Supplemental Materials:

"Choosing more aggressive commitment contracts for others than for the self"

1 Studies 1, 4, and 5: Robustness to Organization Fixed Effects 2

2.1 Study 1: Satellite Condition Method and Results 3

2.2 Study 1: Additional Performance and Persistence Measures 5

3 Study 3: Individual Measures of Effectiveness and Moral Appropriateness..... 8

4.1 Study 4: Ratings of Effectiveness and Moral Appropriateness 9

4.2 Study 4: Regression Models Using Contract-Specific Ratings 10

4.3 Study 4: Logistic Regression Results 13

4.4 Study 4: Blinder-Oaxaca Decomposition 16

5.1 Study 5: Moderation and Mediation Analysis 19

5.2 Study 5: Regression Model with Separate Self and Friend Conditions..... 21

5.3 Study 5: Blinder-Oaxaca Decomposition 22

6 Supplemental Study 1: stickK Contract Analysis..... 24

1 Studies 1, 4, and 5: Robustness to Organization Fixed Effects

For Studies 1, 4, and 5, we examine the self-other difference in contract choice while statistically adjusting for the particular nonprofit organizations used in the choice menus. For each study, we first conducted a logistic regression with contract choice as the outcome variable, and personal-surrogate condition as the independent variable. To test for robustness to the specific organizations (selected by participants) that populated the choice menus we then added organization fixed effects to the model. Table 1.1 displays the average marginal effects (i.e., predicted probabilities) for each model for Studies 1, 4, and 5. Column 1 reports the result of the model without organization fixed effects and Column 2 reports the results with organization fixed effects included. As shown, all self-other differences remain significant when controlling for the specific organizations used in the choice menus.

Table 1.1. *Anti-charity contract choices*

	Study 1		Study 4		Study 5	
	(1)	(2)	(1)	(2)	(1)	(2)
Surrogate Choice	0.291*** (0.04)	0.280*** (0.042)	0.109*** (0.031)	0.104*** (0.031)	0.122*** (0.031)	0.122*** (0.032)
Intercept	0.040* (0.018)	–0.096*** (0.089)	0.387*** (0.022)	0.277*** (0.053)	0.457*** (0.018)	0.310*** (0.050)
Organization Fixed Effects	No	Yes	No	Yes	No	Yes
Adjusted R^2	.136	.154	.011	.024	.012	.019

Note: For each study, Model 1 displays the average marginal effects from a logistic regression that regresses contract choice (0 = pro-charity commitment contract, 1 = anti-charity contract) onto personal-surrogate condition (0 = personal choice, 1 = surrogate choice) without organization fixed effects. Model 2 includes the fixed effects for organizations selected by participants as their pro/anti-charity. Robust standard errors in parentheses. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

2.1 Study 1: Satellite Condition Method and Results

Method

In addition to the participants recruited in Study 1, we recruited 86 participants (42% female, mean age = 29 years, range = 18 to 50 years) from Prolific Academic and assigned them to a satellite condition. First, participants selected the organization they most supported and the organization they most opposed from a list of 10 nonprofit organizations. We then randomly assigned participants to complete 10 rounds of the letter-counting task under either a personally relevant pro-charity or anti-charity contract (based on the organizations previously selected). We informed participants they would start the task with a $2.50 bonus endowment, and their performance on the task would determine the final amount they would receive at the end of the study. For each unsuccessful round, $0.25 of the participant's bonus would be forfeited and donated to either their pro-charity or anti-charity organization, based on condition. No money was lost for each round answered successfully. After completing the 10 rounds, participants completed the same manipulation check as in Study 1 and answered demographic questions.

Results

As a manipulation check of the organizations used to populate the pro-charity and anti-charity commitment contracts, participants reported greater support for pro-charity organizations (M = 2.26, SD = 1.11) than for anti-charity organizations (M = –1.81, SD = 2.02; *paired* $t(85) = 16.56$, $p < .001$, $d = 1.79$).

Participants whose performance was bound to an anti-charity tended to successfully complete more rounds than participants under a pro-charity contract (M = 6.10, SD = 3.55 vs. M = 5.36, SD = 3.47), though the difference was not statistically significant ($t(85) = 0.97$, $p = .337$, $d = 0.21$). Participants in the anti-charity condition also tended to spend more time per

round (in seconds) than did participants in the pro-charity condition (M = 65.55, SD = 47.39 vs. M = 63.07, SD = 51.02), but this difference was also not statistically significant ($t(85) = 0.23$, $p = .816$, $d = 0.05$).

