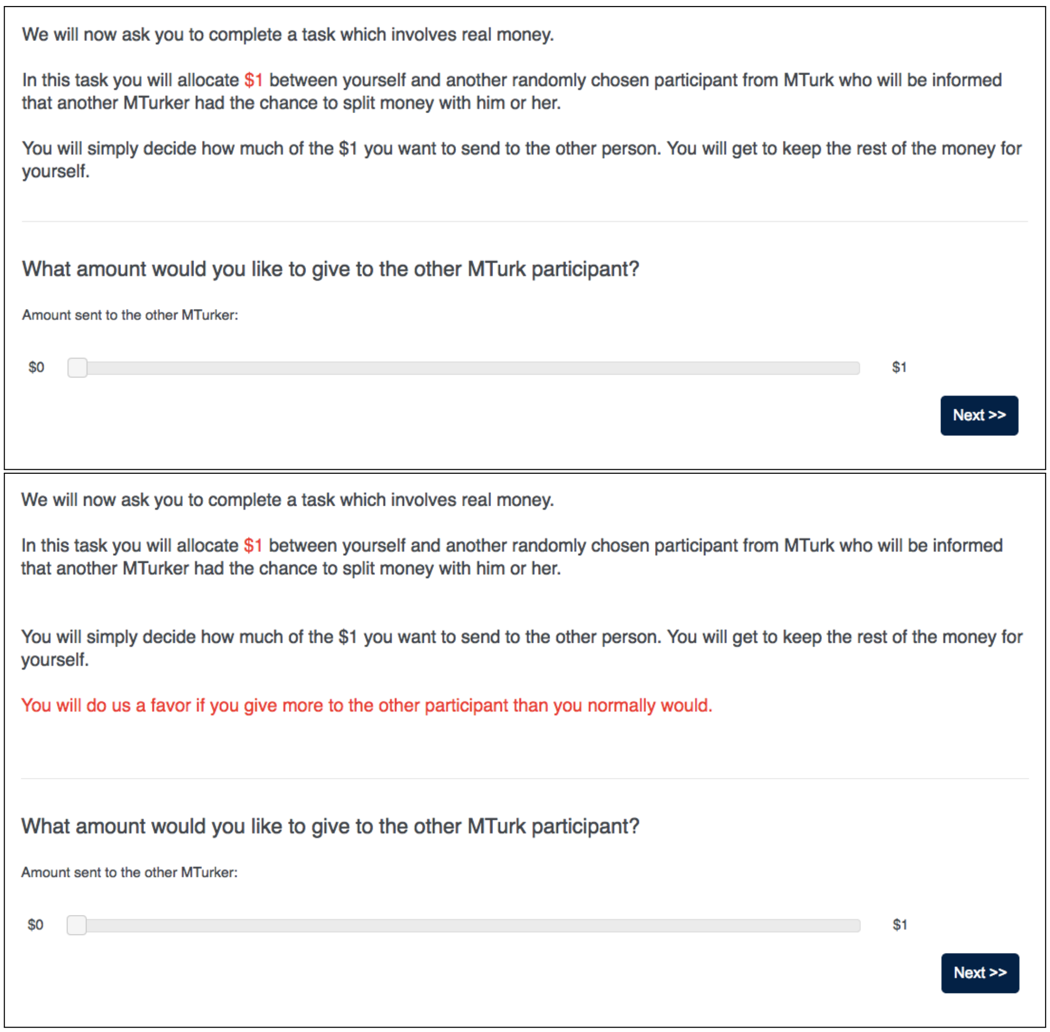
Measuring and Bounding Experimenter Demand

Jonathan de Quidt Johannes Haushofer Christopher Roth

Online Appendix

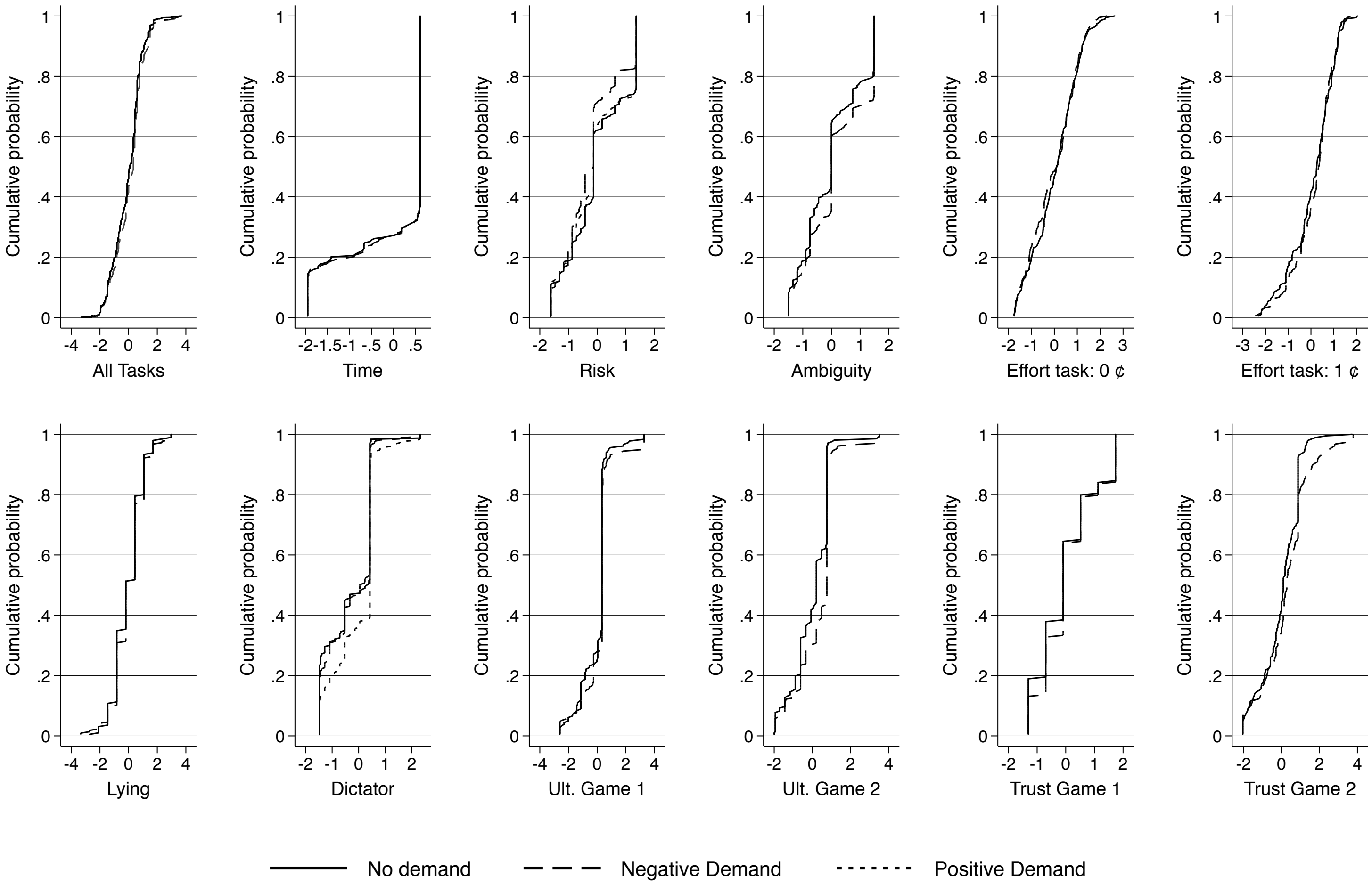
A. Additional Figures and Tables

Figure A1. Interface example



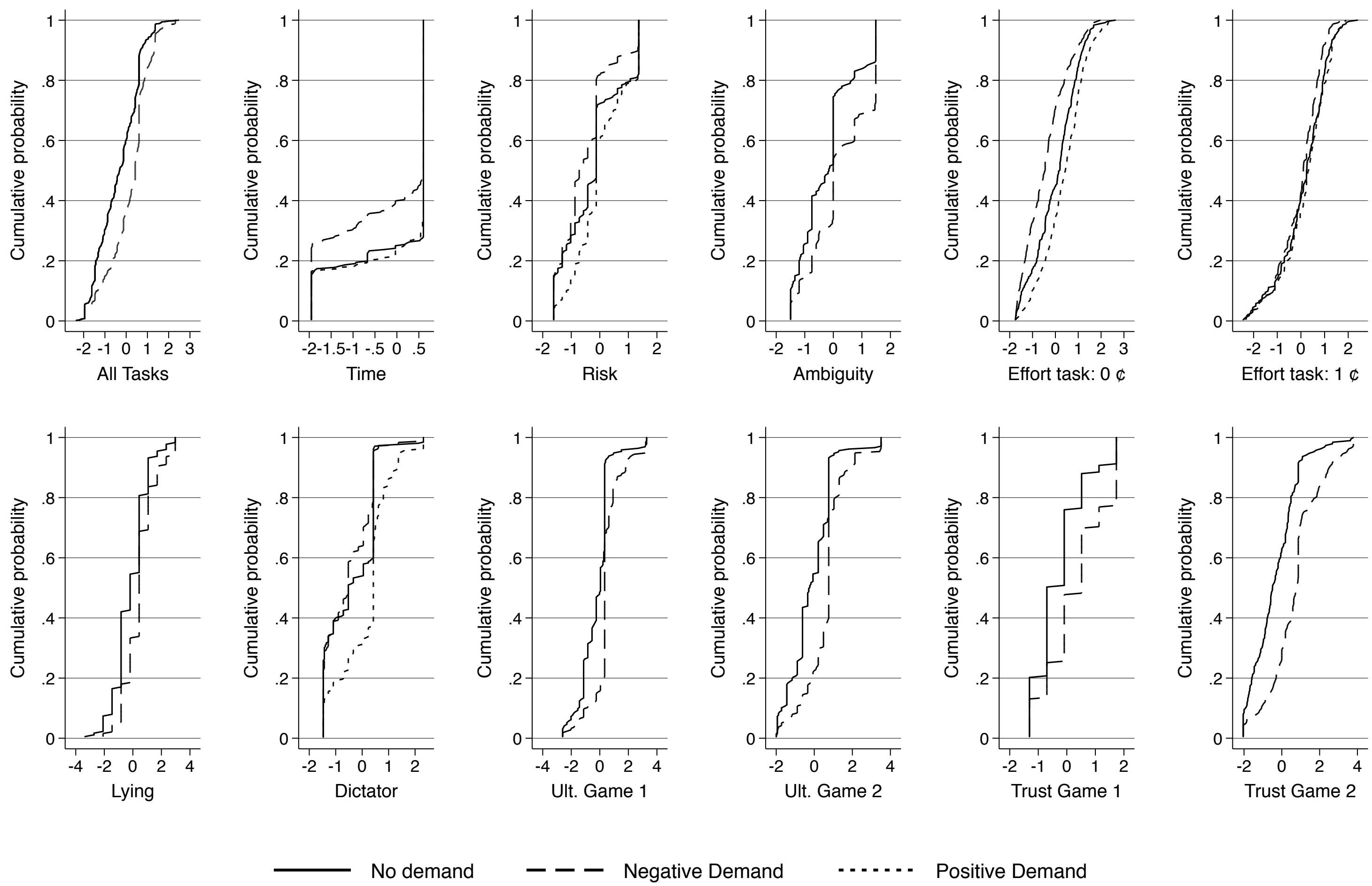
Note: We present two examples of the experimental interface, taken from the dictator game. The first frame corresponds to the real stakes, "no demand" condition, and the second frame to the real stakes, positive demand condition.

Figure A2. Distribution of z-scored actions by task and demand treatment, weak treatments



Note: This figure uses data from incentivized MTurk respondents with weak demand treatments, and displays the cumulative distribution function of z-scored actions by task and demand treatment arm.

Figure A3. Distribution of z-scored actions by task and demand treatment, strong treatments



Note: This figure uses data from incentivized MTurk respondents with strong demand treatments, and displays the cumulative distribution function of z-scored actions by task and demand treatment arm.

Table A1—Controlling for demand

	Conventional	Sensitivity difference	Strong	Weak	
	Treatment Effect	(Strong 1-cent - Strong 0-cent)	Midpoint	positive-positive	negative-negative
Count	540.720	430.009	494.774	588.270	530.001
	(66.763)	(89.477)	(49.321)	(61.499)	(64.532)
Count (z-scored)	0.686	0.546	0.628	0.747	0.673
	(0.085)	(0.114)	(0.063)	(0.078)	(0.082)

Note: This table uses data from the real effort experiments (experiment 3 and experiment 6). We follow the "controlling for demand" procedure outlined in Section III.D to estimate the treatment effect of incentives on effort provision. Column (1) shows the conventional treatment effect estimate (data from experiment 3). Column (2) tests for differences in sensitivity to our strong demand treatments between the 0-cent and 1-cent groups, and finds a significant difference. Therefore in column (3) we apply the "midpoint" technique with strong demand treatments to estimate the treatment effect. Columns (4) and (5) approximate the treatment effect using same-signed weak demand treatments. We apply the "ironing" procedure described in section III.B when constructing these estimates. Count is the raw-score of points scored in the real effort task. Count (z-scored) uses the mean and standard deviation from the negative demand condition. Robust standard errors in parentheses. Note that strong and weak treatment data were collected in separate experiments.

Table A2—Results from the Within Design

	Dictator			Risk		
	Within	Between	Difference	Within	Between	Difference
Panel A: Unconditional Means						
Positive demand	0.384	0.434	-0.050	0.560	0.550	0.010
	(0.017)	(0.015)	(0.023)	(0.021)	(0.020)	(0.029)
No demand	0.273	0.282	-0.010	0.448	0.466	-0.018
	(0.011)	(0.015)	(0.019)	(0.015)	(0.022)	(0.027)
Negative demand	0.195	0.251	-0.056	0.318	0.373	-0.055
	(0.014)	(0.014)	(0.020)	(0.019)	(0.019)	(0.027)
Panel B: Sensitivity (positive - negative)						
Raw data	0.189	0.183	0.006	0.242	0.177	0.065
	(0.022)	(0.021)	(0.031)	(0.029)	(0.027)	(0.040)
Z-score	0.794	0.745	0.048	0.709	0.520	0.188
	(0.093)	(0.086)	(0.127)	(0.084)	(0.080)	(0.116)
Panel C: Monotonicity						
Positive - Neutral (z-score)	0.514	0.617	-0.103	0.377	0.248	0.129
	(0.044)	(0.088)	(0.129)	(0.041)	(0.087)	(0.124)
Negative - Neutral (z-score)	-0.380	-0.128	-0.251	-0.427	-0.272	-0.155
	(0.045)	(0.086)	(0.123)	(0.042)	(0.084)	(0.119)
Observations	499	770	1269	500	728	1228

Note: This table uses data from the within design (experiment 7) and incentivized choices from the dictator game and the investment game in experiment 1. These experiments employ strong demand treatments. Panel A displays the unconditional means by task and demand treatment arm. Panel B displays the estimates of sensitivity. Panel C tests Monotonicity. Note that estimates from Panel C do not add up to the sensitivity estimates from Panel B as sensitivity is estimated between participants while monotonicity tests are within-participant.

Table A3—Confidence intervals for bounds on natural actions

	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Weak Demand											
Interval	[0.766, 0.770]	[0.472, 0.524]	[0.499, 0.557]	[0.343, 0.331]	[0.469, 0.484]	[0.530, 0.537]	[0.318, 0.382]	[0.443, 0.470]	[0.362, 0.413]	[0.430, 0.455]	[0.348, 0.398]
95% CI on interval	[0.713, 0.822]	[0.431, 0.569]	[0.452, 0.605]	[0.316, 0.358]	[0.444, 0.507]	[0.508, 0.559]	[0.291, 0.409]	[0.418, 0.497]	[0.338, 0.440]	[0.380, 0.500]	[0.324, 0.432]
95% CI on parameter	[0.716, 0.821]	[0.438, 0.561]	[0.459, 0.597]	[0.316, 0.358]	[0.448, 0.504]	[0.510, 0.557]	[0.296, 0.405]	[0.422, 0.493]	[0.342, 0.436]	[0.387, 0.494]	[0.328, 0.426]
Observations	422	739	390	388	381	412	758	360	411	352	346
Panel B: Strong Demand											
Interval	[0.659, 0.795]	[0.373, 0.550]	[0.428, 0.583]	[0.255, 0.405]	[0.449, 0.492]	[0.510, 0.606]	[0.251, 0.434]	[0.404, 0.520]	[0.337, 0.474]	[0.350, 0.535]	[0.288, 0.469]
95% CI on interval	[0.603, 0.842]	[0.336, 0.589]	[0.384, 0.629]	[0.233, 0.427]	[0.428, 0.515]	[0.483, 0.630]	[0.222, 0.464]	[0.377, 0.545]	[0.309, 0.501]	[0.308, 0.581]	[0.258, 0.503]
95% CI on parameter	[0.612, 0.834]	[0.342, 0.583]	[0.391, 0.622]	[0.236, 0.424]	[0.432, 0.511]	[0.487, 0.626]	[0.227, 0.459]	[0.381, 0.541]	[0.314, 0.496]	[0.314, 0.574]	[0.263, 0.498]
Observations	727	728	404	731	714	365	770	409	421	382	371

Note: This table uses data from incentivized MTurk respondents with strong and weak demand treatments. It first presents estimated bounds on the natural action, then 95 percent confidence intervals on those bounds, then 95 percent confidence intervals on the parameter (natural action) contained in the bounds.

Table A4—Confidence intervals for bounds on treatment effects

	Treatment Effect: Score in Effort Task
Weak treatments	
Interval	[530.001, 588.270]
95% CI on interval	[403.430, 708.890]
95% CI on parameter	[410.310, 701.645]
Observations	769
Strong treatments	
Interval	[177.421, 948.978]
95% CI on interval	[55.158, 1074.708]
95% CI on parameter	[74.817, 1054.492]
Observations	1445

Note: This table uses data from incentivized MTurk respondents with weak and strong demand treatments (experiments 3 and 6). It first presents estimated bounds on the treatment effect of incentives on effort, then 95 percent confidence intervals on those bounds, then 95 percent confidence intervals on the parameter (treatment effect) contained in the bounds.

Table A5—Results from the Within Design: Compliers and Defiers

	Dictator			Risk		
	All	Compliers	Defiers	All	Compliers	Defiers
Positive - Neutral (z-score)	0.514	0.777	-0.402	0.377	0.704	-0.601
	(0.044)	(0.055)	(0.122)	(0.041)	(0.052)	(0.100)
Observations	265	179	7	247	146	16
Negative - Neutral (z-score)	-0.380	-0.796	1.028	-0.427	-0.721	0.529
	(0.045)	(0.059)	(0.329)	(0.042)	(0.049)	(0.199)
Observations	234	122	8	253	161	16

Note: This table uses data from the within design (experiment 7). The outcome variable is the change in standardized action between task 1 and task 2. We separately present the results for the whole sample, compliers, and defiers.

Table A6—Within Design: Defier-corrected bounds and confidence intervals

	Risk	Dictator
Panel A: Standard Bounds		
Interval	[0.318, 0.560]	[0.195, 0.384]
95% CI on interval	[0.280, 0.602]	[0.167, 0.418]
95% CI on parameter	[0.286, 0.595]	[0.172, 0.412]
Observations	500	499
Panel B: Adjusted Bounds		
Interval	[0.308, 0.571]	[0.185, 0.392]
95% CI on interval	[0.271, 0.613]	[0.158, 0.425]
95% CI on parameter	[0.277, 0.606]	[0.163, 0.420]
Observations	500	499

Note: This table uses data from the within design (experiment 7). In Panel A we compute our standard bounds and confidence intervals. In Panel B we compute the adjusted bounds which take into account defier behavior.

Table A7—Belief about the experimental objective in response to the weak demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.832	0.757	0.760	0.788	0.979	0.779	0.540	0.698	0.688	0.612	0.669
	(0.026)	(0.028)	(0.031)	(0.030)	(0.010)	(0.028)	(0.032)	(0.034)	(0.032)	(0.036)	(0.038)
No demand		0.620					0.321				
		(0.030)					(0.030)				
Negative demand	0.603	0.370	0.330	0.277	0.358	0.467	0.234	0.238	0.383	0.112	0.083
	(0.034)	(0.031)	(0.034)	(0.032)	(0.035)	(0.036)	(0.026)	(0.032)	(0.034)	(0.024)	(0.020)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.229	0.388	0.430	0.511	0.621	0.312	0.306	0.461	0.304	0.500	0.585
	(0.042)	(0.042)	(0.046)	(0.044)	(0.036)	(0.046)	(0.041)	(0.047)	(0.047)	(0.044)	(0.043)
Z-score	0.471	0.776	0.909	1.117	1.240	0.627	0.678	0.994	0.633	1.092	1.417
	(0.087)	(0.084)	(0.096)	(0.095)	(0.073)	(0.092)	(0.091)	(0.101)	(0.098)	(0.095)	(0.104)
		[0.001]					[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)		0.274					0.485				
		(0.082)					(0.096)				
		[0.001]					[0.001]				
Negative - Neutral (z-score)		-0.501					-0.193				
		(0.087)					(0.088)				
		[0.001]					[0.009]				
Observations	422	739	390	388	381	412	758	360	411	352	346

Note: This table uses data from incentivized MTurk respondents with weak demand treatments. The outcome variables take value one if the respondents believed that the experimenter wanted a high action. Panel A displays mean beliefs with standard errors in the positive, negative and no-demand conditions respectively. Panel B presents the raw and z-scored sensitivity of beliefs to our demand treatments. Panel C displays the response to our positive and negative demand treatments separately, when "no demand" choices were also collected. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets, adjusting across tests within each task.

