 (difference -0.044, -0.084 to -0.007; $d = -0.31$), 27 lower and 13 higher. These results call for caution. Both designs confound the before/after contrast with order, time and repetition, and Bartling et al. (2023) argue that repetition, not market interaction, accounts for the erosion found in market experiments. Only a control without competition between the two blocks can attribute the drop to competition.

The drop does, however, bound how much of the CONS deviation time or repetition could explain. On the scale of the tournament predictions it corresponds to a shift of about -1.0 in every log valuation, between a third and a half of the shift that accounts for CONS play. Since the tournament lies between the two IDM blocks, a baseline that interpolates between them corresponds to a shift of about -0.5; the equilibrium computed from it still leaves over-bidding of +38%. Attributing the entire drop to time or repetition rather than to competition, a shift of -1.0, still leaves +26% (AON: -3% and -4%).

6.5 Norms

Norms do not follow behavior (Figure 6). Giving more is rated more appropriate in the individual context (the allocation 2/6 at +1.46, 4/2 at −0.10), and the same ordering holds in both competitive contexts: 2/6 is rated +1.47 and +1.60 for a player expecting to win under AON and CONS, and 4/2 at +0.07 and +0.22. Only the prospect of losing lowers the ratings, and it lowers the more selfish allocation too. Competition does not make proxy-oriented choices appear more appropriate, even as participants make them.

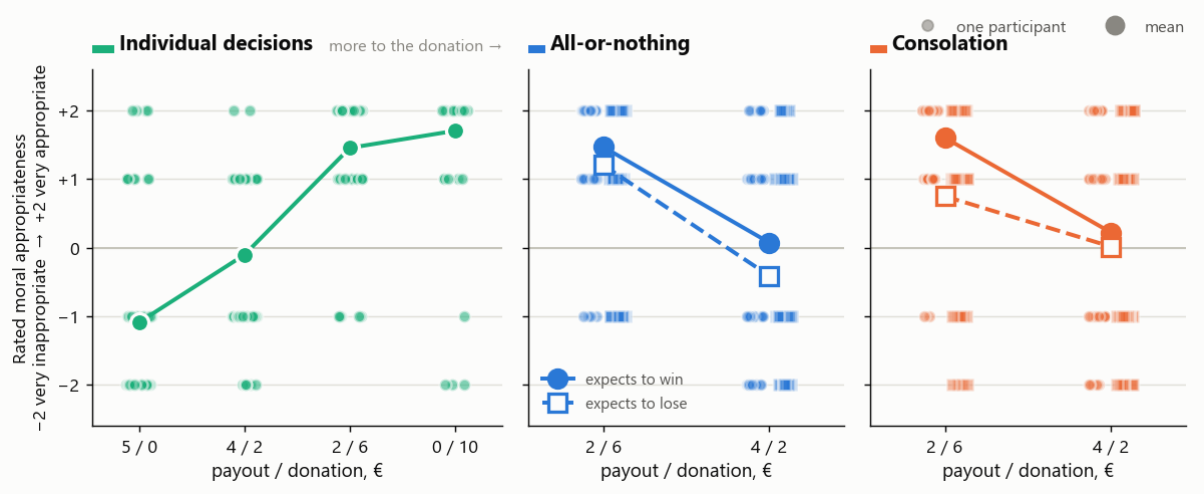

Figure 6: Social norms by context. Appropriateness ratings (−2 to +2) of allocations (personal payout / donation, in euros), elicited with a bonus for matching the modal rating. In the competitive contexts each allocation is rated for a player who expects to win and for one who expects to lose. All 68 participants.

The ratings also show that participants treat the choice as a moral one. Since the scale has no midpoint, every rating takes a side. In the individual context, 84% believe that the majority would rate full proxy orientation (5/0) as morally inappropriate. In the competitive contexts only milder allocations were rated; even the 4/2 allocation, less proxy-oriented than most contest choices, is rated inappropriate by 46% and 43% for a player expecting to win under AON and CONS, and by 66% and 51% for one expecting to lose. Most participants thus chose, in the contests, allocations that they believe a large part of their peers consider morally inappropriate.

7 Discussion

From whether to who and how much. The experimental literature on competition and morals asks a group-level, qualitative question: does competition affect moral behavior? Its answers vary with design (Huber et al., 2023), and without a theory of how much erosion a given design should produce, that variation is hard to interpret. Our potential contribution is to turn the question into individual-level, quantitative predictions: given each person's measured moral valuation and the structure of the competition, who should erode, and by how much. The method that makes this possible combines individual moral elicitations with contest theory as a tool: each participant's moral cost is measured beforehand, and the equilibrium of each group's actual game is computed from these measurements alone. In the pilot, the predictions hold for the level under AON, for which participants resist competition, and for how they respond to their rivals; they fail, in a direction the contest literature anticipates, for the level under CONS.

Incentives are the mechanism, not a confound. About half of the designs in Huber et al. (2023) remove incentive effects through their choice of controls, to isolate an effect of competition as such. This is coherent if the question is whether competition matters beyond its incentives. But the concern that motivates the question (Shleifer, 2004; Braganza, 2022) is arguably whether morals erode *given* the incentives that competition creates. Designing those incentives out answers a different question, and invites the misleading conclusion that competition has limited effects on moral behavior. Our approach keeps the incentives and predicts their consequences instead. Under AON the strategic incentives of the contest alone predict that play collapses toward the proxy, from a no-contest bid of €3.11 to an equilibrium bid of €4.43, and that is what we observe, although survivors are paid no more than without competition. This erosion is the rational response to the rules; a design that controls for incentives would not see it, yet it is the case that matters in institutions that select on a proxy. The approach then separates this incentive-driven erosion from erosion beyond it, which is where the CONS condition departs from the benchmark.

Erosion beyond equilibrium. Because the costs are measured, the excess erosion under CONS has a size in the model's own terms: relative to their elicited moral costs, participants act as if winning were worth about nine times its monetary prize, or as if their moral valuation during competition were about a ninth of the one elicited individually. Over-bidding relative to Nash is the most robust regularity of contest experiments (Sheremeta, 2013), here found with moral rather than monetary costs. Which of the two readings operates is an additional question; the lower valuations measured after competing point to the second, but lack the control that would attribute them to competition.