As detailed in our preregistration, we also examined whether task persistence (measured in seconds spent per round) decayed more rapidly over time under anti-charity versus pro-charity contracts. Using OLS, we regressed time spent per round onto commitment contract (0 = pro-charity, 1 = anti-charity), round order (treated as a continuous measure from round 1 to round 10), and the interaction term between contract and round order. Standard errors were clustered by participants. We did not observe a reliable interaction effect between contracts and round order, indicating that degradation in task persistence over time did not reliably differ across contract type ($b = –2.42$, $SE = 3.01$, $p = .422$). We also find a null interaction effect when using 2.5% trimmed means on time spent per round ($b = 0.94$, $SE = 1.13$, $p = .408$).

2.2 Study 1: Additional Performance and Persistence Measures

Performance over Task Rounds

We examined whether task performance (i.e., success rate per round) decayed more rapidly over time under anti-charity versus pro-charity contracts that participants chose for themselves in Study 1. Using logistic regression, we regressed performance per round (0 = unsuccessful, 1 = successful) onto commitment contract choice (0 = pro-charity, 1 = anti-charity), round number (treated as a continuous measure from round 1 to round 10), and the interaction term between contracts and round. Standard errors were clustered by participants. We do not observe a reliable interaction effect between contract type and round order, indicating that degradation in performance over time did not reliably differ across type of contract chosen ($b = 0.06$, $SE = 0.10$, $p = .567$).

Persistence over Task Rounds

We next examined whether task persistence (measured in seconds spent per round) decayed more rapidly over time under anti-charity versus pro-charity contracts. Using OLS, we regressed time spent per round onto commitment contract choice (0 = pro-charity, 1 = anti-charity), round number (treated as a continuous measure from round 1 to round 10), and the interaction term between contract type and round. Standard errors were clustered by participants. We observe a significant interaction between contract choice and task round ($b = 6.28$, $SE = 2.77$, $p = .024$). Persistence did not reliably decline over time for those participants who selected an anti-charity contract ($b = 0.70$, $SE = 4.07$, $p = .864$), whereas persistence strongly decreased over time for those who selected a pro-charity contract ($b = –5.58$, $SE = 0.83$, $p < .001$). That is, participants who selected an anti-charity contract spent a consistent amount of

time on each round of the task whereas those who selected a pro-charity contract spent significantly less time on each round as they progressed through the task.

However, when using 2.5% trimmed means for time spent per round (as specified in our preregistration analysis), the interaction term is no longer significant ($b = 2.26$, $SE = 2.13$, $p = .288$). We also measured persistence using a seven-second cut-off where we defined any round completed in over seven seconds as the participant "trying." Using this alternative measure of persistence, we again fail to observe a reliable interaction between contract choice and task round ($b = 0.12$, $SE = 0.17$, $p = .458$).

Performance by Personal-Surrogate Conditions

All participants completed ten rounds of the letter-finding task, but performance was only incentivized for participants in the personal choice condition; participants in the surrogate choice condition were asked to complete the same task, but their payment was determined only by the performance of their partner. Thus, we can compare performance in the two conditions to examine the role of incentives. Participants whose bonus payment was directly tied to their own performance (i.e., personal choice condition) correctly completed more rounds ($M = 5.54$, $SD = 3.42$) than participants whose performance was not incentivized (i.e., surrogate choice condition; $M = 4.69$, $SD = 3.52$), though the difference was only marginally significant ($t(250) = 1.95$, $p = .052$, $d = 0.24$).

We also tested for differences in performance over time across personal-surrogate conditions. Using logistic regression, we regressed performance per round (0 = unsuccessful, 1 = successful) onto personal-surrogate condition (0 = personal choice, 1 = surrogate choice), round number (treated as a continuous measure from round 1 to round 10), and the interaction term between condition and round. Standard errors were clustered by participants. We fail to find

a reliable interaction effect between personal-surrogate condition and round order, indicating that the degradation in performance over time did not reliably differ across conditions ($b = -0.00$, $SE = 0.03$, $p = .868$).

Persistence by Personal-Surrogate Condition

We analyzed persistence by condition, measured in average seconds spent per round. Participants whose bonus payment was determined by their own performance spent significantly more time on each round ($M = 66.31$, $SD = 45.14$) than those whose own performance was unincentivized ($M = 53.44$, $SD = 36.07$; $t(250) = 2.50$, $p = .013$, $d = 0.31$). We also examined whether time spent per round changed over time differed across the two conditions, using a similar OLS regression to the one we used to test for performance difference between contract types. We find a reliable interaction effect between personal-surrogate condition and task round ($b = 2.35$, $SE = 0.89$, $p = .008$). Surprisingly, there was a steeper decline in persistence per round over time for participants in the (incentivized) personal choice condition ($b = -5.33$, $SE = 0.69$, $p < .001$) than for participants in the (unincentivized) surrogate choice condition ($b = -2.98$, $SE = 0.68$, $p < .001$). However, the interaction between condition and round is no longer significant when using 2.5% trimmed means for time spent per round ($b = 0.80$, $SE = 0.51$, $p = .114$). We also measured persistence using a seven-second cut-off where we defined any round completed in over seven seconds as the participant "trying." Using this alternative measure of persistence, we again fail to observe a reliable interaction between condition and task round, ($b = 0.01$, $SE = 0.04$, $p = .855$).