Table A8—Belief about the experimental objective in response to the strong demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.802	0.705	0.701	0.776	0.942	0.815	0.651	0.572	0.664	0.407	0.385
	(0.025)	(0.030)	(0.032)	(0.027)	(0.015)	(0.028)	(0.029)	(0.034)	(0.032)	(0.035)	(0.036)
No demand	0.720	0.537		0.485	0.888		0.355				
	(0.029)	(0.032)		(0.032)	(0.020)		(0.030)				
Negative demand	0.622	0.424	0.335	0.296	0.511	0.562	0.244	0.309	0.357	0.295	0.217
	(0.031)	(0.031)	(0.033)	(0.029)	(0.033)	(0.037)	(0.028)	(0.033)	(0.033)	(0.034)	(0.030)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.181	0.281	0.366	0.480	0.431	0.252	0.407	0.263	0.306	0.112	0.168
	(0.040)	(0.043)	(0.046)	(0.039)	(0.036)	(0.047)	(0.040)	(0.047)	(0.047)	(0.049)	(0.047)
Z-score	0.372	0.563	0.773	1.050	0.861	0.507	0.901	0.567	0.637	0.245	0.406
	(0.083)	(0.087)	(0.098)	(0.086)	(0.073)	(0.094)	(0.089)	(0.102)	(0.097)	(0.106)	(0.114)
	[0.001]	[0.001]		[0.001]	[0.001]		[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)	0.169	0.337		0.635	0.108		0.654				
	(0.079)	(0.088)		(0.092)	(0.050)		(0.092)				
	[0.023]	[0.001]		[0.001]	[0.011]		[0.001]				
Negative - Neutral (z-score)	-0.203	-0.226		-0.415	-0.754		-0.247				
	(0.089)	(0.089)		(0.095)	(0.077)		(0.090)				
	[0.022]	[0.003]		[0.001]	[0.001]		[0.002]				
Observations	727	728	404	731	714	365	770	409	421	382	371

Note: This table uses data from incentivized MTurk respondents with strong demand treatments. The outcome variables take value one if the respondents believed that the experimenter wanted a high action. Panel A displays mean beliefs with standard errors in the positive, negative and no-demand conditions respectively. Panel B presents the raw and z-scored sensitivity of beliefs to our demand treatments. Panel C displays the response to our positive and negative demand treatments separately, when "no demand" choices were also collected. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets, adjusting across tests within each task.

Table A9—Belief about the experimental hypothesis in response to the weak demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.788	0.749	0.704	0.782	0.963	0.871	0.464	0.726	0.732	0.628	0.682
	(0.028)	(0.028)	(0.033)	(0.030)	(0.014)	(0.023)	(0.032)	(0.033)	(0.031)	(0.036)	(0.038)
No demand		0.534					0.160				
		(0.031)					(0.024)				
Negative demand	0.458	0.261	0.222	0.231	0.326	0.528	0.106	0.354	0.359	0.296	0.182
	(0.034)	(0.029)	(0.030)	(0.030)	(0.034)	(0.036)	(0.019)	(0.036)	(0.034)	(0.035)	(0.028)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.331	0.488	0.482	0.552	0.637	0.343	0.358	0.373	0.372	0.333	0.500
	(0.044)	(0.040)	(0.044)	(0.042)	(0.037)	(0.042)	(0.037)	(0.049)	(0.046)	(0.050)	(0.047)
Z-score	0.681	0.978	0.982	1.244	1.286	0.706	0.836	0.825	0.750	0.731	1.161
	(0.092)	(0.080)	(0.090)	(0.096)	(0.074)	(0.087)	(0.086)	(0.108)	(0.092)	(0.110)	(0.109)
		[0.001]					[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)		0.431					0.708				
		(0.084)					(0.092)				
		[0.001]					[0.001]				
Negative - Neutral (z-score)		-0.547					-0.128				
		(0.084)					(0.071)				
		[0.001]					[0.024]				
Observations	422	739	390	388	381	412	758	360	411	352	346

Note: This table uses data from incentivized MTurk respondents with weak demand treatments. The outcome variables take value one if the respondents believed that the experimenter expected a high action. Panel A displays mean beliefs with standard errors in the positive, negative and no-demand conditions respectively. Panel B presents the raw and z-scored sensitivity of beliefs to our demand treatments. Panel C displays the response to our positive and negative demand treatments separately, when "no demand" choices were also collected. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets, adjusting across tests within each task.

Table A10—Belief about the experimental hypothesis in response to the strong demand treatments

	Belief: Time	Belief: Risk	Belief: Ambiguity Aversion	Belief: Effort 0 cent bonus	Belief: Effort 1 cent bonus	Belief: Lying	Belief: Dictator Game	Belief: Ult. Game 1	Belief: Ult. Game 2	Belief: Trust Game 1	Belief: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.727	0.595	0.593	0.727	0.934	0.788	0.454	0.670	0.659	0.578	0.588
	(0.028)	(0.033)	(0.034)	(0.029)	(0.016)	(0.030)	(0.030)	(0.032)	(0.032)	(0.035)	(0.037)
No demand	0.682	0.484		0.423	0.855		0.143				
	(0.030)	(0.032)		(0.032)	(0.023)		(0.022)				
Negative demand	0.639	0.420	0.400	0.267	0.576	0.625	0.186	0.284	0.435	0.290	0.243
	(0.031)	(0.031)	(0.035)	(0.028)	(0.033)	(0.037)	(0.025)	(0.032)	(0.035)	(0.034)	(0.031)
Panel B: Sensitivity (Positive - Negative)											
Raw data	0.089	0.175	0.193	0.459	0.358	0.163	0.268	0.386	0.224	0.288	0.345
	(0.042)	(0.045)	(0.049)	(0.040)	(0.036)	(0.047)	(0.039)	(0.046)	(0.047)	(0.049)	(0.048)
Z-score	0.183	0.351	0.393	1.036	0.722	0.336	0.624	0.855	0.451	0.634	0.801
	(0.087)	(0.090)	(0.100)	(0.091)	(0.074)	(0.097)	(0.092)	(0.101)	(0.095)	(0.107)	(0.112)
	[0.118]	[0.001]		[0.001]	[0.001]		[0.001]				
Panel C: Monotonicity											
Positive - Neutral (z-score)	0.093	0.222		0.685	0.159		0.725				
	(0.085)	(0.091)		(0.097)	(0.056)		(0.087)				
	[0.268]	[0.015]		[0.001]	[0.001]		[0.001]				
Negative - Neutral (z-score)	-0.090	-0.128		-0.350	-0.563		0.101				
	(0.090)	(0.089)		(0.096)	(0.080)		(0.077)				
	[0.268]	[0.052]		[0.001]	[0.001]		[0.069]				
Observations	727	728	404	731	714	365	770	409	421	382	371

Note: This table uses data from incentivized MTurk respondents with strong demand treatments. The outcome variables take value one if the respondents believed that the experimenter expected a high action. Panel A displays mean beliefs with standard errors in the positive, negative and no-demand conditions respectively. Panel B presents the raw and z-scored sensitivity of beliefs to our demand treatments. Panel C displays the response to our positive and negative demand treatments separately, when "no demand" choices were also collected. Robust standard errors are in parentheses. False-discovery rate adjusted p-values are in brackets, adjusting across tests within each task.

Table A11—Overview of studies varying anonymity in dictator games

Study	**Description of the treatment**	**Sample**	**Sample Size**	**Stake Size**	**Effect Size**	**Statistical significance**
Hoffman et al. (1994)	Double blind compared to single blind	Student sample from the University of Arizona	101	$10	61 percent reduction in giving	p<0.01
Hoffman, McCabe and Smith (1996)	Double blind compared to single blind	Student sample from the University of Arizona	114	$10	37 percent reduction in giving	p<0.01
Bolton, Katok and Zwick (1998)	Double blind compared to single blind	Student sample at Penn State University	60	$5	22 percent increase in giving	p>0.1
Barmettler, Fehr and Zehnder (2012)	Double blind compared to single blind	Student samples from the University of Zurich (UZH)	103	20 Swiss Frank ($22)	16.8 percent reduction in giving	p>0.1
Cilliers, Dube and Siddiqi (2015)	Presence of non-foreign experimenter vs.presence of a white foreign experimenter	Poor households from Sierra Leone	708	4000 Leones (approximately $1)	16 percent reduction in giving	p<0.01
Our estimates based on demand treatments						
de Quidt et al. (2018)	Weak negative demand treatment compared to weak positive demand treatment	MTurk respondents	515	$1	17 percent reduction in giving	p<0.01
de Quidt et al. (2018)	Strong negative demand treatment compared to strong positive demand treatment	MTurk respondents	511	$1	42 percent reduction in giving	p<0.01

Note: This table provides an overview of dictator game studies which vary the anonymity of experimenter-subject interactions and the presence of a foreign (white) experimenter. Our estimates of treatment effects for the studies by Hoffman et al. (1994) and Hoffman, McCabe and Smith (1996) are based on inspection of the cumulative distribution functions and and probability distribution functions reported in the paper (details of our calculations are available upon request). These papers did not report mean behavior across treatment arms. In Hoffman et al. (1994) we compare behavior in "Double Blind treatment 1" and "Double Blind treatment 2" to behavior in the "Dictator random entitlement, exchange". In Hoffman, McCabe and Smith (1996) we compare behavior in "Double Blind treatment 1" and "Double Blind treatment 2" to behavior in the "Single Blind 1" condition. In Bolton, Katok and Zwick (1998) we compare behavior in the "Anonymity" condition to behavior in the "6card1game" condition. In Barmettler, Fehr and Zehnder (2012) we compare behavior in the "Double Anonymity" condition to behavior in the "Single Anonymity" condition. In Cilliers, Dube and Siddiqi (2015) we compare behavior when a white foreigner was or was not present in the session. The average reduction in giving across the studies using equal weights is a 21.76 percent, or 20.37 percent when weighted by sample size.

Table A12—Overview of standard deviations across tasks

	Time	Risk	Ambiguity Aversion	Effort 0 cent bonus	Effort 1 cent bonus	Lying	Dictator Game	Ultimatum Game 1	Ultimatum Game 2	Trust Game 1	Trust Game 2
Panel A: Weak Demand											
Positive demand	0.385	0.348	0.341	0.193	0.165	0.170	0.222	0.189	0.194	0.314	0.217
No demand	.	0.339	.	.	.	.	0.234	.	.	.	.
Negative demand	0.389	0.317	0.334	0.190	0.178	0.158	0.226	0.170	0.182	0.329	0.172
Panel B: Strong Demand											
Positive demand	0.379	0.331	0.340	0.177	0.179	0.172	0.267	0.184	0.202	0.334	0.234
No demand	0.386	0.340	.	0.182	0.184	.	0.246	.	.	.	.
Negative demand	0.437	0.322	0.319	0.176	0.162	0.183	0.229	0.189	0.209	0.291	0.205

Note: This table uses data from incentivized MTurk respondents with weak and strong demand treatments and displays the standard deviations across the different demand treatment arms.

B. Theoretical Appendix

B1. Proof of Proposition 1 (Monotonicity)

We require that $a^+(\zeta) \geq a^L(\zeta) \geq a^-(\zeta)$. We are therefore interested in the sign of $\phi(E[h|h^T, h^L(\zeta)] - E[h|h^L(\zeta)])$. We have:

$$\begin{aligned}\phi(E[h|h^T, h^L(\zeta)] - E[h|h^L(\zeta)]) &= \phi\left(\frac{h^L(\zeta)p^L(\zeta) + h^T p^T}{1 + h^L(\zeta)p^L(\zeta)h^T p^T} - h^L(\zeta)p^L(\zeta)\right) \\ &= \phi h^T p^T \frac{(1 - h^L(\zeta)^2 p^L(\zeta)^2)}{1 + h^L(\zeta)p^L(\zeta)h^T p^T}\end{aligned}$$

Because we assumed that $p^L(\zeta) < 1$, this expression has the same sign as $\phi h^T p^T$. We want to show that $\phi(E[h|h^T = 1, h^L(\zeta)] - E[h|h^L(\zeta)]) \geq 0$ and $\phi(E[h|h^T = -1, h^L(\zeta)] - E[h|h^L(\zeta)]) \leq 0$. This follows trivially when $p^T = 0$. When $p^T > 0$ if follows if and only if $\phi \geq 0$.

B2. Proof of Proposition 2 (Bounding)

In the Bayesian model, given $\phi \geq 0$ (Monotonicity), the action is larger or smaller than $a(\zeta)$ when $\phi E[h|h^T, h^L] \geq 0$ or $\phi E[h|h^T, h^L] \leq 0$ respectively. Given that $\phi \geq 0$, we need $E[h|h^T = 1, h^L] \geq 0$ and $E[h|h^T = -1, h^L] \leq 0$. This is guaranteed if h^T and h^L have the same sign, so we simply need to check whether it holds when the demand treatment and latent demand are in opposite directions, i.e. $E[h|h^T = 1, h^L = -1] \geq 0$ and $E[h|h^T = -1, h^L = 1] \leq 0$. Given our restriction $p^L(\zeta) < 1$, inspection of (7) reveals that these conditions hold if and only if $p^T \geq p^L(\zeta)$, i.e. the decision-maker perceives the demand treatment as at least as informative about h as the latent demand signal.

B3. Conditions for Monotone Sensitivity

Assumption 3 (Monotone Sensitivity) assumes that sensitivity $S(\zeta) = a^+(\zeta) - a^-(\zeta)$ is (strictly) monotone in the size of the latent demand effect $\left|a^L(\zeta) - a(\zeta)\right|$. Here we examine cases under which that is and is not the case. We assume throughout that Assumptions 1 and 2 hold.

Variation driven by ϕ.

We are interested in how ϕ affects latent demand $(d\left|a^L(\zeta) - a(\zeta)\right|/d\phi)$ and sensitivity $(dS(\zeta)/d\phi)$. From (5) we obtain:

$$\frac{d(a^L(\zeta) - a(\zeta))}{d\phi} = -\frac{h^L(\zeta)p^L(\zeta)}{v_{11}(a^L(\zeta), \zeta)}$$

which has the same sign as $h^L(\zeta)$, allowing us to write $\frac{d\left|a^L(\zeta)-a(\zeta)\right|}{d\phi} = -\frac{p^L(\zeta)}{v_{11}(a^L(\zeta),\zeta)} \geq 0$.

Turning to sensitivity, we have:

$$\begin{aligned}\frac{dS(\zeta)}{d\phi} &= \frac{da^+(\zeta)}{d\phi} - \frac{da^-(\zeta)}{d\phi} \\ &= -\frac{1}{v_{11}(a^+(\zeta),\zeta)}\frac{h^L(\zeta)p^L(\zeta)+p^T}{1+h^L(\zeta)p^L(\zeta)p^T} + \frac{1}{v_{11}(a^-(\zeta),\zeta)}\frac{h^L(\zeta)p^L(\zeta)-p^T}{1-h^L(\zeta)p^L(\zeta)p^T}\end{aligned}$$

By Assumption 2, $h^L(\zeta)p^L(\zeta)+p^T \geq 0$ and $h^L(\zeta)p^L(\zeta)+p^T \leq 0$, so both terms are positive, i.e. $\frac{dS(\zeta)}{d\phi} \geq 0$. Therefore Monotone Sensitivity holds and any set of environments that differ only in ϕ constitutes a comparison class, i.e. for such environments, sensitivity is informative about the magnitude of latent demand effects.

EXAMPLE 2: *Suppose participant pool A is more concerned for pleasing the experimenter than participant pool B. Then latent demand effects and sensitivity will be larger in magnitude in participant pool A.*

Variation driven by v.

Suppose that ζ can be separated into a parameter, z, and a remainder term, ζ', that v is differentiable in z and that ϕ, h^L and p^L do not depend on z. z could be a preference parameter (e.g. risk aversion) or a design parameter (e.g. the scale of incentives). We write $U(a,\zeta',z) = v(a,\zeta',z) + a\phi(\zeta')E[h|\zeta']$ and modify the first-order conditions accordingly.

$$\begin{aligned}\frac{d(a^L(\zeta',z)-a(\zeta',z))}{dz} &= \frac{da^L(\zeta',z)}{dz} - \frac{da(\zeta',z)}{dz} \\ &= -\left[\frac{v_{13}(a^L(\zeta',z),\zeta',z)}{v_{11}(a^L(\zeta',z),\zeta',z)} - \frac{v_{13}(a(\zeta',z),\zeta',z)}{v_{11}(a(\zeta',z),\zeta',z)}\right] \\ \frac{dS(\zeta',z)}{dz} &= -\left[\frac{v_{13}(a^+(\zeta',z),\zeta',z)}{v_{11}(a^+(\zeta',z),\zeta',z)} - \frac{v_{13}(a^-(\zeta',z),\zeta',z)}{v_{11}(a^-(\zeta',z),\zeta',z)}\right]\end{aligned}$$

It is clear from inspecting these conditions that we need to know how v_{13}/v_{11} varies with a, i.e.:

$$\frac{d\frac{v_{13}(a,\zeta',z)}{v_{11}(a,\zeta',z)}}{da} = \frac{v_{11}(a,\zeta',z)v_{113}(a,\zeta',z) - v_{111}(a,\zeta',z)v_{13}(a,\zeta',z)}{v_{11}(a,\zeta',tz)}$$

It is difficult to make general statements about these objects for general utility functions, so we focus attention on two special cases of interest.

Multiplicative separability.

Suppose that $v(a,\zeta',z)=\nu(a,\zeta')f(z)$ and define z such that $f'(z)>0$. Then

$$\begin{aligned}\frac{d\left(a^L(\zeta',z)-a(\zeta',z)\right)}{dz} &= -f'(z)\left[\frac{\nu_1(a^L(\zeta',z),\zeta')}{\nu_{11}(a^L(\zeta',z),\zeta')}-\frac{\nu_1(a(\zeta',z),\zeta')}{\nu_{11}(a(\zeta',z),\zeta')}\right]\\ &= -f'(z)\frac{\nu_1(a^L(\zeta',z),\zeta')}{\nu_{11}(a^L(\zeta',z),\zeta')}\end{aligned}$$

Since by concavity $\nu_1(a,\zeta')>0$ for $a<a(\zeta',z)$ and $\nu_1(a,\zeta')<0$ for $a>a(\zeta',z)$, we have $\frac{d\left|a^L(\zeta',z)-a(\zeta',z)\right|}{dz}\leq 0$. Similarly

$$\frac{dS(\zeta)}{dz}=-f'(z)\left[\frac{\nu_1(a^+(\zeta',z),\zeta')}{\nu_{11}(a^+(\zeta',z),\zeta')}-\frac{\nu_1(a^-(\zeta',z),\zeta')}{\nu_{11}(a^-(\zeta',z),\zeta')}\right]$$

Since $\nu_1(a^+(\zeta',z),\zeta')\leq 0$ and $\nu_1(a^-(\zeta',z),\zeta')\geq 0$, we have $\frac{dS(\zeta)}{dz}\leq 0$. Therefore Monotone Sensitivity holds and any set of environments that varies only in z is a valid comparison set.

Intuitively, this case captures changes in the slope of payoffs that leave the optimal natural action unchanged. For example, an increase in the scale of incentives that makes the payoff function "more concave" around the natural action makes deviating from the natural action more costly and so decreases the magnitude of latent demand and sensitivity.

EXAMPLE 3 (Belief scoring): *Consider a belief-reporting task rewarded by a quadratic scoring rule. A risk-neutral participant reports a belief, a, which is the probability of an event A. He is paid $\frac{z}{2}\left[1-(\mathbb{I}[A]-a)^2\right]$ where $\mathbb{I}[A]=1$ if A is true and 0 otherwise. The utility function is $U(a,\zeta',z)=\frac{z}{2}\left[1-\mu(1-a)^2-(1-\mu)(-a)^2\right]+a\phi(\zeta')E[h|\zeta']$, so $f(z)=z$. The optimal action solves $z\left[\mu(1-a^*)-(1-\mu)a^*\right]+\phi(\zeta')E[h|\zeta']=0$ or $a^*=\mu+\frac{\phi(\zeta')E[h|\zeta']}{z}$. Increases in z are equivalent to decreases in ϕ and decrease both the magnitude of latent demand effects, and sensitivity.*

Additive separability.