Limitations. This is a pilot, and all results are exploratory. Condition order is confounded with session; group sizes vary for accidental reasons, and neither theory nor data suggest a clear group-size effect beyond noisy pairs (Appendix G, Figures 8 and 9). Several choices were made after looking at the data: the opt-out outcome in Figure 4, the parameterization of the calibration, the design points shown in Figure 5, and the survival analysis (Appendix I). All p-values treat participants as independent and are descriptive. Four alternative readings of the CONS deviation remain open to different degrees. *Mere repetition* (Bartling et al., 2023) cannot explain it: even a baseline that attributes the entire fall in valuations between the two IDM blocks to time or repetition leaves over-bidding of about a quarter (Section 6.4). *Logit decision noise* (Anderson et al., 1998; Gneezy and Smorodinsky, 2006) can reproduce the mean CONS bid only as random play, which contradicts both the ranking of CONS play by valuations and the AON result (Appendix D). *Perceived crowding-out*, if participants treat the rivals' donations they see as paid, could be of the size of the deviation and shares the sign of the rivals effect (Appendix E). And the IDM block may *overstate moral costs*, for instance through experimenter demand, although the curves rank tournament play well and valuations fall after competing in two instruments.

Open questions. The pilot answers whether measured valuations and contest theory predict play; it cannot yet say why CONS play exceeds the benchmark. A condition without competition between the two IDM blocks would separate competition from repetition, demand and the sight of rivals' choices; varying the money prize would separate a value of winning from a lower moral valuation, since only the latter makes play sensitive to the prize; and counterbalancing condition order within sessions would remove its confound with session. A longer first IDM block would sharpen the curvature estimate on which every prediction rests.

8 Related literature

Tournaments and relative performance pay. Rank-order tournaments were introduced as optimal labor contracts (Lazear and Rosen, 1981), and experiments since Bull et al. (1987) and Orrison et al. (2004) test their incentive effects on effort; real-effort designs in the tradition of Gneezy et al. (2003) compare tournament with piece-rate pay, most prominently to study who selects into and performs under competition (Niederle and Vesterlund, 2011); see Carpenter and Matthews (2023) for a review. In most of these designs winners are paid a multiple of the piece rate, so competition and higher-powered incentives move together. Our tournaments pay survivors exactly the IDM payout, and the moral cost is part of the proxy choice itself rather than an effort cost.

Competition, cheating and sabotage. A large experimental strand adds an unethical action to the tournament: misreporting performance, lying, or sabotaging rivals. Competition raises cheating relative to piece rates (Schwieren and Weichselbaumer, 2010; Cadsby et al., 2010), sabotage responds to the prize spread and the number of winners (Harbring and Irlenbusch, 2008, 2011), also in the field (Balafoutas et al., 2012), lowers output (Carpenter et al., 2010) and differs by gender (Dato and Nieken, 2014), and lying responds to tournament incentives (Conrads et al., 2014). Dishonesty grows with the size of the competition (Chui et al., 2021), with information about rivals over repeated rounds (Benistant et al., 2022), and after winning (Schurr and Ritov, 2016). Dato et al. (2024) separate the incentive effect of competition on lying from its behavioral effect, the distinction our Discussion argues should be predicted rather than designed out. In this strand the unethical act is a separate, typically binary choice whose moral cost is neither measured nor modeled. In ours it is the degree to which the ranked proxy is pursued at the expense of the goal, its cost is elicited for each participant, and the predicted amount of erosion follows from equilibrium.

Competition and moral behavior beyond the contest. Competition can spill over into later social behavior: it lowers subsequent cooperation (Buser and Dreber, 2016) and interacts with helping in teams (Danilov et al., 2019), while social norms can override competitive pressure in labor markets (Fehr et al., 1998). In markets, the erosion of moral behavior (Shleifer, 2004; Falk and Szech, 2013) depends on market design and on where it is measured (Kirchler et al., 2016; Sutter et al., 2020); it has been decomposed into social information, framing and diffusion of responsibility (Irlenbusch and Saxler, 2019), linked to the replacement logic (Ziegler et al., 2024), attributed to repetition rather than market interaction (Bartling et al., 2023), and found to be limited by clear social norms (Bartling and Özdemir, 2023); see also Bartling et al. (2015). Beyond the laboratory, market exposure has been related to moral values across societies (Enke, 2023), and the moral effects of markets have been formalized theoretically (Dewatripont and Tirole, 2024). Markets can shape preferences (Bowles, 1998) and crowd out intrinsic motivation (Frey and Jegen, 2001). Across designs the average effect of competition on moral behavior is small and heterogeneous (Huber et al., 2023). We study the case in which the moral cost is incurred by optimizing the very measure on which agents are ranked, and we predict behavior person by person from measured valuations.

Proxy failure and intrinsic motivation. Rewarding a measurable proxy diverts effort from the unmeasured goal (Goodhart, 1984; Kerr, 1975; Holmström and Milgrom, 1991; Manheim and Garrabrant, 2018; John et al., 2024); mission-motivated agents respond to incentives in specific ways (Besley and Ghatak, 2005; Bénabou and Tirole, 2016). In science, selection on publication-based proxies has been modeled as a driver of low power and irreproducibility (Smaldino and McElreath, 2016; Higginson and Munafò, 2016; Braganza, 2020; Tiokhin et al.,

2021), and competition for grants can dissipate much of their value (Gross and Bergstrom, 2019). Braganza (2022) generalizes the mechanism to competitive societal systems.