3 Study 3: Individual Measures of Effectiveness and Moral Appropriateness

Table 3.1. *Ratings of moral appropriateness and effectiveness items by contract*

	Pro-Charity		Anti-Charity				
	M	*SD*	*M*	*SD*	*t*	*p*	*d*
Appropriate	5.63	1.41	2.81	1.69	14.24	< .001	1.16
Immoral (RC)	5.76	1.68	3.18	1.80	12.66	< .001	1.03
Effective	4.41	1.69	5.03	1.91	2.79	.006	0.23
Motivating	4.37	1.78	5.43	1.83	4.65	< .001	0.38

Note: RC = reverse coded; reported scores are calculated based on the reverse coding rather than the original coding. All items were measured on 7-point scales; 1 = *not at all*, 7 = *extremely*.

4.1 Study 4: Ratings of Effectiveness and Moral Appropriateness

Table 4.1. *Ratings of moral appropriateness and effectiveness by contract*

	Pro-Charity		Anti-Charity				
	M	*SD*	*M*	*SD*	*t*	*p*	*d*
Appropriateness$_{personal}$	5.63	1.95	3.87	2.28	13.06	< .001	0.58
Appropriateness$_{surrogate}$	5.87	1.74	4.02	2.07	15.25	< .001	0.68
Effectiveness$_{personal}$	4.44	1.73	4.86	2.12	3.62	< .001	0.16
Effectiveness$_{surrogate}$	3.98	1.59	5.04	1.69	9.53	< .001	0.43

Note: Appropriateness was measured using an average of three items: (i) how morally inappropriate would it be for [them/the other person] to use the contract; (ii) how morally negative were the potential consequences of using the contract, and (iii) how unethical it would be for [them/the other person] to take the risks associated with using the contract. Effectiveness was measured using an average of three items: (i) how effective the contract would be at helping [them/the other person] keep their commitment to exercise three times each week; (ii) how likely [they/the other person] would be to keep their commitment, and (iii) how successful [they/the other person] would be in keeping their commitment. All of the items were measured from 1 = *not at all* to 7 = *extremely*.

4.2 Study 4: Regression Models Using Contract-Specific Ratings

For the analysis reported in the main paper, we calculated difference scores between the indices of contract effectiveness for anti-charity and pro-charity contracts, and difference scores between the indices of contract appropriateness for anti-charity and pro-charity contracts. Here, we report the mediation and moderation analyses using each separate index of appropriateness and effectiveness as predictor variables, which returns similar results to those reported in the main text. Tables 4.2 and 4.3 display the results using a linear probability model. As shown in Table 4.2 (moderation analysis), only the interaction between effectiveness ratings of pro-charity contracts and condition is statistically significant. As shown in Table 4.3 (mediation analysis), both effectiveness ratings reliably mediate the self-other difference, while only the appropriateness rating of anti-charity contracts mediates the self-other difference.

Table 4.2. *Study 4 moderation results using separate indices of moral appropriateness and effectiveness*

	(1)	(2)	(3)	(4)	(5)	(6)
Surrogate Choice	0.109*** (0.031)	0.100*** (0.031)	0.042 (0.027)	–0.199 (0.116)	0.126 (0.100)	0.030 (0.128)
Appropriateness$_{pro\text{-}charity}$		0.011 (0.008)		–0.006 (0.011)		–0.007 (0.010)
Appropriateness$_{anti\text{-}charity}$		0.038*** (0.007)		0.030** (0.009)		0.028*** (0.008)
Effectiveness$_{pro\text{-}charity}$			–0.103*** (0.007)		–0.089*** (0.010)	–0.089*** (0.010)
Effectiveness$_{anti\text{-}charity}$			0.105*** (0.006)		0.100*** (0.008)	0.100*** (0.008)
Appropriateness$_{pro\text{-}charity\ x\ Surrogate\ Choice}$				0.039* (0.017)		0.029 (0.015)
Appropriateness$_{anti\text{-}charity\ x\ Surrogate\ Choice}$				0.018 (0.014)		0.007 (0.012)
Effectiveness$_{pro\text{-}charity\ x\ Surrogate\ Choice}$					–0.030* (0.015)	–0.030* (0.015)
Effectiveness$_{anti\text{-}charity\ x\ Surrogate\ Choice}$					0.008 (0.014)	0.000 (0.014)
Intercept	0.387*** (0.022)	0.175** (0.058)	0.334*** (0.049)	0.305*** (0.075)	0.295*** (0.059)	0.226** (0.084)
Adjusted-R^2	.011	.039	.286	.044	.287	.306