Suppose that $v(a,\zeta',z)=v(a,\zeta')+af(z)$ and define z such that $f'(z)>0$. Then:

$$\frac{d(a^L(\zeta',z)-a(\zeta',z))}{dz}=-f'(z)\left[\frac{1}{\nu_{11}(a^L(\zeta',z),\zeta')}-\frac{1}{\nu_{11}(a(\zeta',z),\zeta')}\right]$$

and

$$\frac{dS(\zeta)}{dz}=-f'(z)\left[\frac{1}{\nu_{11}(a^+(\zeta',z),\zeta')}-\frac{1}{\nu_{11}(a^-(\zeta',z),\zeta')}\right]$$

What matters in this case is how the concavity of v (and therefore ν) with respect to a varies with a. Suppose $\nu_{111} < 0$, so ν_{11} is decreasing in a, i.e. concavity is increasing. Then $\frac{dS(\zeta)}{dz} < 0$, i.e. increases in z decrease sensitivity. If $a^L(\zeta', z) > a(\zeta', z)$ then $\frac{d(a^L(\zeta',z)-a(\zeta',z))}{dz} < 0$ and if $a^L(\zeta', z) < a(\zeta', z)$ then $\frac{d(a^L(\zeta',z)-a(\zeta',z))}{dz} > 0$, so $\frac{d\left|a^L(\zeta',z)-a(\zeta',z)\right|}{dz} < 0$ and Monotone Sensitivity holds. Monotone Sensitivity also holds (with the inequalities reversed) for $\nu_{111} > 0$.

EXAMPLE 4 (Effort provision): *A participant performs a real-effort task for piece rate z with cost of effort $C(a)$, $C' > 0$,$C'' > 0$, $C''' > 0$. $U(a, \zeta', z) = za - C(a) + a\phi(\zeta')E[h|\zeta']$. The optimal action a^* solves $z - C'(a^*) + \phi(\zeta')E[h|\zeta'] = 0$. As z increases, a^* increases and responsiveness to latent demand or demand treatments decreases.*

Variation driven by inattention.

Suppose that with some probability ξ the participant is an attentive type who pays careful attention to the decision-making environment, and with probability $1-\xi$, he is inattentive. When inattentive, he takes some action $a^I(\zeta)$. $a^I(\zeta)$ might be equal to $a(\zeta)$, in which case the participant is only inattentive to experimenter demand, but it might differ if the participant is also inattentive to other design features.

While until now we have treated the actions as those of a representative agent, for this analysis it is more natural to work with expected or average actions over a sample. Denote by $\bar{a}(\zeta) = \xi a(\zeta) + (1-\xi)a^I(\zeta)$ the expected natural action, define $\bar{a}^L(\zeta), \bar{a}^+(\zeta), \bar{a}^-(\zeta)$ equivalently and let $\bar{S}(\zeta) = \bar{a}^+(\zeta) - \bar{a}^-(\zeta)$. The latent demand effect is now equal to $\left|\bar{a}^L(\zeta) - \bar{a}(\zeta)\right| = \xi\left|a^L(\zeta) - a(\zeta)\right|$, while $\bar{S}(\zeta) = \xi S(\zeta)$. Hence, if the variation in latent demand effects is driven by variation in attention, ξ, Monotone Sensitivity will hold, and any set of environments that varies only in participant attentiveness is a valid comparison set. Note that since we have assumed the participant is inattentive to both latent demand and the demand treatment, Bounding will hold if $p^T \geq p^L$ as before.

Variation driven by beliefs.

Consider changes to the environment that influence behavior only by altering participants' beliefs about the experimenter's objective, i.e. we consider variation in $h^L(\zeta)p^L(\zeta)$. Call this term H. $a(\zeta)$ is unaffected, so:

$$\frac{d(a^L(\zeta) - a(\zeta))}{dH} = -\frac{\phi(\zeta)}{v_{11}(a^L(\zeta), \zeta)} \geq 0$$

and therefore $\frac{d\left|a^L(\zeta)-a(\zeta)\right|}{dH} = -\frac{\phi(\zeta)}{v_{11}(a^L(\zeta),\zeta)} \times \text{sign}(a^L(\zeta) - a(\zeta)) = -\frac{\phi(\zeta)h^L(\zeta)}{v_{11}(a^L(\zeta),\zeta)}$ which is positive when $h^L(\zeta) = 1$ (because an increase in H means the participant's beliefs are shifting toward certainty that the experimenter wants a high action) and negative when $h^L(\zeta) = -1$ (because the participant is becoming more uncertain about the experimenter's wishes).

Next we turn to demand treatment effects. First we derive the response of the participant's posterior:

$$\frac{d\frac{H+h^T p^T}{1+Hh^T p^T}}{dH} = \frac{\left(1+Hh^T p^T\right) - \left(H + h^T p^T\right) h^T p^T}{\left(1+Hh^T p^T\right)^2}$$
$$= \frac{1-\left(h^T p^T\right)^2}{\left(1+Hh^T p^T\right)^2} = \frac{1-p^{T2}}{\left(1+Hh^T p^T\right)^2}$$

So:

$$\frac{dS(\zeta)}{dH} = -\phi(\zeta)(1-p^{T2})\left[\frac{1}{\left(1+Hp^T\right)^2 v_{11}(a^+(\zeta),\zeta)} - \frac{1}{\left(1-Hp^T\right)^2 v_{11}(a^-(\zeta),\zeta)}\right]$$

The sign of this expression depends on the sign of H and how v_{11} changes with a. However, it is straightforward to see that Monotone Sensitivity *will not* hold in general, and in fact sensitivity will tend to be higher when latent demand is weaker. To see this, consider the simple case where v_{11} is constant. Then we have:

$$\frac{dS(\zeta)}{dH} = -\frac{\phi(\zeta)(1-p^{T2})}{v_{11}}\left[\frac{\left(1-Hp^T\right)^2 - \left(1+Hp^T\right)^2}{\left(1+Hp^T\right)^2\left(1-Hp^T\right)^2}\right]$$
$$= -\frac{\phi(\zeta)(1-p^{T2})}{v_{11}}\left[\frac{-4Hp^T}{\left(1+Hp^T\right)^2\left(1-Hp^T\right)^2}\right]$$

which is positive when $h^L = -1$ and negative when $h^L = 1$, i.e. it has the opposite sign to $\frac{d\left|a^L(\zeta)-a(\zeta)\right|}{dH}$. The reason is that as H approaches zero, the participant becomes more uncertain about the experimenter's wishes and is therefore very responsive to the new information in the demand treatments. Meanwhile as H approaches 1 or -1, the participant is very confident about the value of h. Although his confidence can be undermined by a demand treatment in the opposite direction, he responds little to a demand treatment that confirms his beliefs, so sensitivity is low.

B4. Defiers

In this section we discuss defiance. We first derive a special case that illustrates how valid bounds can be obtained even when some participants defy the experimenter. Then we present three examples where defier behavior causes our key assumptions to break down.

Because our concern is with bounding rather than point identification, the method is able to tolerate some defiance. To illustrate, suppose that v is homogeneous across individuals, quadratic in a, and normalized such that $v_1(a,\zeta) = b(\zeta) - a$ where b is a constant. The natural action is equal to b for all individuals. Beliefs and ϕ are heterogeneous across individuals, indexed by i. For compactness, label the beliefs $H_i^L := h_i^L p_i^L$, $H_i^+ := (H_i^L + p^T)/(1 + H_i^L p^T)$ and $H_i^- := (H_i^L - p^T)/(1 - H_i^L p^T)$. Under our assumptions, the actions of interest are given by:

$$a_i^L = b + \phi_i H_i^L \qquad a_i^+ = b + \phi_i H_i^+ \qquad a_i^- = b + \phi_i H_i^-$$

Then, for Bounding to hold on average in the population, we require $E[\phi_i H_i^+] \geq 0 \geq E[\phi_i H_i^-]$, where expectations are over participants. If $p^T \geq p_i^L$ for all individuals, then $H_i^+ \geq 0 \geq H_i^-$, so the conditions are equivalent to weighted averages of ϕ having the correct sign, where the weights are the beliefs. A special case of interest is that where all individuals have identical H_i^+ and H_i^- (this is the case if latent demand (H_i^L) is the same for all individuals, or if $p^T = 1$). Then both conditions reduce to $E[\phi_i] \geq 0$, i.e. Bounding holds if the average participant is a complier.[1]

Now we provide three simple examples where defier behavior causes our key assumptions to break down. First we show that it is possible for Bounding to hold without Monotonicity, second that it is possible for Monotonicity to hold without Bounding, and third that both can fail while retaining well-ordered bounds.

Let all decision makers share $v(a) = -a^2$, so the natural action $a = 0$. 2/3 of the population are compliers with $\phi = \phi_C = 1$ and 1/3 are defiers with $\phi = \phi_D = -1$. Latent demand signals are assumed common within complier/defier groups but different between compliers and defiers. They equal $H_C^L = h_C^L p_C^L$ and $H_D^L = h_D^L p_D^L$ respectively, with corresponding beliefs following the demand treatments equal to:

$$H_i^+ = \frac{H_i^L + p^T}{1 + H_i^L p^T} \qquad H_i^- = \frac{H_i^L - p^T}{1 - H_i^L p^T}$$

We retain the assumption of common p^T. Then the observed average actions under latent demand, positive and negative demand treatments are:

$$E[a^L] = \frac{1}{3}(2H_C^L - H_D^L) \qquad E[a^+] = \frac{1}{3}(2H_C^+ - H_D^+) \qquad E[a^-] = \frac{1}{3}(2H_C^- - H_D^-)$$

Our first example shows that Bounding can hold without Monotonicity. Thus a Monotonicity failure does not imply a failure of Bounding, but it is a warning sign of the presence of defiers.

[1] For Monotonicity to hold on average we require $E[\phi_i(H_i^+ - H_i^L)] \geq 0 \geq E[\phi_i(H_i^L - H_i^-)]$. Since $H_i^+ - H_i^L > 0$ and $H_i^L - H_i^- < 0$, these conditions require that a weighted average of ϕ has the correct sign, where the weights are the belief *changes* induced by the demand treatments. Violations of Monotonicity or, in the extreme case, reversed bounds ($a^- > a^+$), are clear cause for concern. However it is possible for Monotonicity to hold on average while Bounding fails and vice versa.

EXAMPLE 5 (Bounding without Monotonicity): *Suppose* $H_C^L = 0.5$, $H_D^L = -0.5$ *and* $p^T = 1$. *Then* $E[a^L] = 0.5$, $E[a^+] = 1/3$ *and* $E[a^-] = -1/3$. *Therefore* $E[a^-] < a < E[a^+] < E[a^L]$.

Our second example shows that Bounding can fail while Monotonicity holds. This is possible in the basic model if $p^T < p^L$, but can also be caused by defiance.

EXAMPLE 6 (Monotonicity without Bounding.): *Suppose* $H_C^L = 0.5$, $H_D^L = -0.5$ *and* $p^T = 0.5$. *Then* $E[a^L] = 0.5$, $E[a^+] = \frac{8}{15}$ *and* $E[a^-] = \frac{4}{15}$. *Thus* $a < E[a^-] < E[a^L] < E[a^+]$.

Our third example shows that both Bounding and Monotonicity can fail, while still producing well-ordered bounds (i.e. $a^+ > a^-$).

EXAMPLE 7 (No Bounding or Monotonicity.): *Let* $H_C^L = 0.75$, $H_D^L = 0$, *and* $p^T = 0.75$. *Then* $E[a^L] = 0.5$, $E[a^+] = \frac{39}{100}$ *and* $E[a^-] = 1/4$. *Therefore* $a < E[a^-] < E[a^+] < E[a^L]$.

B5. Extension: learning about ϕ

A possible interpretation of our demand treatments is that they signal not only the direction of the experimenter's objective, but the salience or intensity of her preference over objectives. For instance, "do me a favor" suggests that the choice is important. In this section we extend the model to incorporate this feature, allowing ϕ to depend upon a belief about the "importance" of the objective. We assume that the decision-maker responds more strongly to experimenter demand when they believe that complying with the objective it is more important, and that this belief depends both on latent demand and the demand treatments. Specifically, we now assume that the decision-maker's preferences are:

$$U(a, \zeta) = v(a, \zeta) + a\phi(\zeta)E[gh|\zeta]$$

where $g \in \{0, 1\}$ captures whether conforming to h is important (1) or unimportant (0) to the experimenter. ϕ remains the decision-maker's preference for pleasing the experimenter, which is now scaled by g, i.e. the decision-maker internalizes the perceived importance of the objective. We assume that g and h are believed independent (i.e. direction and importance are independent), so $E[gh|\zeta] = E[g|\zeta]E[h|\zeta]$. We also assume for simplicity is that the decision-maker's prior $E[g] = 0.5$.

Now, ζ contains two signals, $h^L(\zeta)$, defined as before, and $g^L(\zeta) \in \{0, 1\}$, where $E[g|g^L(\zeta)] = E[g|g^L(\zeta), \zeta]$ (i.e. g^L is a sufficient statistic). g^L is believed to equal g with probability $q^L(\zeta) < 1$ and pure independent noise otherwise. We show below that $E[g|g^L(\zeta)] = \frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)$.

Similarly, a demand treatment is now two signals (h^T, g^T), where h^T is defined as before and $g^T \in \{0, 1, \emptyset\}$. $g^T = \emptyset$ corresponds to the case where no treatment is used, $g^T = 0$ signals to the participant that their action is not important to the experimenter, and $g^T = 1$ signals that it is.

Conditional on sending a demand treatment, g^T is believed to equal g with probability q^T and otherwise be pure noise independent of all other signals. We show below that the Bayesian posterior is:

$$E[g|g^T, g^L(\zeta)] = \frac{\frac{1}{2} + q^L(\zeta)\left(g^L(\zeta) - \frac{1}{2}\right) + q^T\left(g^T - \frac{1}{2}\right) + q^T q^L(\zeta)\left(\mathbb{I}[g^T = g^L(\zeta)] - \frac{1}{2}\right)}{1 + 2q^T q^L(\zeta)\left(\mathbb{I}[g^T = g^L(\zeta)] - \frac{1}{2}\right)}$$

We assume that g^T can be varied independently of h^T and will be held constant within a typical pair of positive and negative demand treatments.

For Bounding to hold, we now need:

$$\phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 0, h^L(\zeta)] \leq 0 \leq \phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 1, h^L(\zeta)]$$

Since $E[g|g^T, g^L(\zeta)] \geq 0$ our Bounding condition does *not* depend on how the demand treatments affect beliefs about g, all we require is $\phi(\zeta) \geq 0$ and $p^T \geq p^L(\zeta)$ as before.[2]

However, beliefs about g do affect the width of the bounds: sensitivity is increasing in $E[g|g^T, g^L(\zeta)]$. The tightest bounds are obtained when $E[g|g^T, g^L(\zeta)] = 0$, which obtains when $g^T = 0$ and $q^T = 1$. More generally, the bounds are tightened by signaling that acting according to the experimenter's objective is not important ($g^T = 0$), or if $g^T = 1$ by minimizing q^T. We suspect that it may be difficult in practice to both strongly signal the direction of the objective (large p^T), which is required for Bounding, and that the objective is not important ($g^T = 0$), so reasonable demand treatments are likely to be those that strongly signal a directional objective while keeping salience low, i.e. large p^T and small q^T with $g^T = 1$.

[2]For Monotonicity to hold, we require

$$\phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 0, h^L(\zeta)] \leq \phi(\zeta)E[g|g^L(\zeta)]E[h|h^L(\zeta)] \leq \phi(\zeta)E[g|g^T, g^L(\zeta)]E[h|h^T = 1, h^L(\zeta)]$$

We can write

$$\phi(\zeta)\frac{E[h|h^T = 0, h^L(\zeta)]}{E[h|h^L(\zeta)]} \leq \phi(\zeta)\frac{E[g|g^L(\zeta)]}{E[g|g^T, g^L(\zeta)]} \leq \phi(\zeta)\frac{E[h|h^T = 1, h^L(\zeta)]}{E[h|h^L(\zeta)]}$$

We see that $\phi(\zeta) \geq 0$ is necessary but not sufficient for Monotonicity, we also need that $E[g|g^T, g^L(\zeta)]$ is neither "too big" nor "too small" relative to $E[g|g^L(\zeta)]$. Intuitively, if $g^T = 1$ the demand treatments shift all actions further away from the natural action, while if $g^T = 0$. all actions are shifted toward the natural action. $g^T = 1$ and $p^T \geq p^L$ are sufficient for Monotonicity to hold.

Derivation of $E[g|g^L(\zeta)]$ and $E[g|g^T, g^L(\zeta)]$

Let the prior belief be $\frac{1}{2}$.

$$E[g|g^L = y] = Pr(g = 1|g^L = y) = \frac{A}{B}$$
$$A = Pr(g^L = y|g = 1)Pr(g = 1)$$
$$B = Pr(g^L = y|g = 1)Pr(g = 1) + Pr(g^L = y|g = 0)Pr(g = 0)$$

Since $Pr(g = j|g^L = y) = \frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = j]$ and $Pr(g = j) = \frac{1}{2}$ we have

$$\begin{aligned} A &= \frac{1}{2}\left(\frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = 1]\right) \\ &= \frac{1}{2}\left(\frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)\right) \\ B &= \frac{1}{2}\left[\left(\frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = 1]\right) + \left(\frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = 0]\right)\right] = \frac{1}{2} \end{aligned}$$

Therefore, $E[g|g^L(\zeta)] = \frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right)$.

Turning to $E[g|g^T, g^L(\zeta)]$, we have assumed that when $g^T = \emptyset$, $E[g|g^T, g^L] = E[g|g^L]$. After observing $g^T \neq \emptyset$, the participant forms a posterior:

$$E[g|g^T, g^L] = Pr(g = 1|g^T, g^L) = \frac{A}{B}$$
$$A = Pr(g^T = x|g = 1, g^L = y)Pr(g = 1|g^L = y)$$
$$\begin{aligned} B = &Pr(g^T = x|g = 1, g^L = y)Pr(g = 1|g^L = y) \\ &+ Pr(g^T = x|g = 0, g^L = y)Pr(g = 0|g^L = y) \end{aligned}$$

Using the following

$$Pr(g^T = x|g = j, g^L = y) = \frac{1}{2}(1 - q^T) + q^T \mathbb{I}[x = j]$$
$$Pr(g = j|g^L = y) = \frac{1}{2}(1 - q^L) + q^L \mathbb{I}[y = j]$$

we have:

$$
\begin{aligned}
A &= \left(\frac{1}{2}(1-q^T) + q^T\mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-q^L) + q^L\mathbb{I}[y=1]\right)\\
&= \left(\frac{1}{2}(1-q^T) + q^T g^T\right)\left(\frac{1}{2}(1-q^L) + q^L g^L\right)\\
&= \frac{1}{2}q^L g^L + \frac{1}{2}q^T g^T - \frac{1}{2}q^T q^L\left(g^L(1-g^T) + g^T(1-g^L)\right)\\
&+ \frac{1}{4}(1-q^T)(1-q^L)\\
&= \frac{1}{2}q^L g^L + \frac{1}{2}q^T g^T - \frac{1}{2}q^T q^L\left(\mathbb{I}[g^L \neq g^T]\right)\\
&+ \frac{1}{4} - \frac{1}{4}q^T - \frac{1}{4}q^L + \frac{1}{4}q^T q^L\\
&= \frac{1}{2}q^L\left(g^L - \frac{1}{2}\right) + \frac{1}{2}q^T\left(g^T - \frac{1}{2}\right) - \frac{1}{2}q^T q^L\left(1 - \mathbb{I}[g^L = g^T]\right)\\
&+ \frac{1}{4} + \frac{1}{4}q^T q^L\\
&= \frac{1}{4} + \frac{1}{2}q^L\left(g^L - \frac{1}{2}\right) + \frac{1}{2}q^T\left(g^T - \frac{1}{2}\right)\\
&+ \frac{1}{2}q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
B &= \left(\frac{1}{2}(1-q^T) + q^T\mathbb{I}[x=1]\right)\left(\frac{1}{2}(1-q^L) + q^L\mathbb{I}[y=1]\right)\\
&+ \left(\frac{1}{2}(1-q^T) + q^T\mathbb{I}[x=0]\right)\left(\frac{1}{2}(1-q^L) + q^L\mathbb{I}[y=0]\right)\\
&= \left(\frac{1}{2}(1-q^T) + q^T g^T\right)\left(\frac{1}{2}(1-q^L) + q^L g^L\right)\\
&+ \left(\frac{1}{2}(1-q^T) + q^T(1-g^T)\right)\left(\frac{1}{2}(1-q^L) + q^L(1-g^L)\right)\\
&= \frac{1}{2}(1-q^T)q^L + \frac{1}{2}(1-q^L)q^T + \frac{1}{2}(1-q^T)(1-q^L)\\
&+ q^T q^L\mathbb{I}[g^T = g^L]\\
&= \frac{1}{2} + q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)
\end{aligned}
$$

Therefore,

$$E[g|g^T, g^L] = \frac{\frac{1}{2} + q^L\left(g^L - \frac{1}{2}\right) + q^T\left(g^T - \frac{1}{2}\right) + q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)}{1 + 2q^T q^L\left(\mathbb{I}[g^T = g^L] - \frac{1}{2}\right)}$$

B6. *Richer beliefs and correlated signals*

Researchers sometimes give experimental participants instructions like "there are no right or wrong answers" or "we are only interested in what you think is the best choice." This can be thought of as a demand treatment that demands participants choose the natural action, $a(\zeta)$.