Contests. Rank-order competition with a private cost of competing connects to contests with non-monetary costs such as doping and cheating (Berentsen, 2002; Kräkel, 2007; Gilpatric, 2011). All-pay auctions with complete information have mixed or asymmetric equilibria (Baye et al., 1996; Barut and Kovenock, 1998; Siegel, 2009), and the proxy corner acts as a bid cap (Che and Gale, 1998). Experiments document over-bidding and heterogeneous, often bimodal play (Dechenaux et al., 2015; Sheremeta, 2013; Müller and Schotter, 2010), which logit equilibria with noisy decisions can reproduce (Anderson et al., 1998; Gneezy and Smorodinsky, 2006). Under AON the agent's own moral value is part of the prize, which relates the design to electoral competition between policy-motivated candidates (Wittman, 1983; Calvert, 1985), and the all-pay versus winner-pay contrast relates it to charitable fundraising (Goeree et al., 2005).

References

Simon P. Anderson, Jacob K. Goeree, and Charles A. Holt. Rent seeking with bounded rationality: An analysis of the all-pay auction. *Journal of Political Economy*, 106(4):828–853, 1998.

James Andreoni. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal*, 100(401):464–477, 1990.

Loukas Balafoutas, Florian Lindner, and Matthias Sutter. Sabotage in Tournaments: Evidence from a Natural Experiment. *Kyklos*, 65(4):425–441, 2012. doi: 10.1111/kykl.12000.

Björn Bartling and Yagiz Özdemir. The limits to moral erosion in markets: Social norms and the replacement excuse. *Games and Economic Behavior*, 138:143–160, 2023. doi: 10.1016/j.geb.2022.12.001.

Björn Bartling, Roberto A. Weber, and Lan Yao. Do markets erode social responsibility? *The Quarterly Journal of Economics*, 130(1):219–266, 2015. doi: 10.1093/qje/qju031.

Björn Bartling, Ernst Fehr, and Yagiz Özdemir. Does market interaction erode moral values? *The Review of Economics and Statistics*, 105(1):226–235, 2023.

Yasar Barut and Dan Kovenock. The symmetric multiple prize all-pay auction with complete information. *European Journal of Political Economy*, 14(4):627–644, 1998. doi: 10.1016/S0176-2680(98)00029-9.

Michael R. Baye, Dan Kovenock, and Casper G. de Vries. The all-pay auction with complete information. *Economic Theory*, 8(2):291–305, 1996. doi: 10.1007/BF01211819.

Roland Bénabou and Jean Tirole. Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, 124(2), 2016. doi: 10.1086/684853.

Julien Benistant, Fabio Galeotti, and Marie Claire Villeval. Competition, information, and the erosion of morals. *Journal of Economic Behavior & Organization*, 204:148–163, 2022. doi: 10.1016/j.jebo.2022.10.008.

Aleksander Berentsen. The economics of doping. *European Journal of Political Economy*, 18 (1):109–127, 2002.

Timothy Besley and Maitreesh Ghatak. Competition and incentives with motivated agents. *American Economic Review*, 95(3):616–636, 2005.

Samuel Bowles. Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions. *Journal of Economic Literature*, 36(1):75–111, 1998.

Oliver Braganza. A simple model suggesting economically rational sample-size choice drives irreproducibility. *PLOS ONE*, 15(3):e0229615, 2020. doi: 10.1371/journal.pone.0229615.

Oliver Braganza. Proxyeconomics, a theory and model of proxy-based competition and cultural evolution. *Royal Society Open Science*, 9(2):211030, 2022. doi: 10.1098/rsos.211030.

Clive Bull, Andrew Schotter, and Keith Weigelt. Tournaments and Piece Rates: An Experimental Study. *Journal of Political Economy*, 95(1):1–33, 1987. doi: 10.1086/261439.

Thomas Buser and Anna Dreber. The Flipside of Comparative Payment Schemes. *Management Science*, 62(9):2626–2638, 2016. doi: 10.1287/mnsc.2015.2257.

C. Bram Cadsby, Fei Song, and Francis Tapon. Are You Paying Your Employees to Cheat? An Experimental Investigation. *The B.E. Journal of Economic Analysis & Policy*, 10(1), 2010. doi: 10.2202/1935-1682.2481.

Randall L. Calvert. Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence. *American Journal of Political Science*, 29(1):69–95, 1985. doi: 10.2307/2111212.

Jeffrey Carpenter, Peter Hans Matthews, and John Schirm. Tournaments and Office Politics: Evidence from a Real Effort Experiment. *American Economic Review*, 100(1):504–517, 2010. doi: 10.1257/aer.100.1.504.

Jeffrey P. Carpenter and Peter Hans Matthews. Tournaments and Competition. In Klaus F. Zimmermann, editor, *Handbook of Labor, Human Resources and Population Economics*, pages 1–36. Springer International Publishing, 2023. doi: 10.1007/978-3-319-57365-6_120-1.

Yeon-Koo Che and Ian L. Gale. Caps on political lobbying. *American Economic Review*, 88(3): 643–651, 1998.

Celia Chui, Maryam Kouchaki, and Francesca Gino. "Many others are doing it, so why shouldn't I?": How being in larger competitions leads to more cheating. *Organizational Behavior and Human Decision Processes*, 164:102–115, 2021. doi: 10.1016/j.obhdp.2021.01.004.

Julian Conrads, Bernd Irlenbusch, Rainer Michael Rilke, Anne Schielke, and Gari Walkowitz. Honesty in tournaments. *Economics Letters*, 123(1):90–93, 2014. doi: 10.1016/j.econlet.2014.01.026.

Anastasia Danilov, Christine Harbring, and Bernd Irlenbusch. Helping under a combination of team and tournament incentives. *Journal of Economic Behavior & Organization*, 162: 120–135, 2019. doi: 10.1016/j.jebo.2019.03.013.

Simon Dato and Petra Nieken. Gender differences in competition and sabotage. *Journal of Economic Behavior & Organization*, 100:64–80, 2014. doi: 10.1016/j.jebo.2014.01.012.

Simon Dato, Eberhard Feess, and Petra Nieken. Lying in competitive environments: Identifying behavioral impacts. *European Economic Review*, 170:104844, 2024. doi: 10.1016/j.euroecorev.2024.104844.

Emmanuel Dechenaux, Dan Kovenock, and Roman M. Sheremeta. A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18(4): 609–669, 2015. doi: 10.1007/s10683-014-9421-0.