Note: Results from linear probability models, with robust standard errors in parentheses. The dependent variable in all models is choice of commitment contract (0 = pro-charity, 1 = anti-charity). "Surrogate choice" indicates whether participants were assigned to choose a commitment contract for themselves or for another person (0 = personal choice, 1 = surrogate choice). "Appropriateness$_{pro\text{-}charity}$" and "appropriateness$_{anti\text{-}charity}$" represent the appropriateness index for each type of contract, measured using three items from 1 = *not at all* to 7 = *extremely*. "Effectiveness$_{pro\text{-}charity}$" and "effectiveness$_{anti\text{-}charity}$" represent the effectiveness index for each type of contract, measured using three items also from 1 = *not at all*, 7 = *extremely*. Models 4–5 include separate interaction terms between surrogate choice condition and contract-specific ratings of appropriateness or effectiveness, respectively. Model 6 includes all interaction terms. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 4.3. *Study 4 mediation results using separate indices of moral appropriateness and effectiveness*

	(1)	(2)	(3)	(4)
Surrogate Choice	0.109*** (0.031)	0.100** (0.031)	0.042 (0.027)	0.037 (0.027)
Appropriateness$_{pro\text{-}charity}$		0.011 (0.008)		0.006 (0.007)
Appropriateness$_{anti\text{-}charity}$		0.038*** (0.007)		0.031*** (0.007)
Effectiveness$_{pro\text{-}charity}$			–0.103*** (0.007)	–0.103*** (0.007)
Effectiveness$_{anti\text{-}charity}$			0.105*** (0.006)	0.102*** (0.006)
Intercept	0.387*** (0.022)	0.175** (0.058)	0.334*** (0.049)	0.193** (0.064)
Adjusted R^2	.011	.039	.286	.304
Indirect Effects				
Appropriateness$_{pro\text{-}charity}$		0.003 [–0.001, 0.010]		0.001 [–0.001, 0.007]
Appropriateness$_{anti\text{-}charity}$		0.006 [–0.004, 0.017]		0.005 [–0.004, 0.015]
Effectiveness$_{pro\text{-}charity}$			0.047 [0.025, 0.071]	0.047 [0.025, 0.071]
Effectiveness$_{anti\text{-}charity}$			0.019 [–0.006, 0.044]	0.019 [–0.005, 0.045]

Note: Result from linear probability models in which the dependent variable is choice of commitment contract (0 = pro-charity, 1 = anti-charity), chosen for the self or other. Model 1 shows the basic self-other difference when selecting a contract for the self versus another person (0 = personal, 1 = surrogate). Model 2 shows the results of adding the appropriateness indices for pro-charity and anti-charity contracts as simultaneous mediators. Model 3 shows the results of adding the effectiveness indices as simultaneous mediators. Model 4 adds the indices of both appropriateness and effectiveness as simultaneous mediators. For all models, standard errors are reported in parentheses, and bootstrapped 95% bias corrected and accelerated confidence intervals are reported in brackets. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

4.3 Study 4: Logistic Regression Results

Below, we report the results of our mediation and moderation analyses using logistic regression rather than linear probability models. For each mediation analysis, we also adjust the test procedure to account for potential scaling confounds that can arise when comparing different models using binary choice data ("KHB adjustment"; Karlson et al., 2012). Table 4.4 reports the results from the moderation analysis and Table 4.5 reports the results from the mediation analysis (similar to Tables 2 and 3 in the main text). As shown in the tables, we find a similar pattern of results using logistic regression to those reported in the paper using linear probability models.