It is straightforward to analyze such treatments in our framework. In this section, we extend the model to allow h to take three values: $\{-1, 0, 1\}$, where $h = 0$ captures the case where the experimenter wants the participant to choose the natural action. We call the action following $h^T = 0$, $a^0(\zeta)$.

For simplicity we assume that the participant's prior belief is that each possibility is equally likely (i.e. is true with probability 1/3), so $E[h] = 0$. ϵ and η are also believed to take each value with probability 1/3 and are independent. $h^L \in \{-1, 0, 1\}$ and $h^T \in \{-1, 0, 1, \emptyset\}$ and p^L and p^T are defined as before. We maintain the assumption that the participant infers nothing when the experimenter does not send a demand treatment ($h^T = \emptyset$).

We show below that the beliefs can be written as:

$$E[h|h^L] = p^L h^L \tag{B1}$$

$$E[h|h^T = \emptyset, h^L] = p^L h^L \tag{B2}$$

$$E[h|h^T, h^L] = \frac{\frac{1}{3}(1 - p^T)p^L h^L + \frac{1}{3}(1 - p^L)p^T h^T + p^T p^L h^T \mathbb{I}[h^T = h^L]}{\frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^T = h^L]} \tag{B3}$$

Bounding holds if $E[h|h^T = 1, h^L] \geq 0$ and $E[h|h^T = -1, h^L] \leq 0$. It is straightforward to check that the condition is the same as before: $p^T \geq p^L$.

What purpose, then, do $h^T = 0$ treatments serve? It is natural to think that demanding participants to take the natural action will eliminate demand effects, but under our assumptions, $h^T = 0$ does not in general elicit the natural action. Instead latent demand still influences the participant's action. We have:

$$E[h|h^T = 0, h^L] = \frac{\frac{1}{3}(1 - p^T)p^L h^L}{\frac{1}{3}\left(1 - p^T p^L\right) + p^T p^L \mathbb{I}[h^L = 0]}$$

This expression equals zero if $p^T = 1$ (the demand treatment is perfectly informative), or $p^L h^L = 0$ (no latent demand), otherwise it has the same sign as $p^L h^L$. One interpretation is that while the participant takes at face value the experimenter's demand to choose the natural action, he might be unaware of the influence of other design features that nudge him in one direction or another.

Despite this negative result, $h^T = 0$ treatments can still be useful. First, they are informative about the *sign* of the bias due to latent demand. This is because $E[h|h^T = 0, h^L] \in [\min\{E[h|h^L], 0\}, \max\{E[h|h^L], 0\}]$ and therefore $a^0(\zeta) \in [\min\{a^L(\zeta), a(\zeta)\}, \max\{a^L(\zeta), a(\zeta)\}]$.[3] The action taken when $h^T = 0$ lies between the natural action and the action induced by latent demand, because the demand treatment shifts the participant's posterior toward zero.

Second, they can be used to obtain tighter bounds on $a(\zeta)$ if we know the direction of the latent demand effect. Suppose for example we know that $a^L(\zeta) \geq a(\zeta)$ (either from prior information or because we ran a treatment with $h^T = 0$ and verified that $a^0(\zeta) \leq a^L(\zeta)$). Then, the interval $[a^-(\zeta), a^0(\zeta)]$ gives a valid and tighter bound on $a(\zeta)$ than $[a^-(\zeta), a^+(\zeta)]$. Formally $a(\zeta) \in [a^-(\zeta), a^0(\zeta)] \subseteq [a^-(\zeta), a^+(\zeta)]$.[4]

Finally, there is one important case in which $h^T = 0$ perfectly recovers the natural action, i.e. $a^0(\zeta) = a(\zeta)$. Suppose that instead of assuming that the signals h^T and h^L contain independent shocks, the participant perceives that h^L is a noisy signal of h^T. Formally, he believes that with probability $p^L < 1$, $h^L = h^T$ and with probability $(1 - p^L)$, $h^L = \epsilon$. Then, when h^T and h^L disagree, he knows that h^L is pure noise, when they agree h^L contains no more information than h^T. Hence, the participant disregards h^L after observing h^T and $E[h|h^T, h^L] = p^T h^T$. Then, sending $h^T = 0$ recovers the natural action: $E[h|h^T = 0, h^L] = 0, \forall h^L$. An advantage of our bounds is that they are valid whether or not h^T or h^L are perceived as independent, in other words they are conservative relative to the approach of simply measuring $a^0(\zeta)$.

To summarize, unless the demand treatment is perceived as fully informative ($p^T = 1$), signaling $h^T = 0$ does *not* induce the participant to take the natural action, i.e. $a^0(\zeta) \neq a(\zeta)$. The intuition is that such a treatment does not eliminate all of the influence of latent demand – the decision-maker views both signals as informative and weighs them against one another, so the posterior belief lies between 0 and $E[h|h^L]$. However, because signaling $h^T = 0$ moves actions toward the natural action it can be informative about the *direction* of latent demand. In contrast, in an alternative formulation with non-independent signals, where participants perceive the demand treatments to contain the same information as latent demand but less noise, signaling $h^T = 0$ does elicit the natural action. Thus, demanding the natural action does not necessarily obtain bounds that contain the natural action, while a pair of sufficiently informative positive and negative demand treatments does.

Derivation of beliefs with ternary signals

Recall that now $h \in \{-1, 0, 1\}$, $h^L \in \{-1, 0, 1\}$ and $h^T \in \{-1, 0, 1, \emptyset\}$.

To avoid clutter we suppress dependence on ζ. After observing h^L, the participant forms a posterior $E[h|h^L] = Pr(h = 1|h^L) \times 1 + Pr(h = -1|h^L) \times (-1)$. We

[3] To see this, note that $\left|E[h|h^L] - E[h|h^T = 0, h^L]\right| \geq 0$ and both have the same sign.

[4] We thank Liad Weiss for pointing this out to us.

can write this as:

$$
\begin{aligned}
E[h|h^L = y] = Pr(h = 1|h^L = y) - Pr(h = -1|h^L = y) = \frac{A}{B} \\
A = Pr(h^L = y|h = 1)Pr(h = 1) - Pr(h^L = y|h = -1)Pr(h = -1) \\
B = Pr(h^L = y|h = 1)Pr(h = 1) + Pr(h^L = y|h = 0)Pr(h = 0) \\
+ Pr(h^L = y|h = -1)Pr(h = -1)
\end{aligned}
$$

Since $Pr(h = j|h^L = y) = \frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = j]$ and $Pr(h = j) = \frac{1}{3}$ we have

$$
\begin{aligned}
A &= \frac{1}{3}\left[\left(\frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = 1]\right) - \left(\frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = -1]\right)\right] \\
&= \frac{1}{3}p^L\left[\mathbb{I}[y = 1] - \mathbb{I}[y = -1]\right] = \frac{1}{3}p^L h^L \\
B &= \frac{1}{3}\left[\left(\frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = 1]\right) + \left(\frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = 0]\right)\right. \\
&\quad \left. + \left(\frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = -1]\right)\right] = \frac{1}{3}
\end{aligned}
$$

So

$$
E[h|h^L = y] = p^L h^L \tag{B4}
$$

just as before. Turning to beliefs following the demand treatments, as before we assume that when $h^T = \emptyset$, $E[h|h^T, h^L] = E[h|h^L]$. We have:

$$
\begin{aligned}
E[h|h^T, h^L] = Pr(h = 1|h^T, h^L) - Pr(h = -1|h^T, h^L) = \frac{A}{B} \\
A = Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
- Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y) \\
B = Pr(h^T = x|h = 1, h^L = y)Pr(h = 1|h^L = y) \\
+ Pr(h^T = x|h = 0, h^L = y)Pr(h = 0|h^L = y) \\
+ Pr(h^T = x|h = -1, h^L = y)Pr(h = -1|h^L = y, h^L = y)
\end{aligned}
$$

Using

$$
\begin{aligned}
Pr(h^T = x|h = j, h^L = y) = \frac{1}{3}(1 - p^T) + p^T\mathbb{I}[x = j] \\
Pr(h = j|h^L = y) = \frac{1}{3}(1 - p^L) + p^L\mathbb{I}[y = j]
\end{aligned}
$$

we obtain:

$$\begin{aligned}
A &= \left(\frac{1}{3}(1-p^T) + p^T\mathbb{I}[x=1]\right)\left(\frac{1}{3}(1-p^L) + p^L\mathbb{I}[y=1]\right) \\
&- \left(\frac{1}{3}(1-p^T) + p^T\mathbb{I}[x=-1]\right)\left(\frac{1}{3}(1-p^L) + p^L\mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}(1-p^T)p^L\mathbb{I}[y=1] + \frac{1}{3}(1-p^L)p^T\mathbb{I}[x=1] + p^Tp^L\mathbb{I}[x=1]\mathbb{I}[y=1] \\
&- \frac{1}{3}(1-p^T)p^L\mathbb{I}[y=-1] - \frac{1}{3}(1-p^L)p^T\mathbb{I}[x=-1] - p^Tp^L\mathbb{I}[x=-1]\mathbb{I}[y=-1] \\
&= \frac{1}{3}(1-p^T)p^Lh^L + \frac{1}{3}(1-p^L)p^Th^T + p^Tp^Lh^T\mathbb{I}[h^T=h^L]
\end{aligned}$$

$$\begin{aligned}
B &= \left(\frac{1}{3}(1-p^T) + p^T\mathbb{I}[x=1]\right)\left(\frac{1}{3}(1-p^L) + p^L\mathbb{I}[y=1]\right) \\
&+ \left(\frac{1}{3}(1-p^T) + p^T\mathbb{I}[x=0]\right)\left(\frac{1}{3}(1-p^L) + p^L\mathbb{I}[y=0]\right) \\
&+ \left(\frac{1}{3}(1-p^T) + p^T\mathbb{I}[x=-1]\right)\left(\frac{1}{3}(1-p^L) + p^L\mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}(1-p^T)(1-p^L) + \frac{1}{3}p^T(1-p^L)\left(\mathbb{I}[x=1] + \mathbb{I}[x=0] + \mathbb{I}[x=-1]\right) \\
&+ \frac{1}{3}p^L(1-p^T)\left(\mathbb{I}[y=1] + \mathbb{I}[y=0] + \mathbb{I}[y=-1]\right) \\
&+ p^Tp^L\left(\mathbb{I}[x=1]\mathbb{I}[y=1] + \mathbb{I}[x=0]\mathbb{I}[y=0] + \mathbb{I}[x=-1]\mathbb{I}[y=-1]\right) \\
&= \frac{1}{3}\left(1-p^Tp^L\right) + p^Tp^L\mathbb{I}[h^T=h^L]
\end{aligned}$$

So

$$E[h|h^T,h^L] = \frac{\frac{1}{3}(1-p^T)p^Lh^L + \frac{1}{3}(1-p^L)p^Th^T + p^Tp^Lh^T\mathbb{I}[h^T=h^L]}{\frac{1}{3}\left(1-p^Tp^L\right) + p^Tp^L\mathbb{I}[h^T=h^L]} \tag{B5}$$

B7. Computing confidence intervals

Here we describe how we compute demand-robust confidence intervals. We note that this was not included in the pre-analysis plans.

Correction. This section was corrected in August 2019, making the following changes:

1) The original online appendix misstated the equations for the confidence intervals. The equations we gave for 95% confidence intervals on parameters would in fact generate 90% confidence intervals, while the equations for 95%

confidence intervals on *sets* would in fact generate 95% confidence intervals on *parameters*. We have corrected these and clarified the notation.

2) Tables A3, A4, and A6 reflect the corrected confidence interval equations, as well as a correction of duplicated figures in the "Lying" column of Table A3. Quantitatively the changes are small in magnitude since the contribution of sampling uncertainty to the overall width of the confidence intervals is relatively small.

3) We added a discussion of Stoye (2009).

Confidence intervals for actions

Imbens and Manski (2004) show that asymptotically the probability that the estimate for the upper (lower) bound is lower (higher) than the true value can be ignored when making inference. The 95 percent confidence interval for the true demand-free behavior is given by:

$$CI_{95\%}^{a(\zeta)} = \left[a^{-}(\zeta) - \overline{C_N}\frac{\widehat{\sigma^{-}}}{\sqrt{N}},\ a^{+}(\zeta) + \overline{C_N}\frac{\widehat{\sigma^{+}}}{\sqrt{N}}\right]$$

Here, $\widehat{\sigma^{-}} = \sqrt{\widehat{Var(a^{-}(\zeta))}}$ and $\widehat{\sigma^{+}} = \sqrt{\widehat{Var(a^{+}(\zeta))}}$, and $\overline{C_N}$ satisfies

$$\Phi\left(\overline{C_N} + \sqrt{N}\frac{a^{+}(\zeta) - a^{-}(\zeta)}{\max(\widehat{\sigma^{-}},\ \widehat{\sigma^{+}})}\right) - \Phi(-\overline{C_N}) = 0.95.$$

The 95 percent confidence interval for the set $[a^{-}(\zeta), a^{+}(\zeta)]$ is just given by the conventional 95 percent confidence interval:

$$CI_{95\%}^{[a^{-}(\zeta),a^{+}(\zeta)]} = \left[a^{-}(\zeta) - 1.96\frac{\widehat{\sigma^{-}}}{\sqrt{N}},\ a^{+}(\zeta) + 1.96\frac{\widehat{\sigma^{+}}}{\sqrt{N}}\right]$$

Confidence intervals for treatment effects

We also outline how one can compute confidence intervals for the treatment effects $[a(\zeta_1) - a(\zeta_0)]$ and for the set defined by the upper and lower bounds for treatment effects as given by our demand treatments: $[a(\zeta_1) - a(\zeta_0)] \in [a^{-}(\zeta_1) - a^{+}(\zeta_0), a^{+}(\zeta_1) - a^{-}(\zeta_0)]$

For simplicity we denote the lower bound, $[a^{-}(\zeta_1) - a^{+}(\zeta_0)]$, as T^{-} and the upper bound, $[a^{+}(\zeta_1) - a^{-}(\zeta_0)]$, as T^{+}. The 95 percent confidence interval for the true demand-free treatment effect T is given by:

$$CI_{95\%}^{T} = \left[T^{-} - \overline{C_N} \frac{\widehat{\sigma^{T-}}}{\sqrt{N}},\, T^{+} + \overline{C_N} \frac{\widehat{\sigma^{T+}}}{\sqrt{N}} \right].$$

Here, $\widehat{\sigma^{T-}} = \sqrt{\widehat{Var(T^{-})}}$ and $\widehat{\sigma^{T+}} = \sqrt{\widehat{Var(T^{+})}}$, and $\overline{C_N}$ satisfies

$$\Phi\left(\overline{C_N} + \sqrt{N} \frac{T^{+} - T^{-}}{\max(\widehat{\sigma^{T-}},\, \widehat{\sigma^{T+}})} \right) - \Phi(-\overline{C_N}) = 0.95.$$

The 95 percent confidence interval for the set $[T^{-}, T^{+}]$ is as follows:

$$CI_{95\%}^{[T^{-},T^{+}]} = \left[T^{-} - 1.96 \frac{\widehat{\sigma^{T-}}}{\sqrt{N}},\, T^{+} + 1.96 \frac{\widehat{\sigma^{T+}}}{\sqrt{N}} \right].$$

Stoye (2009)

After publication of this paper, we learned of a result due to Stoye (2009). Stoye's Lemma 1 shows that Imbens and Manski (2004) rely on an implicit superefficiency condition: if the true bounds width is zero ($\Delta \equiv UB - LB = 0$, where UB and LB are the upper and lower bounds on the action or treatment effect), the estimated bounds width must also equal zero with probability approaching 1, in finite samples. This condition is not satisfied for our applications since our estimated bounds width is a difference in means of independent samples.

Stoye (2009) provides an alternative algorithm that is valid even in cases such as ours. However, it is complex to apply. We corresponded with Stoye and he gave us an alternative algorithm which is asymptotically equivalent and extremely simple to apply:

1) Estimate $\hat{\Delta}$ and compute the studentized bounds width $\hat{\Delta}/s.e.(\hat{\Delta})$.

2) Compare this object to the "tuning parameter" $\sqrt{\log(N)}/\sqrt{N}$.

3) If larger than the tuning parameter, presume $\Delta > 0$ and report $CI_{95\%} = [LB - 1.645 \times s.e.(LB), UB + 1.645 \times s.e.(UB)]$, i.e. equivalent to the conventional 90% confidence interval on $[LB, UB]$.

4) If smaller than the tuning parameter and $\hat{\Delta} \geq 0$, presume $\Delta = 0$ and report $CI_{95\%} = [LB - 1.96 \times s.e.(LB), UB + 1.96 \times s.e.(UB)]$, i.e. equivalent to the conventional 95% confidence interval on $[LB, UB]$.

5) If smaller than the tuning parameter and $\hat{\Delta} < 0$, presume $\Delta = 0$ and report $CI_{95\%} = [LB - 1.96 \times s.e.(LB), UB + 1.96 \times s.e.(UB)] \cup [\bar{\Delta} \pm 1.96 \times s.e.(\bar{\Delta})]$ where $\bar{\Delta}$ is a variance-weighted average of UB and LB.

Quantitatively, the above algorithm gives very similar results to those based on Imbens and Manski (2004) (results available on request). In most cases we believe it is unlikely that the true $\Delta = 0$. In the theory this only occurs if $\phi = 0$ or $p^T = 0$, such that the demand treatments do not affect behavior.

B8. Controlling for demand

Here we provide derivations for the results in section III.D. We begin with the usual first-order condition, assuming a demand treatment h^T:

$$0 = v_1(a^*(\zeta, h^T), \zeta) + \phi(\zeta)E[h|h^T, h^L(\zeta)]$$

taking the first-order Taylor approximation at the natural action $a(\zeta)$ we obtain:

$$0 \approx \underbrace{v_1(a(\zeta), \zeta)}_{=0} + \phi(\zeta)E[h|h^T, h^L(\zeta)] + [a^*(\zeta, h^T) - a(\zeta)]v_{11}(a(\zeta), \zeta)$$

where the first term is zero by definition of $a(\zeta)$. Rearranging we obtain equation (8):

$$\begin{aligned} a^*(\zeta, h^T) &\approx a(\zeta) - \frac{\phi(\zeta)}{v_{11}(a(\zeta), \zeta)}E[h|h^T, h^L(\zeta)]. \\ &= a(\zeta) + \Phi(\zeta)E[h|h^T, h^L(\zeta)]. \end{aligned}$$

Assume two treatment groups: $\zeta \in \{0, 1\}$, and denote their corresponding demand treatments by h^T_ζ. If no demand treatments are applied ($h^T_1 = h^T_0 = \emptyset$). Since we are interested in cases where $h^T_0 = h^T_1$ we suppress the subscripts. The approximate bias of the treatment effect estimate can be written as:

$$\begin{aligned} Bias &= a^*(1, \emptyset) - a^*(0, \emptyset) - [a(1) - a(0)] \\ &\approx a(1) + \Phi(1)E[h|h^T, h^L(1)] - a(0) - \Phi(0)E[h|h^T, h^L(0)] - [a(1) - a(0)] \end{aligned}$$

Adding and subtracting terms yields:

$$Bias \approx \underbrace{\Phi(1)\left(E[h|h^T, h^L(1)] - E[h|h^T, h^L(0)]\right)}_{\text{Bias due to beliefs}} + \underbrace{(\Phi(1) - \Phi(0))\,E[h|h^T, h^L(0)]}_{\text{Bias due to ``responsiveness''}}$$

Fully informative demand treatments

We now show that when demand treatments are fully informative, one can test for bias due to behavioral responsiveness.

$$\begin{aligned}
&\underbrace{[a^*(1,1) - a^*(1,0)]}_{\text{Sensitivity } (\zeta=1)} - \underbrace{[a^*(0,1) - a^*(0,0)]}_{\text{Sensitivity } (\zeta=0)} \\
&= \underbrace{[a^*(1,1) - a^*(0,1)]}_{\text{Treatment effect } (h^T=1)} - \underbrace{[a^*(1,0) - a^*(0,0)]}_{\text{Treatment effect } (h^T=-1)} \\
&\approx \left[a(1) + \Phi(1)E[h|1, h^L(1)] - a(0) - \Phi(0)E[h|1, h^L(0)]\right] \\
&- \left[a(1) + \Phi(1)E[h|-1, h^L(1)] - a(0) - \Phi(0)E[h|-1, h^L(0)]\right] \\
&= [\Phi(1) \times 1 - \Phi(0) \times 1] - [\Phi(1) \times -1 - \Phi(0) \times -1] \\
&= 2\left(\Phi(1) - \Phi(0)\right)
\end{aligned}$$

Next, we show that averaging the "positive-positive" and "negative-negative" treatment effects approximates the true treatment effect

$$\begin{aligned}
&\frac{1}{2}\left([a^*(1,1) - a^*(0,1)] + [a^*(1,-1) - a^*(0,-1)]\right) \\
\approx& a(1) - a(0) + \frac{1}{2}\left[(\Phi(1) - \Phi(0)) \times 1 + (\Phi(1) - \Phi(0)) \times -1\right] \\
=& a(1) - a(0)
\end{aligned}$$

Less informative treatments

For compactness we define notation $H^L(\zeta) \equiv h^L(\zeta)p^L(\zeta)$ and $H^T \equiv h^T p^T$. Our Bounding assumption implies $\left|H^T\right| \geq \left|H^L(\zeta)\right|$.