Mathias Dewatripont and Jean Tirole. The Morality of Markets. *Journal of Political Economy*, 2024. doi: 10.1086/729445.

Benjamin Enke. Market exposure and human morality. *Nature Human Behaviour*, 7(1):134–141, 2023. doi: 10.1038/s41562-022-01480-x.

Armin Falk and Nora Szech. Morals and markets. *Science*, 340(6133):707–711, 2013. doi: 10.1126/science.1231024.

Ernst Fehr, Erich Kirchler, Andreas Weichbold, and Simon Gächter. When Social Norms Overpower Competition: Gift Exchange in Experimental Labor Markets. *Journal of Labor Economics*, 16(2):324–351, 1998. doi: 10.1086/209891.

Bruno S. Frey and Reto Jegen. Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5):589–611, 2001. doi: 10.1111/1467-6419.00150.

Scott M. Gilpatric. Cheating in contests. *Economic Inquiry*, 49(4):1042–1053, 2011.

Uri Gneezy and Rann Smorodinsky. All-pay auctions—an experimental study. *Journal of Economic Behavior & Organization*, 61(2):255–275, 2006. doi: 10.1016/j.jebo.2004.09.013.

Uri Gneezy, Muriel Niederle, and Aldo Rustichini. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3):1049–1074, 2003. doi: 10.1162/00335530360698496.

Jacob K. Goeree, Emiel Maasland, Sander Onderstal, and John L. Turner. How (not) to raise money. *Journal of Political Economy*, 113(4):897–918, 2005.

Charles A. E. Goodhart. Problems of monetary management: The UK experience. In *Monetary Theory and Practice*. Macmillan, 1984.

Kevin Gross and Carl T. Bergstrom. Contest models highlight inherent inefficiencies of scientific funding competitions. *PLOS Biology*, 17(1):e3000065, 2019. doi: 10.1371/journal.pbio.3000065.

Christine Harbring and Bernd Irlenbusch. How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization*, 65(3-4):682–702, 2008. doi: 10.1016/J.JEBO.2006.03.004.

Christine Harbring and Bernd Irlenbusch. Sabotage in Tournaments: Evidence from a Laboratory Experiment. *Management Science*, 57(4):611–627, 2011. doi: 10.1287/mnsc.1100.1296.

Andrew D. Higginson and Marcus R. Munafò. Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology*, 14(11):e2000995, 2016. doi: 10.1371/journal.pbio.2000995.

Bengt Holmström and Paul Milgrom. Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7:24–52, 1991. doi: 10.1093/jleo/7.special_issue.24.

Christoph Huber, Anna Dreber, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Utz Weitzel, et al. Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences*, 120(23):e2215572120, 2023. doi: 10.1073/pnas.2215572120.

Steffen Huck, Hans-Theo Normann, and Jörg Oechssler. Two are few and four are many: Number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4):435–446, 2004.

Bernd Irlenbusch and David J. Saxler. The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior. *Journal of Behavioral and Experimental Economics*, 80:141–161, 2019. doi: 10.1016/j.socec.2019.04.001.

Yohan J. John, Leigh Caldwell, Dakota E. McCoy, and Oliver Braganza. Dead rats, dopamine, performance metrics, and peacock tails: Proxy failure is an inherent risk in goal-oriented systems. *Behavioral and Brain Sciences*, 47:e67, 2024. doi: 10.1017/S0140525X23002753.

Steven Kerr. On the folly of rewarding A, while hoping for B. *Academy of Management Journal*, 18(4):769–783, 1975. doi: 10.5465/255378.

Michael Kirchler, Jürgen Huber, Matthias Stefan, and Matthias Sutter. Market Design and Moral Behavior. *Management Science*, 62(9):2615–2625, 2016. doi: 10.1287/mnsc.2015.2246.

Matthias Kräkel. Doping and cheating in contest-like situations. *European Journal of Political Economy*, 23(4):988–1006, 2007.

Erin L. Krupka and Roberto A. Weber. Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11 (3):495–524, 2013. doi: 10.1111/jeea.12006.

Edward P. Lazear and Sherwin Rosen. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 89(5):841–864, 1981. doi: 10.1086/261010.

David Manheim and Scott Garrabrant. Categorizing variants of Goodhart's law, 2018.

Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38, 1995.

Wieland Müller and Andrew Schotter. Workaholics and dropouts in organizations. *Journal of the European Economic Association*, 8(4):717–743, 2010.

Muriel Niederle and Lise Vesterlund. Gender and Competition. *Annual Review of Economics*, 3(1):601–630, 2011. doi: 10.1146/annurev-economics-111809-125122.

Alannah Orrison, Andrew Schotter, and Keith Weigelt. Multiperson Tournaments: An Experimental Examination. *Management Science*, 50(2):268–279, 2004. doi: 10.1287/mnsc.1030.0128.

Amos Schurr and Ilana Ritov. Winning a competition predicts dishonest behavior. *Proceedings of the National Academy of Sciences*, 113(7):201515102, 2016. doi: 10.1073/pnas.1515102113.

Christiane Schwieren and Doris Weichselbaumer. Does competition enhance performance or cheating? A laboratory experiment. *Journal of Economic Psychology*, 31(3):241–253, 2010. doi: 10.1016/j.joep.2009.02.005.

Roman M. Sheremeta. Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys*, 27(3):491–514, 2013.

Andrei Shleifer. Does competition destroy ethical behavior? *American Economic Review*, 94 (2):414–418, 2004.

Ron Siegel. All-pay contests. *Econometrica*, 77(1):71–92, 2009. doi: 10.3982/ECTA7537.

Paul E. Smaldino and Richard McElreath. The natural selection of bad science. *Royal Society Open Science*, 3(9):160384, 2016. doi: 10.1098/rsos.160384.

Matthias Sutter, Jürgen Huber, Michael Kirchler, Matthias Stefan, and Markus Walzl. Where to look for the morals in markets? *Experimental Economics*, 23(1):30–52, 2020. doi: 10.1007/s10683-019-09608-z.