Table 4.4. *Study 4 moderation results (logistic regression)*

	(1)	(2)	(3)	(4)
Surrogate Choice	0.443*** (0.123)	0.412** (0.154)	0.075 (0.204)	0.014 (0.219)
Appropriateness		0.084** (0.030)		0.148*** (0.043)
Effectiveness			0.624*** (0.074)	0.668*** (0.080)
Surrogate Choice x Appropriateness		–0.025 (0.045)		0.070 (0.117)
Surrogate Choice x Effectiveness			0.105 (0.112)	–0.063 (0.063)
Intercept	–0.459*** (0.092)	–0.318 (0.106)	–0.945*** (0.131)	–0.669*** (0.166)
Pseudo R^2	.016	.030	.406	.423
Average Marginal Effects				
Self: $b_{\text{appropriateness}}$		0.020** (0.007)		0.025*** (0.006)
Other: $b_{\text{appropriateness}}$		0.015 (0.008)		0.014* (0.007)
Self: $b_{\text{effectiveness}}$			0.109*** (0.007)	0.113*** (0.007)
Other: $b_{\text{effectiveness}}$			0.120*** (0.006)	0.121*** (0.006)

Note: Models report log-odds coefficients from logistic regression, with robust standard errors in parentheses. The dependent variable in all models is choice of commitment contract (0 = pro-charity, 1 = anti-charity). "Surrogate choice" indicates whether participants were assigned to choose a commitment contract for themselves or for another person (0 = personal, 1 = surrogate). "Appropriateness" represents the difference between ratings of appropriateness for anti-charity and pro-charity contracts ($\text{appropriateness}_{\textit{anti-charity}} - \text{appropriateness}_{\textit{pro-charity}}$). "Effectiveness" represents the difference between ratings of effectiveness for anti-charity and pro-charity contracts ($\text{effectiveness}_{\textit{anti-charity}} - \text{effectiveness}_{\textit{pro-charity}}$). Models 2-3 include interaction terms between surrogate choice and differences scores. For Models 2-4, we decompose the interaction terms by reporting the average marginal effect (i.e., predicted probabilities) for each difference score by condition. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 4.5. *Study 4 mediation results (logistic regression)*

	(1)	(2)	(3)	(4)
Surrogate Choice	0.443*** (0.128)	0.455*** (0.129)	0.191 (0.153)	0.200 (0.155)
Appropriateness		0.073** (0.023)		0.119*** (0.032)
Effectiveness			0.676*** (0.056)	0.700*** (0.058)
Intercept	–0.459*** (0.092)	–0.336*** (0.100)	–0.995*** (0.124)	–0.818*** (0.131)
Pseudo R^2	.016	.030	.405	.422
Indirect Effects				
Appropriateness		–0.011 [–0.027, 0.010]		–0.005 [–0.014, 0.004]
Effectiveness			0.074 [0.042, 0.110]	0.069 [0.028, 0.103]
Model Adjustment				
Confounding Ratio		0.98	3.26	3.18
Confounding Percentage		–1.53	69.29	68.57
Rescaling Factor		1.01	1.41	1.44

Note: Each column reports the results of a logistic model in which the dependent variable is a dichotomous choice between a pro- or anti-charity commitment contract (0 = pro-charity contract, 1 = anti-charity contract), chosen for the self or other (personal = 0, surrogate = 1). Column 1 shows the basic surrogate choice effect of contract preference when selecting a contract for the self or for another person. Column 2 shows the results of adding differences in appropriateness indices between pro- and anti-charity contracts as a mediator of the observed difference of contract choice by self-other condition. Column 3 shows the results of adding differences in effectiveness indices between pro- and anti-charity contracts as a mediator of the observed difference of contract choice by self-other condition. Column 4 simultaneously adds both differences in appropriateness and effective indices of pro- and anti-charity contracts as parallel mediators of the difference in self-other contract preference. The difference in contract choice is fully mediated, suggesting differences in how effective people perceive contracts to be for the self versus for others is driving the choice of commitment contract. For all columns standard errors are reported in parentheses and bootstrapped 95% bias corrected and accelerated confidence intervals are reported in brackets. We follow the recommendations of Karlson, Holm, and Breen (2012) and report the confounding ratio (the total effect divided by the direct effect), the confounding percentage (the percent of the total effect explained by the mediator(s)), and rescaling factor (change in the scale parameter when mediator(s) are included) for the models that use the KHB-adjustment (Columns 2–4). Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

4.4 Study 4: Blinder-Oaxaca Decomposition

In this section, we describe the results from a Kitagawa-Blinder-Oaxaca decomposition to simultaneously test the decision-weighting and belief-based accounts proposed in Study 4 (Blinder, 1973; Kitagawa, 1955; Oaxaca, 1973). Specifically, the decomposition quantifies how much of the self-other difference in contract choice is due to (i) mean differences in effectiveness/appropriateness ratings across self-other conditions (often referred to as *model endowments*; this portion quantifies the amount of the self-other gap that is mediated by effectiveness/appropriateness beliefs, and thus consistent with a belief-based account); (ii) differences in the predictiveness of effectiveness/appropriateness on contract choice across self-other conditions (often referred to as *model coefficients*; this portion quantifies the amount of the self-other gap that is moderated by effectiveness/appropriateness beliefs, and thus consistent with a decision-weighting account), and (iii) all other differences not explained.