Consider the expressions for belief differences between treatment and control, first without (Diff^L) and then with (Diff^T) demand treatments. We have

$$\text{Diff}^L \equiv H^L(1) - H^L(0)$$

and:

$$\begin{aligned}
\text{Diff}^T &\equiv \frac{H^L(1) + H^T}{1 + H^L(1)H^T} - \frac{H^L(0) + H^T}{1 + H^L(0)H^T} \\
&= \frac{(1 - H^{T2})}{(1 + H^L(1)H^T)(1 + H^L(0)H^T)} \times \text{Diff}^L
\end{aligned}$$

We want to find conditions under which $\left|\text{Diff}^T\right| < \left|\text{Diff}^L\right|$, which holds if and only if

$$\frac{(1 - H^{T2})}{(1 + H^L(1)H^T)(1 + H^L(0)H^T)} < 1$$

rearranging we obtain:

$$0 < H^T(H^L(1) + H^L(0)) + H^{T2}(1 + H^L(1)H^L(0)) \tag{B6}$$

If $h^T = 1$, (B6) reduces to

$$-\frac{H^L(1) + H^L(0)}{1 + H^L(1)H^L(0)} < H^T$$

while if $h^T = -1$ it reduces to

$$-\frac{H^L(1) + H^L(0)}{1 + H^L(1)H^L(0)} > H^T.$$

Since the left hand side lies in the interval $(-1, 1)$ there always exists a sufficiently strong demand treatment (p^T sufficiently large) that (B6) is satisfied. We now evaluate whether there is more we can say. There are X cases to consider. Assume throughout, without loss of generality, that $\left|H^L(1)\right| > \left|H^L(0)\right|$.

1) Suppose the latent demand beliefs have the same sign as each other ($h^L(1) = h^L(0) = h^L$) and the same sign as the demand treatment ($h^L = h^T$). Then it is easy to verify that (B6) holds for all H^T. Intuitively, when the latent demand beliefs have the same sign, additional information that further reinforces those beliefs has a greater effect on the one that is less certain, reducing the gap between them.

2) The latent demand beliefs have the same sign as each other ($h^L(1) = h^L(0) = h^L$) and the opposite sign to the demand treatment ($h^L = -h^T$). Assume $h^T = 1$ and $h^L = -1$ (the opposite case is symmetric). We know that (B6) holds for sufficiently strong H^T, we will ask if our Bounding assumption is sufficient. We show that it is not, by contradiction. Suppose Bounding holds exactly, i.e. $H^T = -H^L(1)$. Then, by the premise that (B6) is satisfied:

$$-\frac{H^L(1) + H^L(0)}{1 + H^L(1)H^L(0)} < -H^L(1)$$
$$\frac{1 + \frac{H^L(0)}{H^L(1)}}{1 + H^L(1)H^L(0)} < 1$$

which holds if and only if:

$$\begin{aligned} \frac{H^L(0)}{H^L(1)} &< H^L(1)H^L(0) \\ 1 &< H^L(1)^2 \end{aligned}$$

a contradiction since $H^L(1) < 1$. Thus in this case the condition for demand treatments to reduce bias is stronger than Bounding.

3) The latent demand beliefs have opposite signs ($h^L(1) = -h^L(0)$), and the stronger belief ($H^L(1)$) has the same sign as h^T, i.e. $h^T = h^L(1)$. Focus again on the case where $h^T = 1$ (the opposite case is symmetric). We require:

$$-\frac{H^L(1) + H^L(0)}{1 + H^L(1)H^L(0)} < H^T$$

It is easy to see that the condition is always satisfied since the left-hand side is negative.

4) The latent demand beliefs have opposite signs ($h^L(1) = -h^L(0)$), and the stronger belief has the opposite sign to h^T, i.e. $h^T = -h^L(1)$. Focus again on the case where $h^T = 1$ (the opposite case is symmetric). We know that (B6) holds for sufficiently strong H^T, we will ask if our Bounding assumption is sufficient. Thus let $H^T = -H^L(1)$ (bounding holds exactly). We require:

$$\begin{aligned} -\frac{H^L(1) + H^L(0)}{1 + H^L(1)H^L(0)} &< -H^L(1) \\ \frac{1 + \frac{H^L(0)}{H^L(1)}}{1 + H^L(1)H^L(0)} &< 1 \\ \frac{H^L(0)}{H^L(1)} &< H^L(1)H^L(0) \\ 1 &> H^L(1)^2 \end{aligned}$$

which is satisfied. Thus Bounding is sufficient for (B6) to hold.

B9. Structural estimation

This section outlines step by step how the parameters are constructed in our NLLS estimation of the structural model in section III.E.

Data and parameter adjustments

First, we follow DP exactly in rounding effort scores to the nearest 100 (except for those in range $[1, 49]$ which we round to 25). This is because incentives were paid per 100 points, and we wish to avoid modeling effort choices that lie between two 100 point thresholds. We refer the reader to DP for further details.

Second, we make a couple of adjustments pre and post-estimation. First, we divide the rounded scores by 100. In other words, if effort a is measured in points, we compute $a' = a/100$ which is measured in hundreds of points. Second, we multiply the incentive, ζ, which is measured in cents per point, by 100 to express it as $\zeta' = 100\zeta$ which is measured in cents per 100 points. These transformations were helpful in achieving convergence of the estimator, which otherwise occasionally suffered from underflow problems. However they change the interpretation of the parameters. Specifically, the intrinsic motivation parameter s and the preference for pleasing the experimenter, ϕ, will both be measured in units equivalent to cents per 100 points, while the cost function parameters will be expressed for effort measured in hundreds of points.

To aid comparability with DP we therefore re-transform the parameters after estimation. DP present their estimates of incentive parameters (which in our case are s and ϕ) in the same units, cents per 100 points, so we do not need to correct them. k and γ are reported for effort measured in points, so we transform our estimates for comparability. We derive the adjustments as follows. First, for the power cost function, we have:

$$U = (s + \zeta + \phi E[h|h^T, h^L])a - \frac{ka^{1+\gamma}}{1+\gamma}$$

Let $a' = \frac{a}{100}$ and $\zeta' = 100\zeta$. Then:

$$\begin{aligned} U &= \left(s + \frac{\zeta'}{100} + \phi E[h|h^T, h^L]\right) 100a' - \frac{k(100a')^{1+\gamma}}{1+\gamma} \\ &= \left(100s + \zeta' + 100\phi E[h|h^T, h^L]\right) a' - \frac{k(100a')^{1+\gamma}}{1+\gamma} \end{aligned}$$

giving rise to first-order condition:

$$\begin{aligned} 0 &= \left(100s + \zeta' + 100\phi E[h|h^T, h^L]\right) - ka'^{\gamma} 100^{1+\gamma} \\ a' &= \left(\frac{100s + \zeta' + 100\phi E[h|h^T, h^L]}{k100^{1+\gamma}}\right)^{\frac{1}{\gamma}} \\ \log(a') &= \frac{1}{\gamma} \log\left(\frac{s^* + \zeta' + \phi^* E[h|h^T, h^L]}{k^*}\right) \end{aligned}$$

where $s^* = 100s$, $\phi^* = 100\phi$ and $k^* = 100^{1+\gamma}k$. We leave s^* and ϕ^*, (which are in equivalent units to cents per 100 points) untransformed for comparability with DP. In the tables we report $k = k^*/100^{1+\gamma}$ and its standard error, computed via the delta method.

For the exponential cost function we have:

$$\begin{aligned}U &= (s+\zeta+\phi E[h|h^T,h^L])a - \frac{k}{\gamma}\exp(\gamma a)\\ &= (s^*+\zeta'+\phi^* E[h|h^T,h^L])a' - \frac{k}{\gamma}\exp(100\gamma a')\end{aligned}$$

implying first-order condition:

$$\begin{aligned}0 &= s^*+\zeta'+\phi^* E[h|h^T,h^L] - 100k\exp(100\gamma a')\\ a' &= \frac{1}{100\gamma}\log\left(\frac{s^*+\zeta'+\phi^* E[h|h^T,h^L]}{100k}\right)\\ &= \frac{1}{\gamma^*}\log\left(\frac{s^*+\zeta'+\phi^* E[h|h^T,h^L]}{k^*}\right)\end{aligned}$$

where $s^* = 100s$, and $\phi^* = 100\phi$ as before, while $\gamma^* = 100\gamma$, $k^* = 100k$. In the tables we report $\gamma = \gamma^*/100$ and $k = k^*/100$.

Error term

To allow for the observed heterogeneity in effort, we follow DP in assuming heterogeneous effort costs, as follows. Let the cost of effort under power utility equal $ka^{1+\gamma}(1+\gamma)^{-1}\exp(-\gamma\epsilon)$ where $\epsilon \sim N(0,\sigma_\epsilon^2)$. Then our FOC becomes

$$\begin{aligned}0 &= \left(100s+\zeta'+100\phi E[h|h^T,h^L]\right) - ka'^{\gamma}100^{1+\gamma}\exp(-\gamma\epsilon)\\ a' &= \left(\frac{100s+\zeta'+100\phi E[h|h^T,h^L]}{k100^{1+\gamma}}\right)^{\frac{1}{\gamma}}\exp(\epsilon)\\ \log(a') &= \frac{1}{\gamma}\log\left(\frac{100s+\zeta'+100\phi E[h|h^T,h^L]}{k100^{1+\gamma}}\right)+\epsilon\end{aligned}$$

where ϵ becomes the error term in our NLLS routine. For the exponential cost, we follow DP and assume effort cost is $k\gamma^{-1}\exp(\gamma a)\exp(-\gamma\epsilon)$. Then our FOC becomes

$$\begin{aligned}0 &= s^*+\zeta'+\phi^* E[h|h^T,h^L] - 100k\exp(100\gamma a')\exp(-\gamma\epsilon)\\ a' &= \frac{1}{100\gamma}\log\left(\frac{s^*+\zeta'+\phi^* E[h|h^T,h^L]}{100k}\right)+\frac{\epsilon}{100}\\ &= \frac{1}{\gamma^*}\log\left(\frac{s^*+\zeta'+\phi^* E[h|h^T,h^L]}{k^*}\right)+\epsilon^*\end{aligned}$$

where $\epsilon^* = \epsilon/100$ forms the error term in our estimation.

Estimating equation

Finally, in our estimation we sometimes need to estimate the product $\phi^* E[h|h^L]$. We estimate this product directly, then transform by dividing by ϕ^*. Specifically, we estimate the following:

$$
\begin{aligned}
y_i = {} & \frac{1}{\beta_0} \log \left[\zeta_i' + \beta_1 + \beta_2(\text{pos\_demand}_i - \text{neg\_demand}_i)\right. \\
& + \beta_3 \times \text{no\_demand}_i \times \text{incentive\_0c}_i + \beta_4 \times \text{no\_demand}_i \times \text{incentive\_1c}_i \\
& \left. + \beta_5 \times \text{no\_demand}_i \times \text{incentive\_4c}_i\right] - \frac{1}{\beta_0} \log(\beta_6) + \varepsilon_i
\end{aligned}
$$

where $y = \log(a')$ or a' respectively, pos_demand, neg_demand and no_demand are dummies for our positive, negative and no demand treatments, while incentive_Xc is a dummy for the treatment with X cents per 100 points. Parameters are as follows: $\beta_0 = \gamma$ or γ^* respectively, $\beta_1 = s^*$, $\beta_2 = \phi^*$, $\beta_3 = \phi^* E[h|h^L(\zeta = 0)]$, $\beta_4 = \phi^* E[h|h^L(\zeta = 1)]$, $\beta_5 = \phi^* E[h|h^L(\zeta = 4)]$ and $\beta_6 = k^*$. We then compute the three values for $E[h|h^L]$ by dividing by β_2, i.e. β_3/β_2, β_4/β_2 and β_5/β_2. γ and k are computed by the transformations outlined above. Standard errors are computed by the delta method. In the specification where we restrict latent demand to be equal for the 1 cent and 4 cent treatments we impose $\beta_4 = \beta_5$.

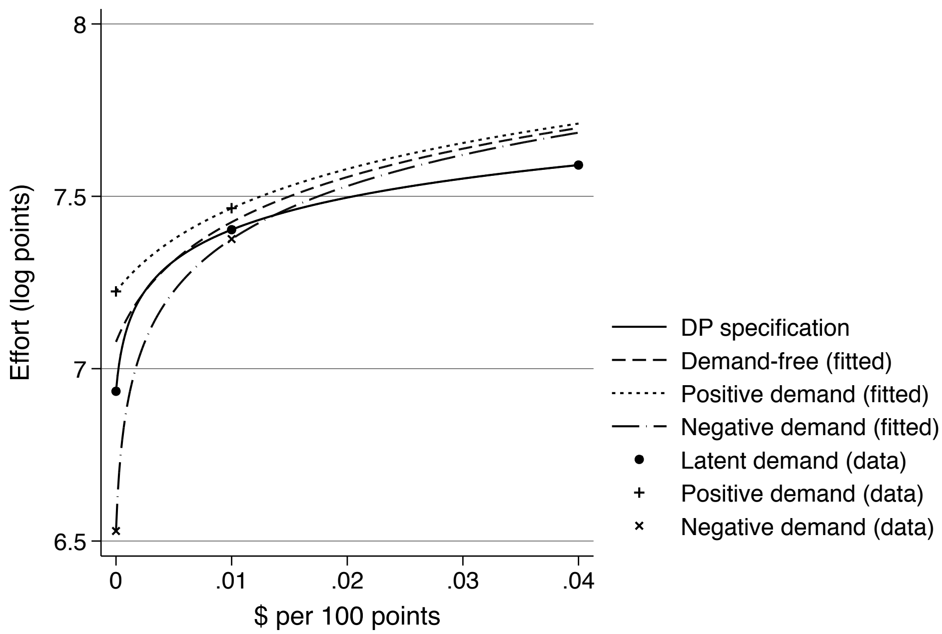
Extrapolation

Our large estimates of $h^L(4)p^L(4)$ reflect an out-of-sample extrapolation as the main model parameters are estimated from the 0 and 1 cent treatment groups. Figure B1 illustrates this for the power cost case. Points correspond to mean (log) effort for each treatment group. The figure then plots (a) predicted effort using the DP specification, which fits the "no demand" data only (parameters taken from Column 1 of table 4), and (b) predicted effort using the exactly identified model (Column 3 of table 4), for each case of zero demand $E[h] = 0$, strong positive demand ($E[h] = 1$) and strong negative demand ($E[h] = -1$). The estimation then recovers the latent demand beliefs by comparing observed effort to predicted effort when demand is zero.

It is clear from the figure that the extrapolation from model (b) to the 4 cent effort treatment is not perfect, and the observed behavior lies outside the limits implied by $E[h] \in [-1, 1]$. This is the reason for the large negative fitted value for beliefs at this point.

Another thing that is clear from the figure is how the curvature of the effort cost function determines the imputed latent beliefs, which may explain the difference in imputed beliefs between the power and exponential cost functions. The sign of imputed beliefs depends on whether the "no demand" point lies above or below the curve, so changes in curvature can flip the sign of these estimates.

Figure B1. Structural estimation: fitted values



Note: Figure displays mean (log) effort for each treatment used in the structural estimation, and fitted values from the estimated models.

B10. Using the method in practice

In this section we work through an example to illustrate the methods we have developed. We do this with the help of Figure B2, which represents the 10 data points available to a researcher who has applied no demand, weak and strong treatments to a control group (ζ_0) and a treatment group (ζ_1). To avoid confusion, we will label the actions under weak positive and negative demand as $a^+(\zeta)$ and $a^-(\zeta)$, while strong (assumed to be fully informative) positive and negative demand actions are defined as $a^{++}(\zeta)$ and $a^{--}(\zeta)$. In our example the conventional treatment effect is given by $a^L(\zeta_1) - a^L(\zeta_0)$ (or point H minus point C).

1) We can use our strong treatments to construct bounds for actions: In Figure B2 the bounds on action one are defined by points A and E for the control group, $[a^{--}(\zeta_0), a^{++}(\zeta_0)]$, and by points F and J for the treatment group, $[a^{--}(\zeta_1), a^{++}(\zeta_1)]$.

2) Similarly, we can construct bounds using the weak treatments, which are defined by points B and D for the control group, $[a^-(\zeta_0), a^+(\zeta_0)]$, and by points G and I for the treatment group, $[a^-(\zeta_1), a^+(\zeta_1)]$. Under the assumption that our demand treatments are more informative than underlying latent demand, these bounds contain the natural action.

3) We can analogously also define strong bounds for treatment effects. The upper bound is given by the comparison between respondents in the treatment group that receive strong positive demand treatments, and respondents in the control group that receive strong negative demand treatments. In Figure B2 this corresponds to the difference between points J and A: $a^{++}(\zeta_1)-a^{--}(\zeta_0)$. The lower bound is given by the comparison between respondents in the treatment group that receive strong negative demand treatments, and respondents in the control group that receive strong positive demand treatments, given by the difference points F and E: $[a^{--}(\zeta_1)-a^{++}(\zeta_0)$. The bounds are formally defined as follows: $[a^{--}(\zeta_1) - a^{++}(\zeta_0), a^{++}(\zeta_1) - a^{--}(\zeta_0)]$.

4) Similarly, we can construct weak bounds for treatment effects, by applying weak instead of strong treatments. The bounds are given by: $[a^-(\zeta_1) - a^+(\zeta_0), a^+(\zeta_1)-a^-(\zeta_0)]$. In Figure B2 the upper and lower bounds are given by the difference between points I and B, and points G and D, respectively.

5) An alternative to creating bounds, is to "control for demand effects". Under the assumption that demand treatments are fully informative, provided responsiveness to demand treatments does not differ across treatment groups, we can point identify treatment effects that are not biased by demand effects. In this specific case, we could apply strong positive or strong negative

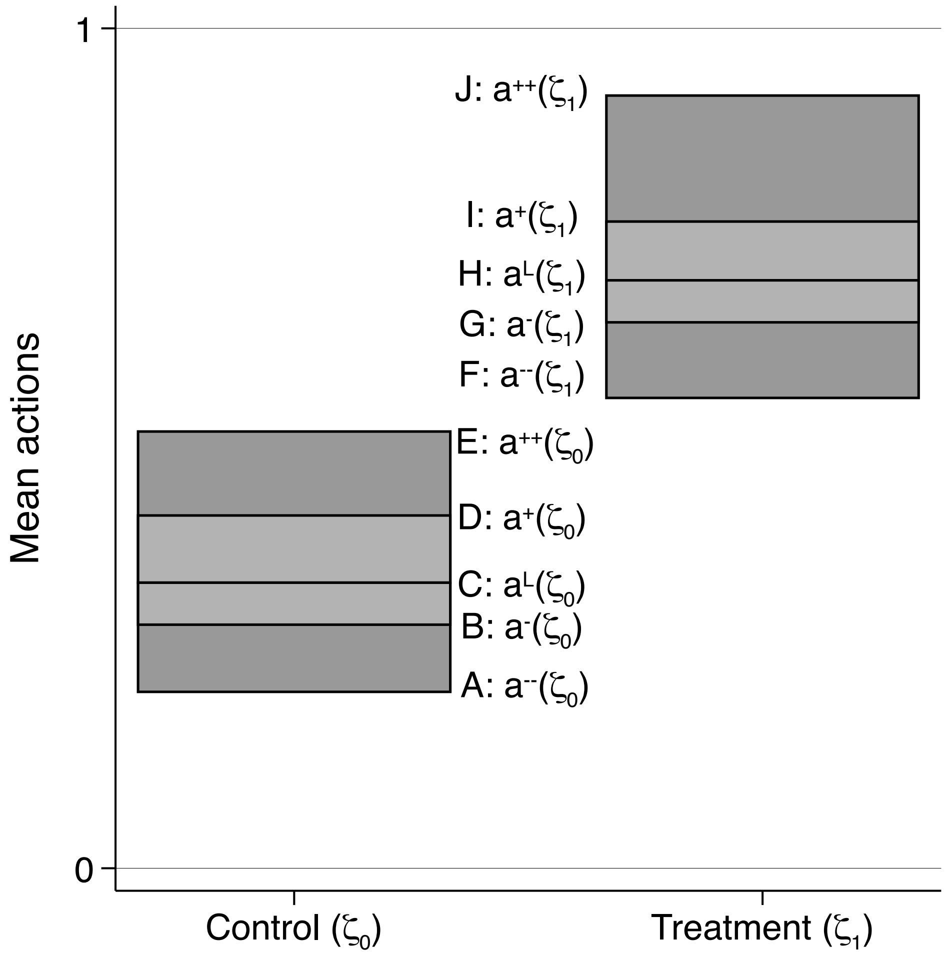
demand treatments to both the treatment and the control group: In Figure B2 the estimates are given by the difference between points J and E: $(a^{++}(\zeta_1) - a^{++}(\zeta_0))$, or F and A: $(a^{--}(\zeta_1) - a^{--}(\zeta_0))$.