Leonid Tiokhin, Minhua Yan, and Thomas J. H. Morgan. Competition for priority harms the reliability of science, but reforms can help. *Nature Human Behaviour*, 5(7):857–867, 2021. doi: 10.1038/s41562-020-01040-1.

Donald Wittman. Candidate motivation: A synthesis of alternative theories. *American Political Science Review*, 77(1):142–157, 1983.

Andreas G. B. Ziegler, Giorgia Romagnoli, and Theo Offerman. Morals in multi-unit markets. *Journal of the European Economic Association*, 22(5):2225–2260, 2024. doi: 10.1093/jeea/jvae001.

A Sample

session	date	participants	groups (n, s)
1	2026-09-17	12	(6,3) × 2
2	2026-09-18	11	(2,1), (3,·), (6,3)
3	2026-09-18	10	(4,2), (6,3)
4	2026-09-25	12	(2,1), (4,2), (6,3)
5	2026-09-28	11	(2,1), (4,2), (5,·)
6	2026-09-28	12	(2,1) × 2, (4,2) × 2
total		68	17 groups

Table 1: Sessions. $(3, \cdot)$ and $(5, \cdot)$: odd groups whose cutoff is drawn each round.

B Estimation

The likelihood treats each individual choice as a draw from $\text{clip}(x_i^* + \varepsilon, 0, 1)$, censored within 0.05 of either end, with x_i^* from (8) under (3) and the random effect integrated by Gauss–Hermite quadrature. The level is weakly identified: the profile likelihood changes by less than 2 log-likelihood units over a three-fold range of $a = e^{\mu}$, along which b and τ rise with μ, while θ stays nearly constant and θ/b varies between 0.052 and 0.064. We therefore impose that the implied pooled mean equals the observed one and report the ratios. Alternative families (a power and a linear marginal value) fit worse by $\Delta\text{AIC} \geq 29.6$.

C Out-of-sample protocol

Predictions use curves fitted on the first IDM block. The benchmarks are: the no-contest optimum (L0); the parameter-free equilibrium (L2); both with a common shift δ in α chosen on five sessions and scored on the sixth; and a no-curves benchmark that predicts every participant with the other sessions' pooled choices in the same condition.

condition	predictor	bias	MAE	rank ρ
AON	no curves	0.00	0.122	—
	L0	+0.28	0.288	0.36
	L2	0.00	**0.112**	0.26
	L2 + δ	−0.01	0.116	0.22
CONS	no curves	0.00	0.244	—
	L0	+0.11	0.183	**0.74**
	L0 + δ	−0.09	**0.163**	0.74
	L2	+0.26	0.292	0.69
	L2 + δ	+0.03	0.207	0.61

Table 2: Held-out prediction of participants' mean position, rounds 2–15. δ is chosen between −2.0 and −3.0 in every fold. On the log score of individual choices the no-curves benchmark is best in both conditions: the position noise of the IDM block spreads the predictive distribution too widely at the corner, so the curves predict means and rankings rather than the shape of each person's choices.

D A value of winning or a lower moral valuation

With the same protocol, a value of winning π_0 chosen on a grid up to €40 selects €40 in every fold and predicts the held-out CONS level with bias +0.02 and MAE 0.183, against +0.04 and 0.207 for δ; on the log score the two are −2.67 and −2.77. The two are nearly observationally

equivalent under CONS (Section 6.2). A per-person analysis that avoids sharing any first-block estimate between the two shifts (split halves of the first block) finds that the tournament shift and the after-block shift covary ($\rho \approx 0.25$), a group difference between participants who play CONS as money-maximizers and the rest rather than a graded relation.

Logit noise. Reading the precision λ of the logit path as behavior rather than as a solver device, we keep each group's realized logit equilibrium at every λ on the path (from 0.1 to 3000 per euro, same curves, position noise and rules) and choose λ by the log-likelihood of the observed choices (rounds 2–15). Choosing uniformly among the 40 options gives a bid of €3.74; with the position noise added, €3.58, almost exactly the observed CONS bid of €3.60. AON choices are fitted best at the Nash end (bias 0.00, over-bidding -2%). CONS choices are fitted best by fully random choice ($\lambda \to 0$), which matches the CONS level but assigns every participant the same distribution, whatever their valuation, whereas observed CONS play is ranked by valuations (rank correlation 0.74). A common λ settles near 0.1 per euro and mispredicts both conditions (AON: bias +0.14, over-bidding +16%; CONS: over-bidding +21%). Logit noise thus cannot reconcile the two conditions, nor the ranking within CONS.

E Altruism toward the total donation

The model assumes that participants value their own donation. Under pure altruism they would value the total donation instead, and rivals' donations could crowd out their own. Taken at face value, this does not change our predictions: the paid round is drawn independently for each participant from all blocks, so the donations of others a participant can expect to be paid are the same in the IDM block and in the tournament, and any crowding-out is already absorbed in the elicited valuation. Moreover, relative to the budget of a large charity, the donations at stake are negligible, so altruism toward the charity's total would imply a nearly constant marginal moral value, whereas the elicited one declines steeply in the participant's own donation. A behavioral variant remains. In the tournament, unlike in the IDM block, participants see their rivals' choices every round; a participant who treats the donations they see as if they were paid would give less when rivals give more. Because the elicited curves are strongly concave, such perceived crowding-out could be of the size of the CONS deviation, and it predicts a rivals effect of the same sign as the strategic one. Under AON it would change little: rivals near the proxy corner give little, only survivors' donations count, and even a common drop in moral valuation large enough to account for CONS play (δ between -2 and -3) leaves the AON prediction essentially unchanged (Appendix C). It would lower moral behavior in the tournament without an erosion of moral valuations. Showing participants each other's choices without competition would test it.