We performed a two-way OLS decomposition with pooled estimates across conditions with bootstrapped standard errors based on 10,000 resamples (Hlavac, 2022; Jann, 2008).[1] Table 4.6 displays the results of the decomposition. We find that of the 10.9 percentage point self-other difference in contract selection rates, 6.5 percentage points can be attributed to differences in model endowments (i.e., differences in beliefs about effectiveness and appropriateness across conditions), 2.6 percentage points to differences in model coefficients (i.e., how effectiveness/appropriateness are weighted differently across conditions), and 1.8 percentage points to the model intercept (i.e., unrelated to effectiveness/appropriateness ratings). Thus, the results suggest that about 60% of the self-other difference is explained by differences in effectiveness/appropriateness beliefs across conditions.

[1] We find virtually identical results when using logistic regression instead of OLS for the decomposition.

Looking further into the component parts of the model endowments and coefficients, mean differences in effectiveness beliefs explain 6.7/10.9 = 61.5% of the overall self-other difference in contract preferences whereas mean differences in appropriateness ratings explain –0.1/10.9 = –0.9% of the overall difference in contract preferences. We also find that differences in effectiveness and appropriateness coefficients across conditions account for 11% and 12% of the overall effect in contract preferences, respectively. Similar to the results reported in the main text, only mean differences in effectiveness ratings across conditions—consistent with a belief-based account—are statistically significant at $p < .05$ and explain the majority of the self-other gap in contract choice.

Table 4.6. *Study 4 results of Kitawaga-Blinder-Oaxaca decomposition*

	b	*SE*	*z*	*p*
Overall difference	0.109	0.031	3.48	< 0.001
Endowments	0.066	0.018	3.79	< 0.001
Coefficients	0.026	0.020	1.64	0.183
Intercept	0.018	0.033	0.54	0.590
Endowments				
Appropriateness	–0.001	0.003	0.49	0.623
Effectiveness	0.067	0.017	3.94	< 0.001
Coefficients				
Appropriateness	0.013	0.018	0.75	0.455
Effectiveness	0.012	0.006	1.73	0.084

Note: Results from a Kitawaga-Blinder-Oaxaca decomposition using OLS regression with pooled estimates and bootstrapped standard errors based on 10,000 resamples. The dependent variable is choice of commitment contract (0 = pro-charity, 1 = anti-charity). The difference between the percentages is decomposed into differences due to model endowments (i.e., differences in beliefs about the appropriateness/effectiveness of the contracts across conditions), differences due to differences in model coefficients (i.e., how appropriateness/effectiveness are weighted differently across conditions), and differences unrelated to appropriateness and effectiveness (captured by the model intercept).

5.1 Study 5: Moderation and Mediation Analysis

Table 5.1. *Study 5 moderation results*

	(1)	(2)	(3)	(4)
Other Choice	0.122*** (0.032)	0.151*** (0.040)	0.033 (0.026)	0.056 (0.041)
Appropriateness		0.052*** (0.006)		0.022*** (0.005)
Effectiveness			0.111*** (0.004)	0.106*** (0.004)
Other Choice x Appropriateness		0.007 (0.011)		0.005 (0.010)
Other Choice x Effectiveness			0.014 (0.007)	0.013 (0.008)
Intercept	0.457*** (0.018)	0.589*** (0.023)	0.336*** (0.014)	0.398*** (0.020)
Adjusted R^2	.012	.094	.440	.454
Average Marginal Effects				
Self: $b_{\text{appropriateness}}$		0.052*** (0.006)		0.022*** (0.005)
Other: $b_{\text{appropriateness}}$		0.059*** (0.010)		0.027*** (0.008)
Self: $b_{\text{effectiveness}}$			0.111*** (0.005)	0.106*** (0.005)
Other: $b_{\text{effectiveness}}$			0.125*** (0.007)	0.119*** (0.007)

Note: Each model reports the results of a linear probability model, with robust standard errors in parentheses. The dependent variable in all models is choice of commitment contract (0 = pro-charity, 1 = anti-charity). "Surrogate choice" indicates whether participants were assigned to choose a commitment contract for themselves or for another person (0 = personal/surrogate-friend, 1 = surrogate-other). "Appropriateness" represents the difference between ratings of appropriateness for anti-charity and pro-charity contracts ($\text{appropriateness}_{\textit{anti-charity}} - \text{appropriateness}_{\textit{pro-charity}}$). "Effectiveness" represents the difference between ratings of effectiveness for anti-charity and pro-charity contracts ($\text{effectiveness}_{\textit{anti-charity}} - \text{effectiveness}_{\textit{pro-charity}}$). Models 2–3 include interaction terms between surrogate choice and differences scores. For models 2–4, we decompose the interaction terms by reporting the average marginal effects (i.e., simple slopes) for each difference score by condition. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 5.2. *Study 5 mediation results*