6) If responsiveness to fully informative demand treatments differs significantly across treatment arms, our point estimates from employing same-signed demand treatments are still biased. However, by the symmetry of the Taylor approximation, we can approximate the treatment effect using the midpoints of the bounds generated the strong demand treatments. In B2 this corresponds to comparing the average of A and E to the average of J and F: $0.5 * [(a^{++}(\zeta_1) + a^{--}(\zeta_1)) - (a^{++}(\zeta_0) + a^{--}(\zeta_0))]$.

7) Our approach of "controlling for demand effects" can also be extended to weak treatments. In Section 3.4 we outline the conditions under which this approach reduces bias. First, we compare respondents in the treatment and control group who all receive weak positive or weak negative demand treatments. In Figure B2 the positive-positive point estimate is defined by points I and D: $(a^{+}(\zeta_1) - a^{+}(\zeta_0))$, while the negative-negative estimate is comes from points F and B: $(a^{-}(\zeta_1) - a^{-}(\zeta_0))$.

8) Finally, fully informative demand treatments can be used to eliminate nuisance parameters due to unobservable beliefs, facilitating the estimation of structural models. Structural estimation leverages points A, E, F, and J to estimate model parameters, and uses those to impute the latent demand beliefs at points C and H.

FIGURE B2. USING THE METHOD IN PRACTICE: EXAMPLE



Note: Figure displays mean actions under different treatment conditions and different demand treatments. Point A is given by respondents in the control group who receive the negative strong demand treatment: $a^{--}(\zeta_0)$; Point B is given by respondents in the control group who receive the negative weak demand treatment: $a^{-}(\zeta_0)$; Point C is defined by respondents in the control group who receive no demand treatment: $a^{L}(\zeta_0)$; Point D is given by respondents in the control group who receive the positive weak demand treatment: $a^{+}(\zeta_0)$; Point E is given by respondents in the control group who receive the positive strong demand treatment: $a^{++}(\zeta_0)$. Points F to J are defined analogously for respondents in the treatment group (ζ_1).

C. Pre-specified Tables and Figures

This section works through the pre-specified analysis for each experiment, presenting summaries of the raw data and conducting hypothesis tests.

1) Pre-analysis plan 1 described experiment 1, which was conducted on MTurk with strong demand treatments and both real and hypothetical stakes, on the dictator game, investment game and convex time budget.
2) Pre-analysis plan 2 described experiment 2, which was conducted on MTurk with weak demand treatments and both real and hypothetical stakes, on the dictator game and investment game.
3) Pre-analysis plan 3 described experiment 3, which was conducted on MTurk with strong demand treatments, real stakes and the real-effort task.
4) Pre-analysis plan 4 described experiment 4, which was conducted on the representative panel with both strong and weak demand treatments, real stakes, and the dictator game and investment game.
5) Pre-analysis plan 5 described experiments 5 and 6, which were conducted on MTurk with strong and weak demand treatments and collected data for the remaining games (experiment 6 collected real-effort data and experiment 5 collected the other games).
6) Pre-analysis plan 6 described experiment 7, which were conducted on MTurk with strong demand treatments, varied within-participant, on the dictator game and investment game.

The majority of the hypothesis tests for each pre-analysis plan are presented in a single table format (e.g. Table C3). The top half of these tables report regression coefficients and standard errors, and the bottom half reports p-values (and adjusted p-values) on the pre-specified hypothesis tests.

When conducting multiple tests within a family of hypotheses we also report false-discovery rate corrected p-values. These are used when a) testing for a positive effect (on actions or beliefs) of the positive demand treatment, negative effect of the negative demand treatment and overall effect; and b) when testing for heterogeneity across games within an experiment.

We deviate from the pre-analysis plans in two minor ways, which are inconsequential for the results.

1) As described in section II.A of the paper, we only pre-specified sample exclusions in the main real-effort experiment 3 (to match those used by DellaVigna and Pope). For consistency, we decided to apply the same restrictions to all other games. The only binding restriction was the dropping of participants who submitted multiple responses in a given experiment, amounting to less than 0.5 percent of our sample.

2) In experiments for which we collected data without demand treatments ("no demand" conditions), we pre-specified to standardize actions by the mean and standard deviation of this group. However, experiments 5 and 6 only collected positive and negative demand conditions. For consistency, therefore, we instead always standardize by the mean and standard deviation of the negative demand treatment group. This amounts to a simple linear transformation of the data.

In addition, some of the analysis in the paper was not described in the pre-analysis plans: the bounding of treatment effects, the computation of confidence intervals on the bounds, and the structural analysis.

C1. Pre-analysis Plan 1

- Table C1 and Figure C1 summarize the means, standard errors, and corresponding 95 percent confidence intervals from experiment 1 across all 18 treatment arms. Table C2 displays the game-level regressions based on the raw data showing the control mean from the "no demand condition" as well as the coefficients on the positive demand treatment indicator and the negative demand treatment indicator.

- Balance tests for this experiment are in Table D1 in Section D and indicate that there are no imbalances.

- Table C3 displays the main effects of the positive and negative demand treatment as well as heterogeneous treatment effects by gender, attention and whether choices are hypothetical or incentivized. This table also summarizes the results for the tests we had pre-specified in the pre-analysis plan.

  - Column 1 of Table C3 shows that people increase their actions in response to the positive demand treatments ($p<0.001$), decrease their actions in response to the negative demand treatments ($p<0.001$) and that the overall response to demand is non-zero ($p<0.001$). False-discovery rate corrected p-values reach the same conclusion.
  - Next, in column 2 of Table C3 we show that there is no significant treatment heterogeneity depending on whether choices are hypothetical or incentivized ($p=0.24$).
  - In column 3, we test whether there are any systematic gender differences in response to demand. Pooling across all tasks, measured sensitivity was higher for women than for men ($p=0.099$).
  - In column (4) we test whether attention moderates the response to demand treatments. We find stronger responses to the demand treatments for more attentive respondents ($p=0.102$).

  - In column (5) we examine heterogeneity across games. We find that overall sensitivity in the dictator game is significantly higher than sensitivity in the time preference measure and the risk game ($p<0.01$). We find no significant difference in sensitivity in the time preference measure and the risk game ($p=0.552$). False-discovery rate corrected p-values reach the same conclusion. An omnibus test of differences across all games highlights that responses significantly differ between games ($p<0.001$).

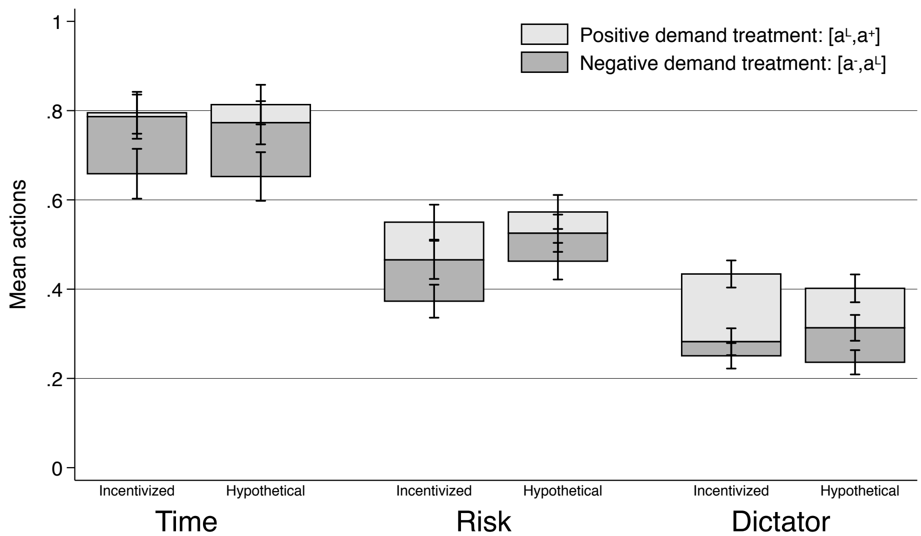
- Table C4 explores how people's beliefs about whether the experimenter wanted (column 1) or expected (column 2) a high action. Table C4 shows the results for the tests we pre-specified.
  - People in the positive demand condition are more likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).
  - People in the negative demand condition are less likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).
  - Overall, people in the positive demand condition are significantly more likely to think that the experimenter wanted a high action or expected a high action compared to people in the negative demand condition ($p<0.001$).
  - False-discovery rate corrected p-values reach the same conclusions.

Table C1—Overview of raw data: experiment 1

	Time		Risk Aversion		Dictator Game	
	Incentivized	Hypothetical	Incentivized	Hypothetical	Incentivized	Hypothetical
Unconditional Means						
Positive demand: Mean	0.795	0.813	0.550	0.573	0.434	0.402
Positive demand: SD	0.379	0.357	0.300	0.316	0.253	0.258
No demand: Mean	0.786	0.773	0.466	0.525	0.282	0.313
No demand: SD	0.386	0.392	0.340	0.335	0.246	0.230
Negative demand: Mean	0.659	0.652	0.373	0.463	0.251	0.236
Negative demand: SD	0.437	0.440	0.300	0.327	0.225	0.206
Observations	727	757	728	764	770	733

Note:
This table summarizes the raw actions from experiment 1 across all 18 treatment arms.

FIGURE C1. OVERVIEW OF RAW DATA: EXPERIMENT 1



Note: This figure summarizes the mean actions and corresponding 95 confidence intervals from experiment 1 across all 18 treatment arms

TABLE C2—GAME-LEVEL REGRESSIONS: EXPERIMENT 1

	Time	Risk Aversion	Dictator Game
Positive demand	0.025	0.067	0.121
	(0.024)	(0.021)	(0.015)
Negative demand	-0.124	-0.079	-0.054
	(0.026)	(0.021)	(0.015)
Control Mean	0.779	0.496	0.297
Observations	1484	1492	1503

Note: This table shows the effect of the positive and negative demand treatment at the game level based on the raw actions (pooling across incentivized and unincentivized choices).

Table C3—Strong Demand (Experiment 1)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.242	0.191	0.280	0.017	0.457
	(0.035)	(0.049)	(0.048)	(0.189)	(0.058)
Negative Demand	-0.247	-0.257	-0.269	-0.206	-0.203
	(0.036)	(0.051)	(0.049)	(0.175)	(0.055)
Positive demand × interactant		0.103	-0.072	0.236	
		(0.070)	(0.070)	(0.192)	
Negative demand × interactant		0.021	0.044	-0.044	
		(0.072)	(0.072)	(0.179)	
Interactant		-0.096	-0.046	-0.082	
		(0.051)	(0.051)	(0.141)	
Positive Demand × Risk					-0.258
					(0.085)
Negative Demand × Risk					-0.031
					(0.083)
Positive Demand × Time					-0.393
					(0.085)
Negative Demand × Time					-0.114
					(0.087)
Constant	-0.149	-0.101	-0.125	-0.070	-0.343
	(0.025)	(0.035)	(0.034)	(0.139)	(0.040)
Interactant		Monetary Incentive	Male	Passed attention check	
Adjusted R^2	0.041	0.041	0.041	0.041	0.052
Positive demand ≤ 0	0.000				
Adjusted p-value	0.010				
Negative demand ≥ 0	0.000				
Adjusted p-value	0.010				
Positive demand = negative demand	0.000				
Adjusted p-value	0.010				
(Positive demand - negative demand)* interaction = 0		0.240	0.099	0.102	
Risk*(pos - neg) = Time*(pos - neg)					0.552
Adjusted p-value					0.283
Risk*(positive demand - negative demand) = 0					0.006
Adjusted p-value					0.011
Time*(positive demand - negative demand) = 0					0.001
Adjusted p-value					0.006
Joint F-test					.001
Observations	4479	4479	4479	4479	4479

Note: This table summarizes the results from experiment 1. The outcome variable (action chosen) is standardized at the game level using the mean and standard deviation of the negative demand group. Robust standard errors are in parentheses. Lower section of the table reports p-values on pre-specified hypothesis tests.

Table C4—Beliefs about the experimental objective and hypothesis: Strong Demand

	Belief: Want High	Belief: Expect High
Positive - Negative	0.278	0.181
	(0.017)	(0.018)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.161	0.143
	(0.017)	(0.018)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.116	-0.038
	(0.018)	(0.018)
Adjusted p-value	[0.001]	[0.006]
Mean (No Demand)	0.543	0.451
Observations	4479	4479

Note: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action and zero if they wanted or expected a low action.

C2. Pre-analysis Plan 2

- Table C5 and Figure C2 summarize the means, standard errors, and corresponding 95 confidence intervals from experiment 2 across all 12 treatment arms. Table C6 displays the game-level regressions based on the raw data showing the control mean from the "no demand condition" as well as the coefficients on the positive demand treatment indicator and the negative demand treatment indicator.

- Balance tests for this experiment are in Table D2 in Section D. We find a slight imbalance for an indicator taking value 1 for those in part-time employment. Table C8 shows the main results controlling for this indicator.

- Table C7 displays the main effects of the positive and negative demand treatment as well as heterogeneous treatment effects by gender, attention and whether choices are hypothetical or incentivized. This table also summarizes the results of the tests we pre-specified.

  - Column (1) of Table C7 shows that people increase their actions in response to the positive demand treatments ($p<0.001$), but do not significantly decrease their actions in response to the negative demand treatments ($p=0.221$). The overall sensitivity in response to demand is non-zero ($p<0.001$). False-discovery rate corrected p-values reach the same conclusions.
  - In column (2) we show that there is no significant treatment heterogeneity depending on whether choices are hypothetical or incentivized ($p=0.313$).
  - In column (3), we find no significant treatment heterogeneity in response to demand between men and women ($p=0.252$).
  - In column (4) we test whether attention moderates the response to demand treatments. We find no significant heterogeneity by attention ($p=0.530$).
  - In column (5) we examine heterogeneity across games. We find that overall sensitivity in the dictator game was significantly higher in the risk game ($p=0.046$).

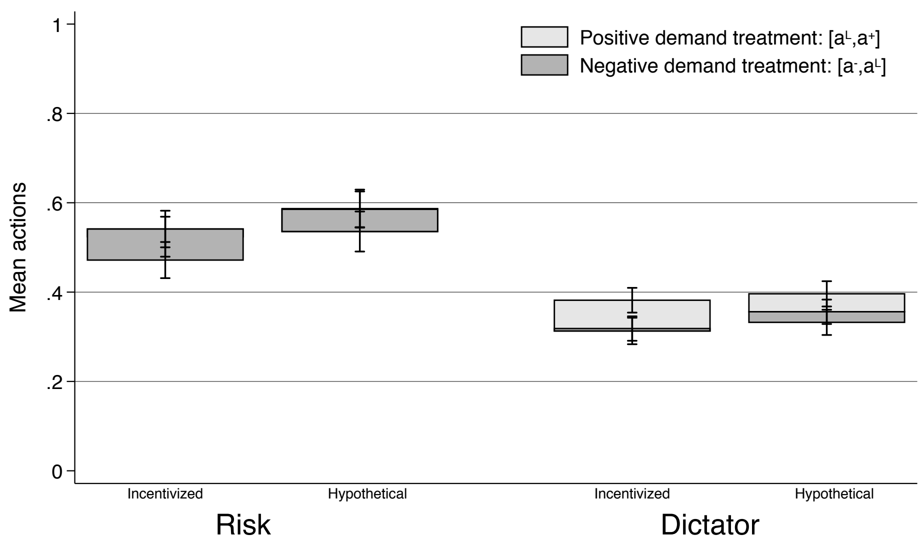
- Table C10 explores how people's beliefs about whether the experimenter wanted (column 1) or expected (column 2) a high action. Table C10 shows the results for the tests we pre-specified.

  - People in the positive demand condition are more likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).

  - People in the negative demand condition are less likely to think that the experimenter wanted a high action (p<0.001) and that the experimenter expected a high action (p<0.001).
  - Overall, people in the positive demand condition are significantly more likely to think that the experimenter wanted a high action or expected a high action compared to people in the negative demand condition (p<0.001).
  - False-discovery rate corrected p-values reach the same conclusions.

- In Table C11 we report results confirming that people in the incentive condition are more likely to believe that the task involved real money (p<0.001).
- In Table C9 we test for differences between strong and weak demand pooling data from experiment 1 and 2. We find that the overall sensitivity to strong demand is significantly higher pooling across games ((p<0.001), as seen in column (1)), for the dictator game ((p<0.001), as seen in column (2)), and for the risk game ((p<0.001), as seen in column (3)). False-discovery rate corrected p-values reach the same conclusions.
- Table C12 shows that there was no differential attrition across treatment arms.

Figure C2. Overview of raw data: Experiment 2



Note: This figure summarizes the mean actions and corresponding 95 confidence intervals from experiment 2 across all 12 treatment arms

Table C5—Overview of raw data: experiment 2

	Risk Aversion		Dictator Game	
	Incentivized	Hypothetical	Incentivized	Hypothetical
Unconditional Means				
Positive demand: Mean	0.524	0.587	0.382	0.396
Positive demand: SD	0.348	0.335	0.222	0.224
No demand: Mean	0.541	0.536	0.313	0.356
No demand: SD	0.339	0.350	0.234	0.215
Negative demand: Mean	0.472	0.585	0.318	0.332
Negative demand: SD	0.317	0.325	0.226	0.219
Observations	739	734	758	719

Note: This table summarizes the raw action data from experiment 2 across all 12 treatment arms.

Table C6—Game-level regressions: experiment 2

	Risk Aversion	Dictator Game
Positive demand	0.017	0.055
	(0.022)	(0.014)
Negative demand	-0.008	-0.009
	(0.021)	(0.014)
Control Mean	0.539	0.334
Observations	1473	1477

Note: This table shows the effect of the positive and negative demand treatment at the game level based on the raw action data (pooling across incentivized and unincentivized choices).

Table C7—Weak Demand (Experiment 2)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.127	0.153	0.085	0.067	0.206
	(0.043)	(0.060)	(0.056)	(0.116)	(0.054)
Negative Demand	-0.032	0.037	-0.029	-0.023	-0.036
	(0.042)	(0.060)	(0.055)	(0.109)	(0.054)
Pos. demand × interactant		-0.054	0.090	0.070	-0.155
		(0.085)	(0.086)	(0.124)	(0.085)
Neg. demand × interactant		-0.138	-0.006	-0.010	0.012
		(0.084)	(0.085)	(0.118)	(0.083)
Interactant		-0.066	-0.032	-0.217	0.192
		(0.060)	(0.061)	(0.081)	(0.060)
Interactant		Monetary Incentive	Male	Passed attention check	Risk
Adjusted R-squared	0.005	0.009	0.004	0.009	0.011
Pos. demand ≤ 0	0.001				
Adjusted p-value	0.010				
Neg. demand ≥ 0	0.221				
Adjusted p-value	0.070				
Pos. demand = neg. demand	0.000				
Adjusted p-value	0.010				
(Pos. - neg.) × interactant = 0		0.313	0.252	0.530	0.046
Observations	2950	2950	2950	2950	2950

Note: This table summarizes the results from experiment 2. The outcome variable (action chosen) is standardized at the game level using the mean and standard deviation of the negative demand group. Robust standard errors are in parentheses. Lower section of the table reports p-values on pre-specified hypothesis tests.