F Condition order

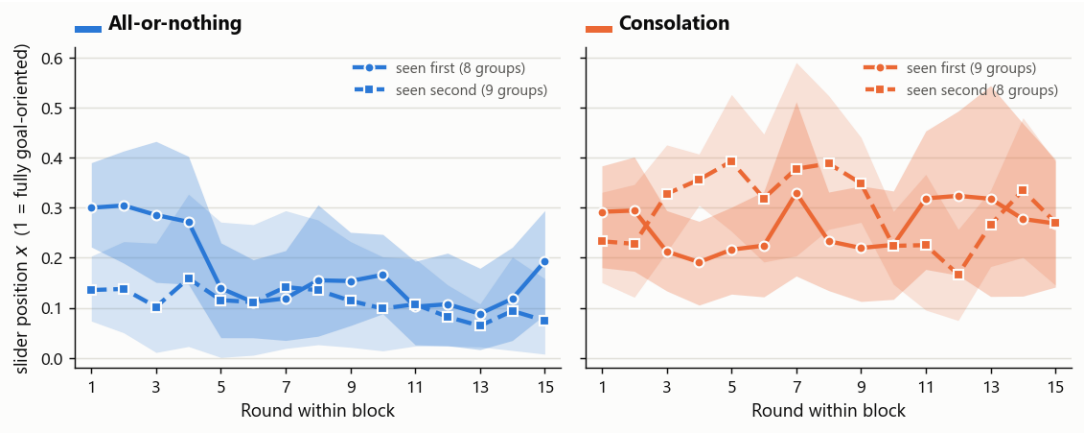

Figure 7: Mean slider position per round, averaged over groups, by whether the group met the condition first or second (8 groups AON first, 9 CONS first). Bands: 95% bootstrap intervals over groups. Order is confounded with session.

G Group size

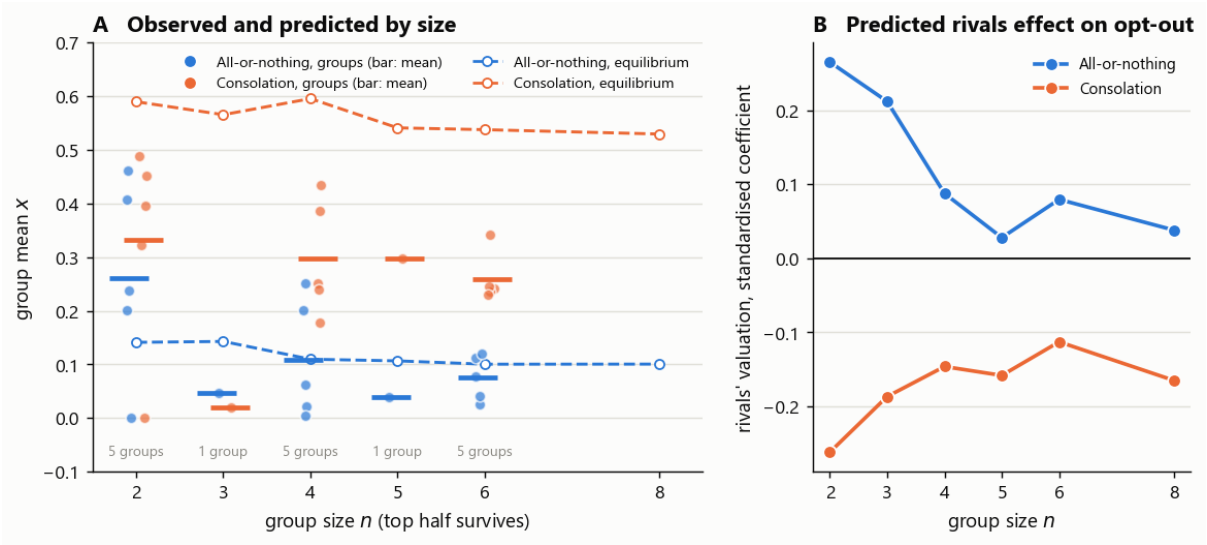

Figure 8: (A) Observed group means by size, with the equilibrium for the pilot's participants re-drawn into random groups of each size (30 draws per size). (B) The equilibrium's rivals effect on opt-out by group size. The level is predicted to depend little on size; the pilot's AON pairs resist the collapse more than larger groups and more than predicted, consistent with tacit cooperation in repeated duopolies (Huck et al., 2004). Sizes 3 and 5 have one group each.

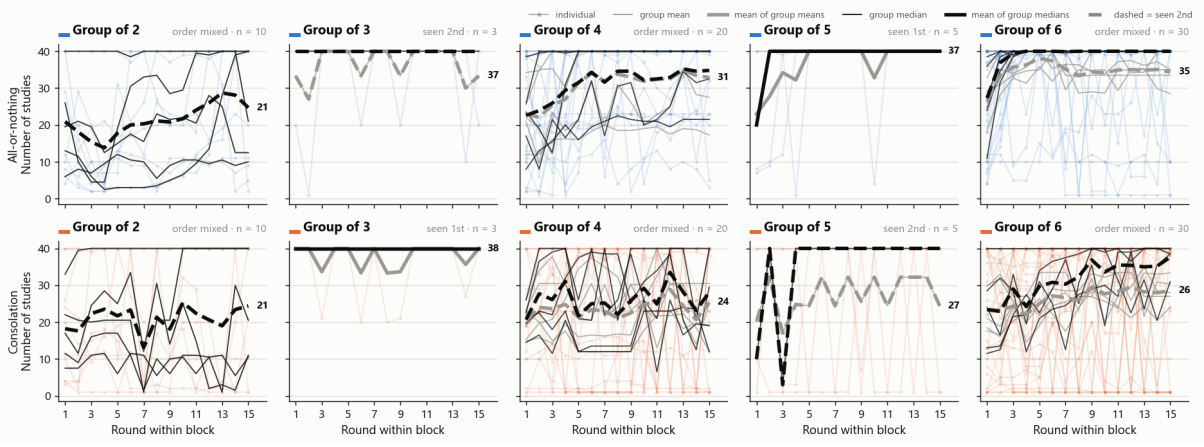

Figure 9: Play by condition and group size: number of studies per round (40 = fully proxy-oriented), rounds aligned on each participant's own order of conditions. Faint lines: participants; thin grey and black lines: each group's mean and median; thick lines: their average over the groups of that size, dashed where the condition was met second. The label above each panel gives whether the participants in it met the condition first, second, or in mixed order, and their number.