	(1)	(2)	(3)	(4)
Other Choice	0.122*** (0.032)	0.133*** (0.030)	0.053* (0.024)	0.062** (0.024)
Appropriateness		0.054*** (0.005)		0.023*** (0.004)
Effectiveness			0.115*** (0.003)	0.109*** (0.004)
Intercept	0.457*** (0.018)	0.594*** (0.022)	0.332*** (0.014)	0.398*** (0.019)
Adjusted R^2	.012	.095	.439	.453
Indirect Effects				
Appropriateness		–0.011 [-0.029, 0.007]		–0.005 [-0.014, 0.002]
Effectiveness			0.069 [0.030, 0.107]	0.065 [0.028, 0.102]

Note: Each model reports the results of a linear probability model, with robust standard errors in parentheses. The dependent variable in all models is choice of commitment contract (0 = pro-charity, 1 = anti-charity). "Surrogate choice" indicates whether participants were assigned to choose a commitment contract for themselves or for another person (0 = personal/surrogate-friend, 1 = surrogate-other). "Appropriateness" represents the difference between ratings of appropriateness for anti-charity and pro-charity contracts ($\text{appropriate}_{anti\text{-}charity} - \text{appropriate}_{pro\text{-}charity}$). "Effectiveness" represents the difference between ratings of effectiveness for anti-charity and pro-charity contracts ($\text{effective}_{anti\text{-}charity} - \text{effective}_{pro\text{-}charity}$). Models 2–4 also display indirect effects, with bootstrapped bias corrected and accelerated confidence intervals (using 10,000 resamples) in brackets. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

5.2 Study 5: Regression Model with Separate Self and Friend Conditions

Table 5.3. *Study 5 mediation results with self and friend conditions separated*

	(1)	(2)	(3)	(4)
Friend-Choice	–0.007 (0.037)	–0.012 (0.035)	–0.053 (0.028)	–0.053 (0.027)
Other-Choice	0.119** (0.037)	0.127*** (0.035)	0.027 (0.027)	0.035 (0.027)
Appropriateness		0.054*** (0.004)		0.023*** (0.004)
Effectiveness			0.116*** (0.003)	0.110*** (0.004)
Intercept	0.460*** (0.026)	0.600*** (0.028)	0.358*** (0.019)	0.424*** (0.023)
Adjusted R^2	.012	.094	.440	.455
Indirect Effects				
Appropriateness$_{Friend}$		0.005 [–0.016, 0.027]		0.002 [–0.006, 0.014]
Appropriateness$_{Other}$		–0.009 [–0.030, 0.012]		–0.004 [–0.015, 0.007]
Effectiveness$_{Friend}$			0.046 [–0.004, 0.095]	0.044 [–0.010, 0.087]
Effectiveness$_{Other}$			0.092 [0.044, 0.138]	0.087 [0.024, 0.140]

Note: Each model reports the results of a linear probability model, with robust standard errors in parentheses. The dependent variable in all models is choice of commitment contract (0 = pro-charity, 1 = anti-charity). "Friend-choice" indicates whether participants were assigned to choose a commitment contract for a close friend (0 = no, 1 = yes). "Other-choice" indicates whether participants were assigned to choose a commitment contract for a stranger (0 = no, 1 = yes). The personal choice condition served as the reference/baseline value. "Appropriateness" represents the difference between ratings of appropriateness for anti-charity and pro-charity contracts (appropriateness$_{anti\text{-}charity}$ − appropriateness$_{pro\text{-}charity}$). "Effectiveness" represents the difference between ratings of effectiveness for anti-charity and pro-charity contracts (effectiveness$_{anti\text{-}charity}$ – effectiveness$_{pro\text{-}charity}$). Models 2–4 also display indirect effects, with bootstrapped bias corrected and accelerated confidence intervals (using 10,000 resamples) in brackets. Significance levels: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

5.3 Study 5: Blinder-Oaxaca Decomposition

In this section, we describe the results from a Kitagawa-Blinder-Oaxaca decomposition to simultaneously test the decision-weighting and belief-based accounts proposed in Study 4 (Blinder, 1973; Kitagawa, 1955; Oaxaca, 1973). Specifically, we use a two-way OLS decomposition with pooled estimates across conditions and bootstrapped standard errors based on 10,000 resamples.[2] Table 5.4 shows the results of the decomposition. We find that of the 12.2 percentage point self-other difference in contract selection rates, 6.1 percentage points can be attributed to differences in endowments (i.e., differences in beliefs about effectiveness and appropriateness across conditions), 0.4 percentage points to differences in coefficients (i.e., how effectiveness/ appropriateness are weighted differently across conditions), and 5.6 percentage points to the intercept (i.e., unrelated to effectiveness/appropriateness ratings). Thus, the results suggest that the largest share of the self-other difference is explained by differences in effectiveness/appropriateness beliefs across conditions.