Table C8—Weak Demand (Experiment 2): controlling for imbalances

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.128	0.154	0.086	0.069	0.208
	(0.043)	(0.060)	(0.056)	(0.115)	(0.054)
Negative Demand	-0.032	0.037	-0.029	-0.023	-0.036
	(0.042)	(0.060)	(0.055)	(0.109)	(0.054)
Pos. demand × interactant		-0.054	0.089	0.069	-0.155
		(0.085)	(0.086)	(0.124)	(0.085)
Neg. demand × interactant		-0.139	-0.006	-0.010	0.012
		(0.084)	(0.085)	(0.118)	(0.083)
Interactant		-0.066	-0.030	-0.217	0.193
		(0.060)	(0.061)	(0.081)	(0.060)
Interactant		Monetary Incentive	Male	Passed attention check	Risk
Adjusted R-squared	0.004	0.009	0.004	0.009	0.011
Pos. demand ≤ 0	0.001				
Adjusted p-value	0.010				
Neg. demand ≥ 0	0.222				
Adjusted p-value	0.070				
Pos. demand = neg. demand	0.000				
Adjusted p-value	0.010				
(Pos. - neg.) × interactant = 0		0.307	0.255	0.538	0.045
Observations	2950	2950	2950	2950	2950

Note: This table summarizes the results from experiment 2. The outcome variable (action chosen) is standardized at the game-level using the mean and standard deviation of the negative demand group. Robust standard errors are in parentheses. Here we control for an indicator taking value 1 for those in part-time employment due to imbalance on this variable. Lower section of the table reports p-values on pre-specified hypothesis tests.

Table C9—Comparing experiments 1 and 2

	(1)	(2)	(3)
Positive Demand=1	0.127	0.206	0.051
	(0.043)	(0.054)	(0.065)
Experiment 1=1	-0.137	-0.140	-0.128
	(0.043)	(0.056)	(0.064)
Positive Demand=1 $\times$ Experiment 1=1	0.203	0.251	0.148
	(0.060)	(0.080)	(0.090)
Negative Demand=1	-0.032	-0.036	-0.024
	(0.042)	(0.054)	(0.063)
Negative Demand=1 $\times$ Experiment 1=1	-0.182	-0.167	-0.211
	(0.059)	(0.077)	(0.088)
Constant	-0.105	-0.203	-0.011
	(0.030)	(0.039)	(0.046)
Sample	All	Dictator Game	Investment
Adjusted R^2	0.034	0.056	0.021
H_0 : (Positive Demand - Negative Demand)*Interaction = 0	0.000	0.000	0.000
Adjusted p-value	0.001	0.001	0.001
Observations	5945	2980	2965

Note: This table uses action data from the investment game and dictator game from experiments 1 (strong demand treatments) and 2 (weak demand treatments), standardized at the game-experiment level using the mean and standard deviation of the negative demand treatment group. The dummy experiment 1 takes value 1 for respondents from experiment 1. Column (1) pools the data from both games, column (2) uses dictator game data and column (3) investment game data. Adjusted p-values are corrected for false-discovery rate across the three tests.

Table C10—Beliefs about the experimental objective and hypothesis: Weak Demand (Experiment 2)

	Belief: Want High	Belief: Expect High
Positive - Negative	0.332	0.402
	(0.021)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.171	0.217
	(0.022)	(0.022)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.161	-0.185
	(0.022)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.485	0.392
Observations	2950	2950

Note: This table uses data from all respondents who completed experiment 2. The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action and zero if they wanted or expected a low action.

Table C11—Beliefs about whether the experiment is incentivized

	(1) Belief: Real Money
Monetary Incentive	0.367
	(0.016)
Control Mean	0.139
R^2	0.153
Observations	2950

Note: This table uses data from all respondents who completed experiment 2. The outcome variable takes value one if the respondent believes that the tasks in the experiment involve real money and value zero otherwise. Monetary incentive takes value 1 for respondents whose choices were incentivized, and takes value 0 for respondents whose choices were hypothetical.

Table C12—Attrition across treatment arms

	(1) Finished
Positive Demand	0.00601
	(0.003)
Negative Demand	0.00104
	(0.003)
Mean (no demand)	0.990
R^2	0.00145
Observations	2964

Note: This table uses data from all respondents who started experiment 2. Finished takes value one for all respondents who completed the experiment.

C3. Pre-analysis Plan 3

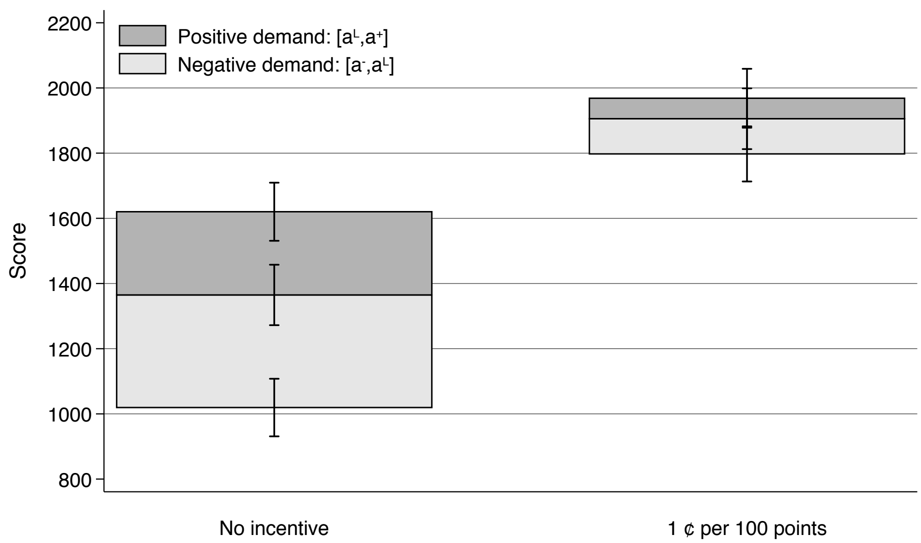
- Balance tests for this experiment are in Table D3 in Section D. We found a slight imbalance on an indicator for Hispanic race. Table C15 shows our main results controlling for this variable.

- Table C13 and Figure C3 summarizes the means, standard errors, and corresponding 95 percent confidence intervals from experiment 3 across all 6 treatment arms (excluding the 4 cent treatment that was used only for structural estimation).

- Table C14 displays the main effects of the positive and negative demand treatment as well as heterogeneous treatment effects by gender, and whether people are paid a 1-cent bonus or no bonus. This table summarizes the results for the main pre-specified tests.
  - Column (1) shows that people increase their effort in response to the positive demand treatments ($p<0.001$), decrease their effort in response to the negative demand treatments ($p<0.001$). Moreover, the overall sensitivity in response to demand is non-zero ($p<0.001$). False-discovery rate corrected p-values reach the same conclusions.
  - In column (2), we test for systematic differences by incentive level in response to the demand treatments. Sensitivity was higher in the no-incentive condition compared to the 1-cent incentive condition ($p<0.001$).
  - In column (3), we test for gender differences in response to the demand treatments. Sensitivity was not significantly different for women than for men ($p=0.946$).

- Table C16 explores beliefs about whether the experimenter wanted (column 1) or expected (column 2) a high action.
  - People in the positive demand condition are more likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).
  - People in the negative demand condition are less likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).
  - Overall, people in the positive demand condition are significantly more likely to think that the experimenter wanted a high action or expected a high action compared to people in the negative demand condition ($p<0.001$).
- False-discovery rate corrected p-values reach the same conclusions.
- In Table C17, we show that there is no differential attrition across treatment arms.

FIGURE C3. OVERVIEW OF RAW DATA: EXPERIMENT 3



Note: This figure summarizes the mean actions (expressed in points scored) and corresponding 95 confidence intervals from experiment 3 across all 6 treatment arms.

Table C13—Overview of raw data: experiment 3

	Effort	
	1-cent bonus	No bonus
Unconditional Means		
Positive demand: Mean	0.492	0.405
Positive demand: SD	0.179	0.177
No demand: Mean	0.476	0.341
No demand: SD	0.184	0.182
Negative demand: Mean	0.449	0.255
Negative demand: SD	0.162	0.176
Observations	714	731

Note: This table summarizes the raw action data from experiment 3 across all 6 treatment arms.

Table C14—Effort (z-scored) with strong demand

	(1)	(2)	(3)
Positive Demand	0.209	0.333	0.313
	(0.061)	(0.085)	(0.107)
Negative Demand	-0.305	-0.450	-0.198
	(0.061)	(0.085)	(0.103)
Positive demand × interactant		-0.249	-0.182
		(0.123)	(0.129)
Negative demand × interactant		0.305	-0.191
		(0.121)	(0.126)
Interactant		0.082	0.136
		(0.088)	(0.093)
Constant	0.068	0.027	-0.009
	(0.044)	(0.061)	(0.078)
Interactant		1-cent incentive	Male
Adjusted R^2	0.046	0.061	0.046
Positive demand ≤ 0	0.000		
Adjusted p-value	0.010		
Negative demand ≥ 0	0.000		
Adjusted p-value	0.010		
Positive demand = negative demand	0.000		
Adjusted p-value	0.010		
(Positive demand - negative demand)* interaction = 0		0.000	0.946
Observations	1445	1445	1445

Note: This table summarizes the results from experiment 3. The outcome variable (action chosen) is standardized at the incentive treatment level using the mean and standard deviation of the negative demand group. Lower section of the table reports p-values on pre-specified hypothesis tests.

TABLE C15—EFFORT (Z-SCORED) WITH STRONG DEMAND: WITH CONTROL FOR IMBALANCE

	(1)	(2)	(3)
Positive Demand	0.221	0.344	0.319
	(0.061)	(0.085)	(0.107)
Negative Demand	-0.299	-0.444	-0.200
	(0.061)	(0.085)	(0.103)
Positive demand × interactant		-0.248	-0.173
		(0.122)	(0.128)
Negative demand × interactant		0.305	-0.177
		(0.120)	(0.126)
Interactant		0.081	0.127
		(0.088)	(0.092)
Constant	0.050	0.009	-0.022
	(0.045)	(0.061)	(0.078)
Interactant		1-cent incentive	Male
Adjusted R^2	0.049	0.064	0.049
Positive demand ≤ 0	0.000		
Adjusted p-value	0.010		
Negative demand ≥ 0	0.000		
Adjusted p-value	0.010		
Positive demand = negative demand	0.000		
Adjusted p-value	0.010		
(Positive demand - negative demand)* interaction = 0		0.000	0.974
Observations	1445	1445	1445

Note: This table summarizes the results from experiment 3. The outcome variable (action chosen) is standardized at the incentive treatment level using the mean and standard deviation of the negative demand group. Here we control for an indicator taking value 1 for Hispanics as we found an imbalance for this variable across demand treatment arms. Lower section of the table reports p-values on pre-specified hypothesis tests.

Table C16—Beliefs: Effort with strong demand

	Belief: Want High	Belief: Expect High
Positive - Negative	0.459	0.414
	(0.027)	(0.028)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.170	0.190
	(0.026)	(0.028)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.289	-0.224
	(0.031)	(0.031)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.688	0.640
Observations	1445	1445

Note: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action and zero if they wanted or expected a low action.

Table C17—Attrition across treatment arms

	(1) Finished
Positive Demand	-0.000449
	(0.009)
Negative Demand	0.00679
	(0.010)
Mean (no demand)	0.990
R^2	0.000366
Observations	1739

Note: Finished takes value one for all respondents who completed the experiment.

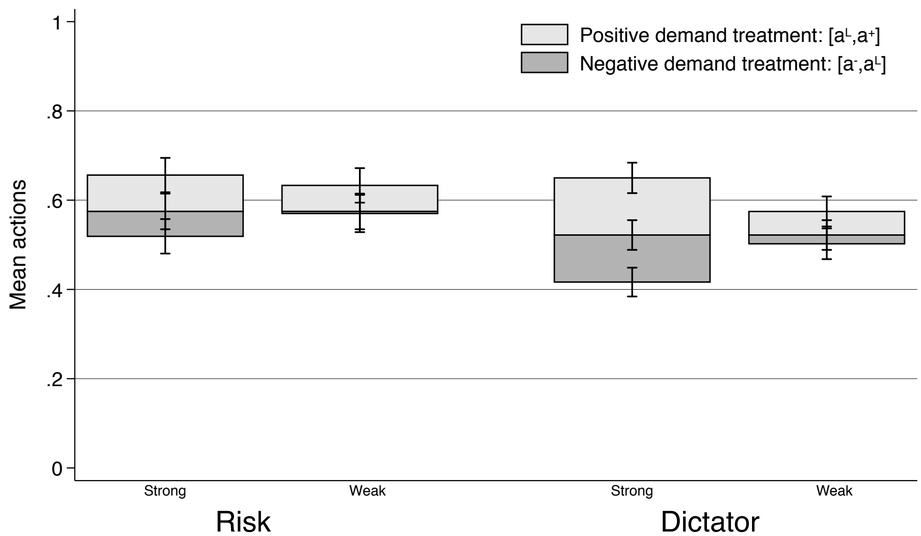
C4. Pre-analysis Plan 4

- Balance tests for this experiment are in Table D4 in Section D and indicate that there are no imbalances.
- Table C18 and Figure C4 summarize the means, standard errors, and corresponding 95 percent confidence intervals from experiment 4 across all treatment arms. Table C19 displays the game-level regressions based on the raw data showing the control mean from the "no demand condition" as well as the coefficients on the positive demand treatment indicator and the negative demand treatment indicator.
- Table C20 displays the main effects of the positive and negative demand treatment as well as heterogeneous treatment effects by strong vs. weak demand treatment, gender, attention and game. This table also summarizes the results for the main pre-specified tests.
  - Column (1) shows that people increase their actions in response to the positive demand treatments ($p<0.001$), decrease their actions in response to the negative demand treatments ($p<0.001$). Moreover, the overall sensitivity in response to demand is non-zero ($p<0.001$). False-discovery rate corrected p-values reach the same conclusions.
  - Pooling across games, column (2) finds that sensitivity was significantly higher in the strong treatments than the weak treatments ($p<0.001$).
  - Pooling across games and demand treatments, column (3) finds sensitivity was significantly higher for women than for men ($p=0.014$).
  - Pooling across games and demand treatments, column (4) finds that sensitivity was significantly higher for attentive respondents than for inattentive respondents ($p<0.001$).
  - Pooling across demand treatments, column (5) finds that sensitivity in the dictator game was significantly higher than sensitivity in the risk game ($p=0.001$).
- Table C22 explores how people's beliefs about whether the experimenter wanted (column 1) or expected (column 2) a high action. Table C22 shows the results for the tests we pre-specified.
  - People in the positive demand condition are more likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).
  - People in the negative demand condition are less likely to think that the experimenter wanted a high action ($p<0.001$) and that the experimenter expected a high action ($p<0.001$).

  - Overall, people in the positive demand condition are significantly more likely to think that the experimenter wanted a high action ($p<0.001$) or expected a high action compared to people in the negative demand condition ($p<0.001$).
  - False-discovery rate corrected p-values reach the same conclusions.

- Table C21 examines demand sensitivity by population and shows that there were no systematic differences ($p=0.602$) when pooling across games.
- Table C23 shows that there was no differential attrition across treatment arms.

Figure C4. Overview of raw data: Experiment 4



Note: This figure summarizes the means and corresponding 95 confidence intervals from experiment 4 across all treatment arms.

Table C18—Overview of raw data: experiment 4

	Risk Aversion		Dictator Game	
	Strong	Weak	Strong	Weak
Unconditional Means				
Positive demand: Mean	0.656	0.633	0.650	0.575
Positive demand: SD	0.341	0.337	0.300	0.292
No demand: Mean	0.575	0.575	0.522	0.522
No demand: SD	0.358	0.358	0.289	0.289
Negative demand: Mean	0.519	0.570	0.416	0.502
Negative demand: SD	0.331	0.351	0.286	0.291
Observations	900	880	896	862

Note: This table summarizes the raw data from experiment 4 across all treatment arms.

Table C19—Game-level regressions: experiment 4

	Risk Aversion	Dictator Game
Positive demand	0.070	0.091
	(0.025)	(0.021)
Negative demand	-0.031	-0.065
	(0.025)	(0.021)
Control Mean	0.575	0.522
Observations	1468	1465

Note: This table shows the effect of the positive and negative demand treatment at the game level based on the raw data (pooling across strong and weak demand treatments).

Table C20—Representative Sample with strong and weak demand treatments (Experiment 4)

	(1)	(2)	(3)	(4)	(5)
Positive Demand	0.281	0.193	0.322	0.244	0.555
	(0.055)	(0.063)	(0.063)	(0.061)	(0.064)
Negative Demand	-0.159	-0.037	-0.224	-0.083	-0.033
	(0.055)	(0.064)	(0.061)	(0.061)	(0.064)
Pos. demand × interactant		0.175	-0.084	0.113	-0.545
		(0.064)	(0.064)	(0.064)	(0.062)
Neg. demand × interactant		-0.237	0.139	-0.222	-0.257
		(0.063)	(0.064)	(0.063)	(0.063)
Interactant		Strong demand treatment	Male	Passed attention check	Risk
Adjusted R-squared	0.031	0.038	0.033	0.035	0.060
Pos. demand ≤ 0	0.000				
Adjusted p-value	0.010				
Neg. demand ≥ 0	0.002				
Adjusted p-value	0.010				
Pos. = neg. demand	0.000				
Adjusted p-value	0.010				
(Pos. - neg.) × interactant = 0		0.000	0.014	0.000	0.001
Observations	2933	2933	2933	2933	2933

Note: This table summarizes the results from experiment 4. The outcome variable (action chosen) is standardized at the game level using the mean and standard deviation of the negative demand group. Lower section of the table reports p-values on pre-specified hypothesis tests.

Table C21—Comparing Representative and MTurk Samples

	(1)	(2)	(3)
Positive Demand=1	0.261	0.423	0.095
	(0.043)	(0.057)	(0.064)
Representative Sample=1	0.522	0.851	0.208
	(0.054)	(0.076)	(0.075)
Positive Demand=1 × Representative Sample=1	0.021	-0.079	0.114
	(0.070)	(0.097)	(0.097)
Negative Demand=1	-0.148	-0.042	-0.251
	(0.042)	(0.056)	(0.061)
Negative Demand=1 × Representative Sample=1	-0.011	-0.202	0.160
	(0.069)	(0.096)	(0.096)
Constant	-0.226	-0.344	-0.111
	(0.031)	(0.041)	(0.045)
Sample	All	Dictator Game	Investment
Adjusted R^2	0.093	0.165	0.041
H_0 : (Positive Demand - Negative Demand)*Repres. Sample = 0	0.602	0.149	0.592
Adjusted p-value	0.813	0.813	0.813
Observations	5928	2993	2935

Note: This table uses data from the incentivized MTurk respondents from experiments 1 and 2 and the representative online panel (experiment 4). Representative Sample is a dummy variable taking value 1 for respondents from the representative online panel and value zero for the MTurk respondents. Lower section of the table reports p-values on pre-specified hypothesis tests.

TABLE C22—BELIEFS ABOUT THE EXPERIMENTAL OBJECTIVE AND HYPOTHESIS: REPRESENTATIVE SAMPLE

	Belief: Want High	Belief: Expect High
Positive - Negative	0.206	0.205
	(0.020)	(0.020)
Adjusted p-value	[0.001]	[0.001]
Positive - Neutral	0.068	0.091
	(0.024)	(0.025)
Adjusted p-value	[0.001]	[0.001]
Negative - Neutral	-0.138	-0.114
	(0.025)	(0.025)
Adjusted p-value	[0.001]	[0.001]
Mean (No Demand)	0.602	0.511
Observations	2933	2933

Note: The outcome variables take value one if the respondents believed that the experimenter wanted (column 1) or expected (column 2) a high action and zero if they wanted or expected a low action.

TABLE C23—ATTRITION ACROSS TREATMENT ARMS: EXPERIMENT 4

	(1) Finished
Positive Demand	0.000710
	(0.004)
Negative Demand	-0.000400
	(0.004)
Mean (no demand)	0.990
R^2	0.0000390
Observations	2952

Note: This table uses data from all respondents who started experiment 4. Finished takes value one for all respondents who completed the experiment.

C5. Pre-analysis Plan 5

This plan encompasses experiments 5 and 6 and pre-specified the collecting of all incentivized MTurk data together by demand treatment type, to present results in single tables and figures.

- Balance tests for the experiments are Tables D5 and D6 in Section D and indicate that there are no imbalances.
- Figure 2 and Tables 2 and 1 (included in the paper) summarize the raw data and sensitivities across games.
- Next, we consider whether sensitivity to weak demand treatments differs across games. In Table C24 we show little evidence of statistically significant differences in sensitivity across all games ($p=0.241$ with effort tasks included, $p=0.437$ when excluded).
- Table C24 shows that there are large differences in sensitivity across games in response to strong demand both when all 11 games are considered and when the effort tasks are excluded ($p<0.001$).
- In Table C25 we conduct a pooled test with all MTurk experiments examining whether sensitivity varies between strong and weak demand treatments. We find a larger response to strong compared to weak demand treatments ($p<0.001$).