H Mean position

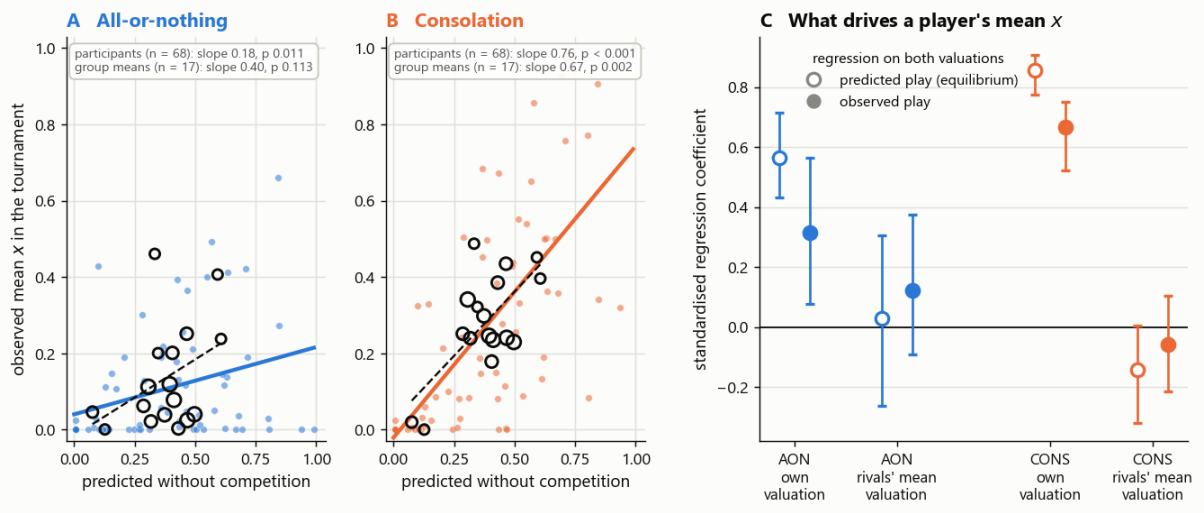

Figure 10: As Figure 4, for each participant's mean slider position instead of their opt-out share. Under CONS the participant slope is 0.76 ($p < 0.001$) and the group-mean slope 0.67 ($p = 0.002$); the rivals effect is -0.14 in the equilibrium and -0.06 in the data.

I Who survives

Figure 11 shows the observed sorting on survival and Figure 12 the benchmark's prediction of it. The benchmark gives each participant a survival probability W_i at the equilibrium of their group's game, computed on the noiseless ranking (Section 5). W_i is a probability, whereas observed survival is a share of 15 rounds with ties broken at random. Round-to-round noise therefore weakens the observed correlation of α_i with survival but not its correlation with W_i, so the latter (-0.70 under AON, -0.87 under CONS) is not comparable with the observed one. Instead we simulate each group's 15 rounds from the equilibrium distributions, 400 times per group, and score the simulated shares like the data. This gives the correlation the benchmark expects and its range across simulations: -0.32 (-0.52 to -0.10) under AON, against an observed

-0.35 (41% of simulations are at or below it), and -0.85 (-0.87 to -0.83) under CONS, against an observed -0.67, weaker than every simulation. With the shift δ the expectation is -0.70 (-0.75 to -0.64) under CONS, matching the observed sorting, and -0.10 under AON, too weak (3% of simulations at or below the observed value), just as the AON level needs no δ. The simulations also give the ceiling of the correlation of W_i with observed survival if the benchmark were exact: 0.45 under AON and 0.98 under CONS, against observed 0.26 and 0.64.

Slopes need no simulation, because noise in survival does not attenuate them. The observed slope of survival on α_i is -0.025 under AON, against -0.017 (-0.024 to -0.012) for W_i, and -0.073 under CONS, against -0.129 (-0.158 to -0.116), or -0.090 (-0.116 to -0.069) with δ. By tercile of group-centered α_i, from least to most moral, participants survived 59%, 49% and 43% of AON rounds (predicted 55%, 52% and 45%) and 73%, 49% and 27% of CONS rounds (predicted 93%, 45% and 11%; with δ, 75%, 52% and 21%). Under CONS the shortfall thus lies at both ends: the most moral third survived more often than the equilibrium predicts, and the least moral third less often.

One caveat applies. α_i is a posterior mean, and its estimation error weakens the observed relations but not the simulated ones, in which α_i is exact. The comparison is thus tilted against the benchmark, and part of the CONS shortfall may be measurement error rather than erosion beyond the equilibrium.