Looking further into the component parts of the endowments and coefficients, mean differences in effectiveness beliefs explain 6.6/12.2 = 54.1% of the overall self-other difference in contract preferences whereas mean differences appropriateness ratings explain –0.5/12.2 = –4.1% of the effect. We also find that differences in effectiveness and appropriateness coefficients across conditions account for 15.6% and –11.5% of the overall effect, respectively. Similar to the results reported in the main text, only mean differences in effectiveness ratings across conditions—consistent with a belief-based account—are statistically significant at $p < .05$ and explain the majority of the self-other gap in contract choice.

[2] Again, we find virtually identical results when using logistic regression instead of OLS to perform the decomposition.

Table 5.4. *Study 5 results of Kitawaga-Blinder-Oaxaca decomposition*

	b	*SE*	*z*	*p*
Overall difference	0.122	0.031	3.87	< .001
Endowments	0.061	0.020	3.00	.003
Coefficients	0.004	0.030	0.15	.882
Intercept	0.056	0.041	1.38	.168
Endowments				
Appropriateness	–0.005	0.004	1.22	.222
Effectiveness	0.066	0.019	3.45	< .001
Coefficients				
Appropriateness	–0.014	0.028	0.52	.600
Effectiveness	0.019	0.012	1.51	.132

Note: Results from a Kitawaga-Blinder-Oaxaca decomposition using OLS regression with pooled estimates and bootstrapped standard errors based on 10,000 resamples. The dependent variable is choice of commitment contract (0 = pro-charity, 1 = anti-charity). The difference between the percentages is decomposed into differences due to model endowments (i.e., differences in beliefs about the appropriateness/effectiveness of the contracts across conditions), differences due to differences in model coefficients (i.e., how appropriateness/effectiveness are weighted differently across conditions), and differences unrelated to appropriateness and effectiveness (captured by the model intercept).

6 Supplemental Study 1: stickK Contract Analysis

We analyzed 17,654 commitment contracts set by 9,046 users on stickK.com from Decembers 28, 2020 to April 29, 2021. When users set a commitment contract on stickK they choose between five different recipient options (Anti-Charity, Charity, Friend or Foe, stickK, and No Stakes). Upon selecting any option besides "No Stakes," the user is required to place money on the line for non-compliance. Users who chose "Anti-Charity" were prompted to also select a specific organization they opposed from a pre-determined list. The most popular option was "No Stakes" (66%), indicating that only approximately one-third of participants selected contracts with potential monetary penalties. Table 6.1 provides descriptive statistics on the number of contracts set by option.

When creating a commitment contract, users had the option to report their progress daily, weekly, or at the end of the contract. At end of each reporting period, users received an email asking whether they had kept their commitment (yes or no). We find that users who set a contract requiring monetary stakes were 65 percentage points more likely to report successfully keeping their commitment (83%) than users who selected "No Stakes" (18%; $z = 330.61$, $p < .001$). Furthermore, users who set anti-charity contracts were more than 4 percentage points more likely to report keeping their commitment (84%) than users who selected any of the other contract options (80%; $z = 13.57$, $p < .001$).

Table 6.1. *Contracts set on stickK.com*

	No Stakes	Anti-Charity	Charity	Friend or Foe	stickK
Number of Contracts	11,645 (66.09%)	2,978 (16.90%)	1,154 (6.55%)	1,243 (7.05%)	601 (3.41%)
Number of Users	6,747 (74.58%)	1,318 (14.57%)	609 (6.73%)	721 (7.97%)	207 (2.28%)

Note: The first row contains descriptive statistics on the number of contracts set. The second row contains descriptive statistics of contract selection by users. Users are allowed to set more than one contract at a time. For the second row, the total percentage exceeds 100 because some users selected more than one contract.

References

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, *8*(4), 436–455.

Hlavac, Marek (2022). *oaxaca: Blinder-Oaxaca decomposition in R* (Version 0.1.5.) [R Package] https://CRAN.R-project.org/package=oaxaca.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, *8*(4), 453-479.

Karlson, K. B., Holm, A., & Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, *42*(1), 286-313.

Kitagawa, E. M. (1955). Components of a difference between two rates. *Journal of the American Statistical Association*, *50*(272), 1168–1194.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, *14*(3), 693–709.