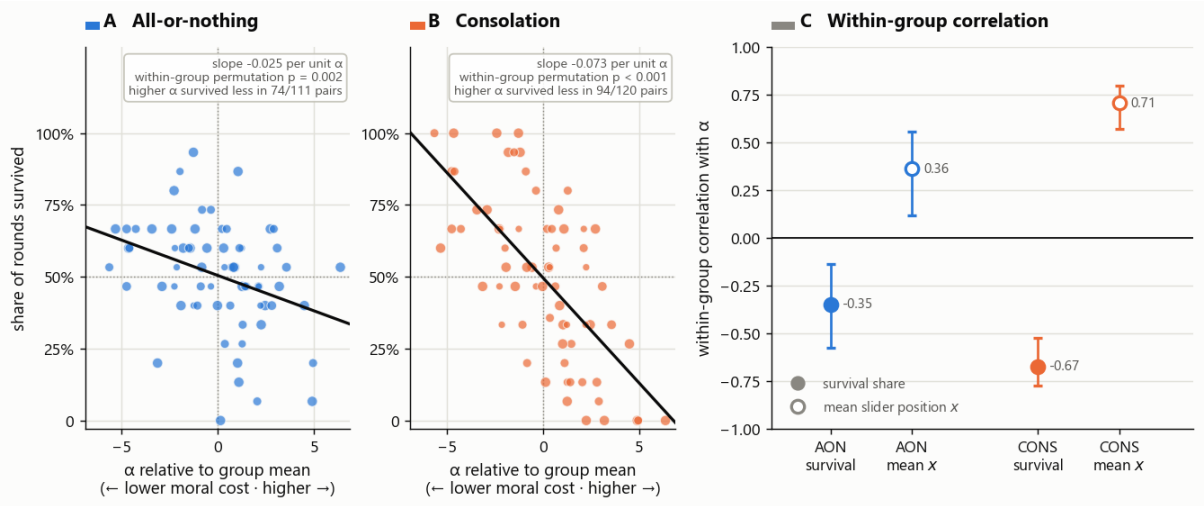

Figure 11: More moral participants survive fewer rounds. (A, B) Each participant's valuation α_i from the first IDM block, relative to their group's mean, against their share of the 15 rounds survived; point size: group size. Line: within-group slope; p: one-sided test of a negative slope, permuting α within groups; pairs: among the within-group pairs that differ in both α and survival, those in which the more moral member survived less often. (C) Within-group correlation of α_i with survival (filled) and, for comparison, with mean slider position (hollow); 95% bootstrap intervals over groups. 68 participants in 17 groups, no parameter from tournament data.

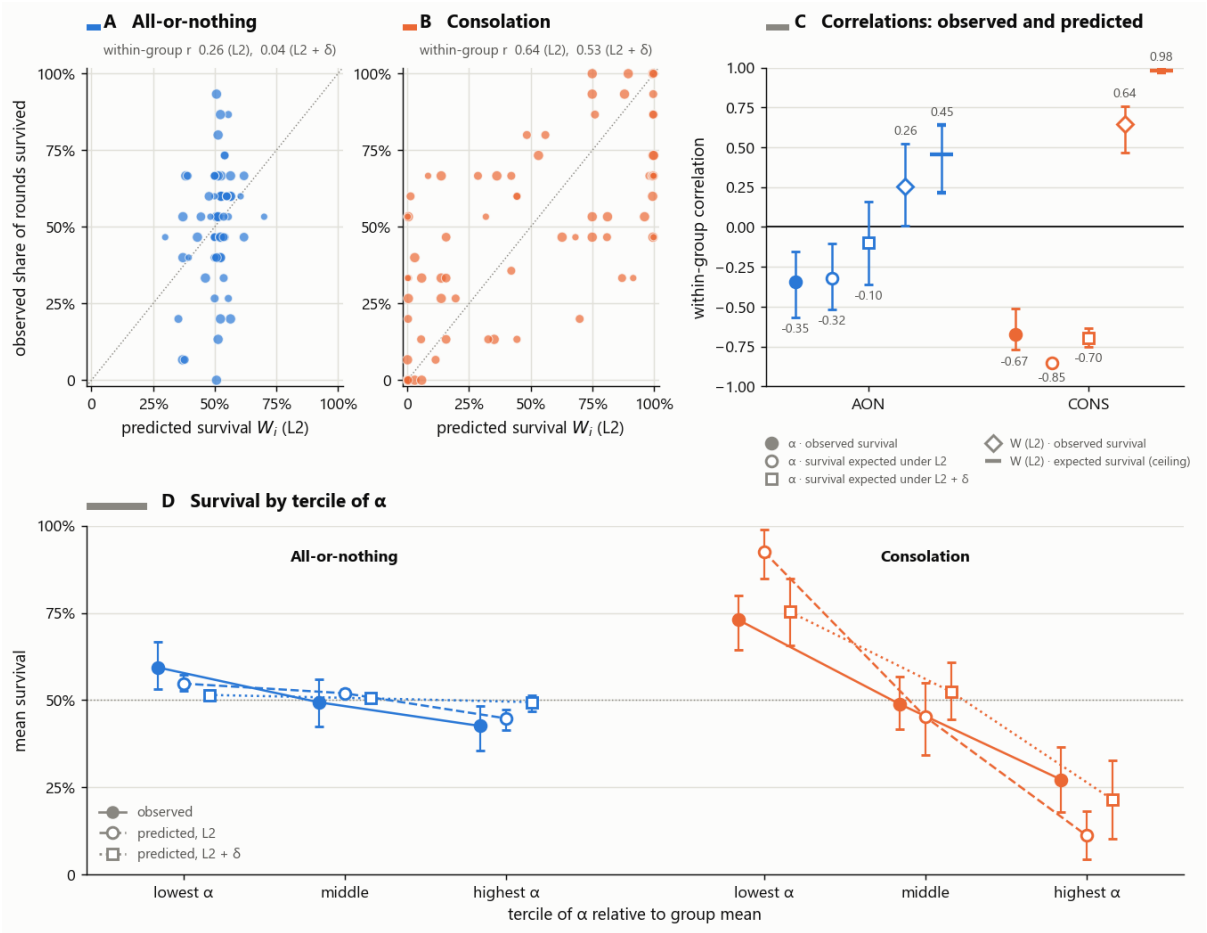

Figure 12: The benchmark's prediction of who survives. (A, B) Each participant's survival probability W_i at the equilibrium of their group's game (L2) against their observed share of rounds survived; above each panel, the within-group correlation for L2 and for L2 with the shift δ (Appendix C). Under AON with δ, all members of 15 of the 17 groups have the same W_i, so that correlation rests on two groups. (C) Within-group correlations of α_i with observed survival (filled circle) and with the survival the benchmark expects over the same 15 rounds (hollow circle: L2; square: L2 + δ), and of W_i with observed survival (diamond), against its ceiling if the benchmark were exact (bar). Expected values and ceilings: mean and 2.5–97.5% range of 400 simulations of each group's rounds. (D) Mean survival by tercile of group-centered α_i, observed (filled) and predicted (hollow: mean W_i). Curves from the first IDM block; no parameter from tournament data except δ, which is chosen on the other sessions. 68 participants in 17 groups; 95% bootstrap intervals over groups.