Table C24—Differences in response to demand across games

	(1)	(2)
Positive Demand=1	1.058	0.289
	(0.133)	(0.125)
Ambiguity	0.149	0.007
	(0.110)	(0.102)
DG	-0.153	-0.248
	(0.103)	(0.089)
Effort: incentive	0.332	0.085
	(0.104)	(0.100)
Effort: no incentive	-0.056	0.049
	(0.105)	(0.101)
Lying	0.241	0.015
	(0.123)	(0.102)
Risk	-0.138	-0.195
	(0.103)	(0.095)
Time	0.100	0.021
	(0.113)	(0.099)
Trust	0.124	0.015
	(0.109)	(0.105)
UG 1	0.137	0.015
	(0.118)	(0.104)
UG 2	0.230	0.015
	(0.118)	(0.100)
Positive Demand=1 × Ambiguity	-0.596	-0.116
	(0.165)	(0.161)
Positive Demand=1 × DG	-0.364	-0.049
	(0.156)	(0.146)
Positive Demand=1 × Effort: incentive	-0.829	-0.211
	(0.158)	(0.156)
Positive Demand=1 × Effort: no incentive	-0.275	-0.352
	(0.157)	(0.161)
Positive Demand=1 × Lying	-0.454	-0.247
	(0.178)	(0.161)
Positive Demand=1 × Risk	-0.530	-0.133
	(0.156)	(0.155)
Positive Demand=1 × Time	-0.709	-0.277
	(0.164)	(0.158)
Positive Demand=1 × Trust	-0.495	-0.213
	(0.165)	(0.163)
Positive Demand=1 × UG 1	-0.374	-0.132
	(0.172)	(0.168)
Positive Demand=1 × UG 2	-0.308	-0.008
	(0.173)	(0.161)
Constant	-0.367	-0.015
	(0.087)	(0.072)
Treatment	Strong	Weak
Adjusted R^2	0.102	0.012
P-value(Omnibus F-Test)	0.000	0.241
Adjusted p-values	0.001	0.191
P-value(Omnibus F-Test): without effort tasks	0.001	0.437
Adjusted p-values	0.001	0.279
Observations	4800	4450

Note: Outcome variable (action chosen) is standardized at the game level. We pool all real stakes MTurk observations across all experiments. Column (1) presents results from the strong demand treatments and column 2 presents results from the weak demand treatments.

Table C25—Differences in response to strong vs. weak demand treatments

	(1) Z-scored behavior
Strong × Positive Demand	0.471
	(0.042)
Positive demand	0.133
	(0.030)
R^2	0.0455
Observations	9250

Note: Outcome variable (action chosen) is standardized at the game level. We pool all real stakes MTurk observations across all experiments.

C6. Pre-analysis Plan 6

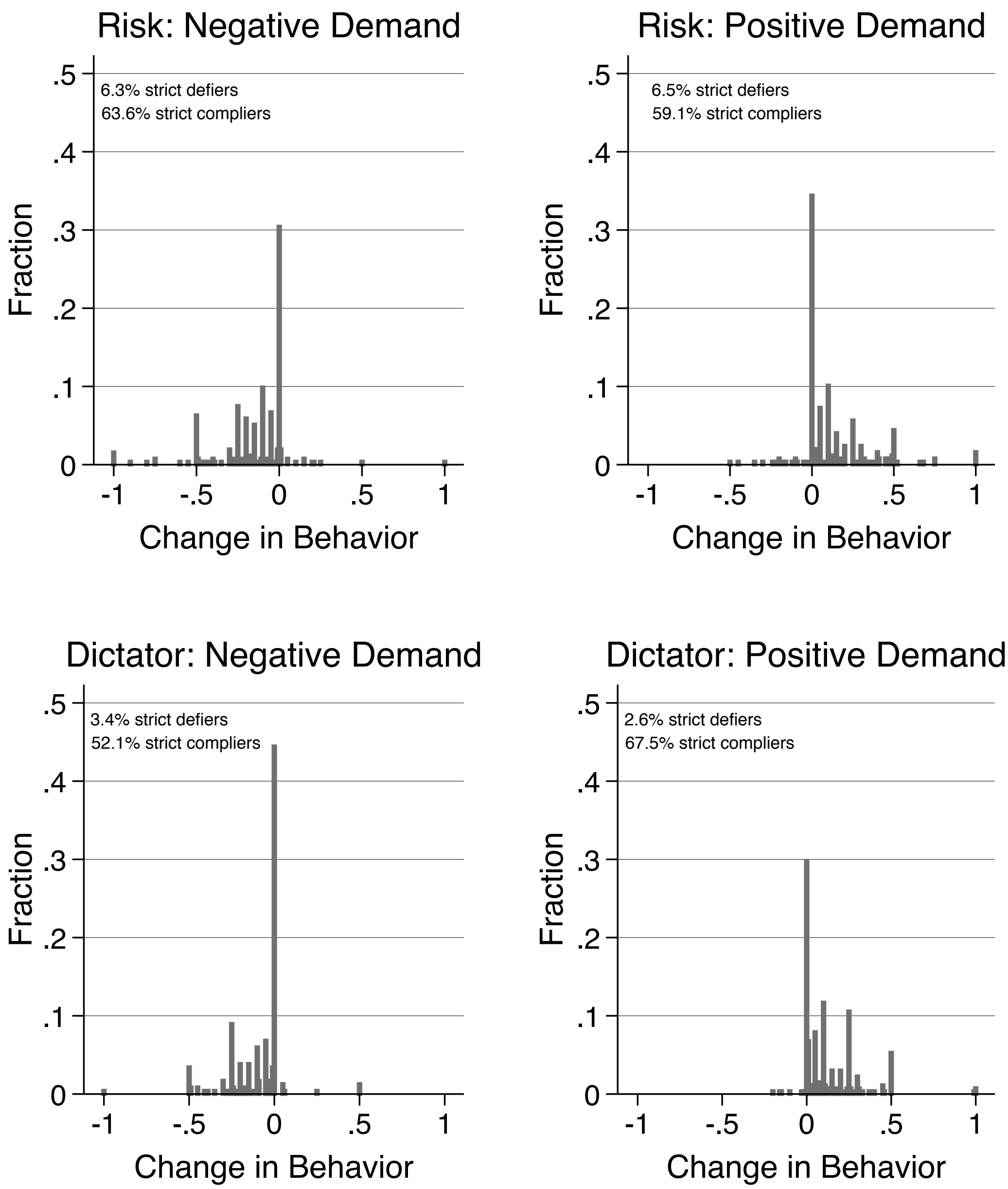
- Balance tests for this experiment are in Table D7 in Section D. We found a slight imbalance for men and people in part-time employment. Therefore, in Table C26 we show the main results controlling for an indicator taking value 1 for men and an indicator taking value 1 for people in part-time employment.
- Figures 3 (included in the paper) and C5 present the raw data graphically. Figure 3 plots task 2 actions against task 1 actions – points above the 45 degree line correspond to increases in actions. Figure C5 plots the distributions of changes of actions between task 1 and task 2 (positive difference means action increased in task 2).
- Table A2 (included in the paper) Panel A summarizes behavior in tasks 1 and 2. The relevant columns are headed "Within," Task 1 choices are labeled "no demand" and Task 2 choices are either labeled Positive or Negative demand. Panel B of Table A2 shows the sensitivities computed from raw or standardized (at the game level) task 2 actions.
- Table A2 compares sensitivity estimates and raw choices from the "within" experiment (experiment 7) and the incentivized MTurk "between experiment" with strong demand treatments (experiment 1). We do not find statistically significant differences in sensitivity.
- Table A5 (in the main web Appendix) documents for each game and demand treatment the number of strict compliers and strict defiers. Defiance rates were very low at around 5 percent.
- Table A5 also displays the average change in action between tasks 1 and 2 for each treatment arm and for compliers and defiers separately.

Table C26—Within Design (Experiment 7): with controls for imbalance

	Dictator			Risk		
	Within	Between	Difference	Within	Between	Difference
Panel A: Unconditional Means						
Positive demand	0.384	0.434	-0.045	0.560	0.550	0.010
	(0.017)	(0.015)	(0.023)	(0.021)	(0.020)	(0.029)
No demand	0.273	0.282	-0.005	0.448	0.466	-0.023
	(0.011)	(0.015)	(0.019)	(0.015)	(0.022)	(0.026)
Negative demand	0.195	0.251	-0.044	0.318	0.373	-0.058
	(0.014)	(0.014)	(0.021)	(0.019)	(0.019)	(0.027)
Panel B: Sensitivity (positive - negative)						
Raw data	0.182	0.190	-0.007	0.244	0.177	0.068
	(0.023)	(0.021)	(0.031)	(0.029)	(0.027)	(0.040)
Z-score	0.763	0.745	-0.005	0.715	0.520	0.197
	(0.095)	(0.086)	(0.126)	(0.084)	(0.080)	(0.117)
Panel C: Monotonicity						
Positive - Neutral (z-score)	0.514	0.617	-0.110	0.377	0.248	0.137
	(0.044)	(0.088)	(0.128)	(0.041)	(0.087)	(0.124)
Negative - Neutral (z-score)	-0.380	-0.128	-0.252	-0.427	-0.272	-0.155
	(0.045)	(0.086)	(0.122)	(0.042)	(0.084)	(0.119)
Observations	499	770	1269	500	728	1228

Note: This table uses data from the within design (experiment 7) and incentivized choices from the dictator game and the investment game in experiment 1. These experiments employ strong demand treatments. Here we control for an indicator taking value 1 for men and another indicator value taking value 1 for people in part-time employment. as we had found an imbalance for this variable across demand treatment arms

Figure C5. Distribution of Responses: Within Design



Note: This figure uses MTurk data from experiment 7 and displays the distribution of changes in behavior (in task 2 compared to task 1) to our strong demand treatments.

D. Balance tables, summary statistics, and attrition

Table D1—Balance Table: Experiment 1 (Strong Demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.511	0.519	0.497	0.639	0.442	0.215	4479
Income	51560.364	52459.736	53429.878	0.387	0.068	0.346	3995
Age	36.226	36.464	36.414	0.560	0.653	0.904	4479
Household Size	3.711	3.649	3.626	0.225	0.097	0.640	4479
White	0.773	0.785	0.774	0.429	0.974	0.451	4479
Black	0.070	0.067	0.071	0.676	0.927	0.612	4479
Hispanic	0.053	0.057	0.055	0.592	0.824	0.757	4479
Asian	0.079	0.063	0.075	0.091	0.703	0.194	4479
Full-time employment	0.484	0.508	0.522	0.194	0.039	0.430	4479
Part-time employment	0.128	0.120	0.115	0.541	0.280	0.632	4479
Unemployed	0.144	0.133	0.130	0.409	0.269	0.770	4479
Bachelor Degree	0.353	0.369	0.389	0.360	0.044	0.264	4479
Conservative	0.230	0.238	0.242	0.641	0.459	0.778	4441
Number of HITs	9393.289	9217.178	8651.406	0.762	0.202	0.321	4479
Joint							

Note: In this table we present evidence on the experimental integrity in experiment 1. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.9110. The p-value of the joint F-test when comparing covariates in the positive and no-demand demand condition is 0.6965. The p-value of the joint F-test when comparing covariates in the negative and no-demand demand condition is 0.2402.

Table D2—Balance Table: Experiment 2 (Weak Demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.466	0.467	0.477	0.949	0.630	0.678	2950
Income	51017.241	51410.405	51949.829	0.757	0.455	0.664	2612
Age	35.909	35.871	35.227	0.940	0.171	0.197	2950
Household Size	3.696	3.686	3.757	0.880	0.346	0.275	2950
White	0.785	0.761	0.749	0.194	0.060	0.568	2950
Black	0.069	0.076	0.076	0.525	0.517	0.994	2950
Hispanic	0.054	0.051	0.057	0.756	0.723	0.507	2950
Asian	0.066	0.070	0.087	0.705	0.083	0.178	2950
Full-time employment	0.493	0.464	0.467	0.199	0.236	0.915	2950
Part-time employment	0.130	0.099	0.125	0.033	0.742	0.071	2950
Unemployed	0.101	0.139	0.129	0.010	0.056	0.493	2950
Bachelor Degree	0.367	0.353	0.376	0.524	0.662	0.283	2950
Conservative	0.273	0.254	0.243	0.342	0.131	0.583	2927
Number of HITs	5854.863	5642.157	5306.841	0.703	0.314	0.529	2950

Note: In this table we present evidence on the experimental integrity in experiment 2. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.7084. The p-value of the joint F-test when comparing covariates in the positive and no-demand demand condition is 0.2332. The p-value of the joint F-test when comparing covariates in the negative and no-demand demand condition is 0.4838.

Table D3—Balance Table: Experiment 3 (Effort Experiment with strong demand)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.558	0.574	0.536	0.598	0.439	0.235	1691
Income	33596.974	32213.115	32457.983	0.170	0.265	0.822	1691
Age	37.444	37.359	36.618	0.906	0.251	0.352	1691
Household Size	3.750	3.783	3.771	0.690	0.790	0.894	1691
White	0.752	0.783	0.761	0.217	0.749	0.411	1691
Black	0.109	0.084	0.084	0.149	0.152	0.999	1691
Hispanic	0.055	0.025	0.046	0.006	0.493	0.070	1691
Asian	0.065	0.072	0.074	0.634	0.555	0.914	1691
Full-time employment	0.509	0.498	0.536	0.707	0.364	0.241	1691
Part-time employment	0.125	0.125	0.107	0.993	0.337	0.387	1691
Unemployed	0.105	0.121	0.107	0.380	0.886	0.502	1691
Bachelor Degree	0.395	0.355	0.370	0.155	0.382	0.623	1691
Republican	0.250	0.289	0.273	0.139	0.382	0.585	1691

Note: In this table we present evidence on the integrity of the randomization in experiment 3. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.9171. The p-value of the joint F-test when comparing covariates in the positive and no-demand demand condition is 0.1012. The p-value of the joint F-test when comparing covariates in the negative and no-demand demand condition is 0.4845.

Table D4—Balance Table: Experiment 4 (Representative Sample)

	No demand	Pos. demand	Neg. demand	P-value(Pos. demand - no demand)	P-value(Neg. demand - no demand)	P-value(Pos. demand - neg. demand)	Observations
Male	0.488	0.485	0.468	0.912	0.446	0.428	2933
Income	68357.264	65309.037	67175.470	0.253	0.662	0.398	2882
Age	47.972	46.899	47.879	0.195	0.911	0.147	2933
Household Size	3.332	3.312	3.333	0.752	0.983	0.692	2926
White	0.801	0.772	0.784	0.159	0.399	0.505	2927
Black	0.070	0.069	0.062	0.968	0.518	0.457	2927
Hispanic	0.051	0.064	0.062	0.269	0.379	0.799	2927
Asian	0.043	0.061	0.062	0.104	0.079	0.866	2927
Full-time employment	0.499	0.485	0.496	0.566	0.888	0.604	2933
Part-time employment	0.074	0.078	0.091	0.768	0.218	0.263	2933
Unemployed	0.068	0.050	0.052	0.132	0.188	0.818	2933
Bachelor Degree	0.331	0.352	0.330	0.368	0.975	0.262	2933
Conservative	0.350	0.352	0.351	0.921	0.962	0.951	2797

Note: In this table we present evidence on the integrity of the randomization in experiment 4. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.7455. The p-value of the joint F-test when comparing covariates in the positive and no-demand demand condition is 0.4909. The p-value of the joint F-test when comparing covariates in the negative and no-demand demand condition is 0.6390.

Table D5—Balance Table: Experiment 5 (Many Task experiment)

	Pos. demand	Neg. demand	P-value(Pos. demand - neg. demand)	Observations
Male	0.453	0.472	0.168	5045
Income	53324.385	52746.322	0.460	4478
Age	37.328	37.207	0.714	5045
Household Size	3.711	3.654	0.159	5045
White	0.770	0.776	0.620	5045
Black	0.078	0.072	0.434	5045
Hispanic	0.048	0.048	0.956	5045
Asian	0.075	0.078	0.773	5045
Full-time employment	0.513	0.516	0.819	5045
Part-time employment	0.115	0.113	0.830	5045
Unemployed	0.126	0.140	0.147	5045
Bachelor Degree	0.376	0.372	0.745	5045
Conservative	0.263	0.257	0.636	5019
Number of HITs	9381.041	8544.300	0.055	5045

Note: In this table we present evidence on the integrity of the randomization in experiment 5. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.1990.

Table D6—Balance Table: Experiment 6 (Effort Experiment with weak demand treatments)

	Pos. demand	Neg. demand	P-value(Pos. demand - neg. demand)	Observations
Male	0.547	0.556	0.803	769
Income	32317.708	32532.468	0.861	769
Age	37.339	37.545	0.807	769
Household Size	3.732	3.681	0.626	769
White	0.755	0.730	0.422	769
Black	0.083	0.083	0.991	769
Hispanic	0.055	0.073	0.306	769
Asian	0.081	0.075	0.780	769
Full-time employment	0.552	0.527	0.491	769
Part-time employment	0.128	0.094	0.132	769
Unemployed	0.125	0.122	0.902	769
Bachelor Degree	0.432	0.379	0.134	769
Conservative	0.266	0.325	0.078	764

Note: In this table we present evidence on the integrity of the randomization in experiment 6. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.2562.

Table D7—Balance Table: Experiment 7 (Within Design)

	Pos. demand	Neg. demand	P-value(Pos. demand - neg. demand)	Observations
Male	0.545	0.610	0.038	999
Income	53645.374	54549.763	0.604	876
Age	34.465	34.439	0.970	999
Household Size	3.533	3.540	0.938	999
White	0.732	0.743	0.696	999
Black	0.078	0.078	0.995	999
Hispanic	0.064	0.049	0.300	999
Asian	0.090	0.109	0.317	999
Full-time employment	0.520	0.569	0.118	999
Part-time employment	0.133	0.092	0.043	999
Unemployed	0.137	0.125	0.592	999
Bachelor Degree	0.408	0.386	0.475	999
Conservative	0.241	0.233	0.776	994

Note: In this table we present evidence on balance for experiment 7. The p-value of the joint F-test when comparing covariates in the positive and negative demand condition is 0.043.

Table D8—Summary Statistics: Pooled across all experiments

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.50	0.50	0.00	0.00	1.00	18866
Income	52105.70	32773.39	45000.00	5000.00	225000.00	17303
Age	38.25	13.02	35.00	17.00	116.00	18866
Household Size	3.63	1.40	3.00	2.00	13.00	18859
White	0.77	0.42	1.00	0.00	1.00	18860
Black	0.07	0.26	0.00	0.00	1.00	18860
Hispanic	0.05	0.22	0.00	0.00	1.00	18860
Asian	0.07	0.26	0.00	0.00	1.00	18860
Full-time employment	0.50	0.50	1.00	0.00	1.00	18866
Part-time employment	0.11	0.32	0.00	0.00	1.00	18866
Unemployed	0.12	0.32	0.00	0.00	1.00	18866
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	18866
Conservative	0.27	0.44	0.00	0.00	1.00	16942
Number of HITs	8215.31	14921.14	2500.00	750.00	75000.00	12474

Note: This table summarizes the main covariates of all respondents across all 6 experiments.

Table D9—Summary Statistics: Experiment 1 (Strong demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.51	0.50	1.00	0.00	1.00	4479
Income	52481.85	26617.99	55000.00	5000.00	100000.00	3995
Age	36.37	11.26	33.00	19.00	88.00	4479
Household Size	3.66	1.40	3.00	2.00	11.00	4479
White	0.78	0.42	1.00	0.00	1.00	4479
Black	0.07	0.25	0.00	0.00	1.00	4479
Hispanic	0.06	0.23	0.00	0.00	1.00	4479
Asian	0.07	0.26	0.00	0.00	1.00	4479
Full-time employment	0.50	0.50	1.00	0.00	1.00	4479
Part-time employment	0.12	0.33	0.00	0.00	1.00	4479
Unemployed	0.14	0.34	0.00	0.00	1.00	4479
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	4479
Conservative	0.24	0.43	0.00	0.00	1.00	4441
Number of HITs	9091.59	15766.32	2500.00	750.00	75000.00	4479

Note: This table summarizes the main covariates of all respondents in experiment 1.

Table D10—Summary Statistics: Experiment 2 (Weak demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.47	0.50	0.00	0.00	1.00	2950
Income	51460.57	26145.92	55000.00	5000.00	100000.00	2612
Age	35.67	11.09	33.00	19.00	81.00	2950
Household Size	3.71	1.43	3.00	2.00	13.00	2950
White	0.77	0.42	1.00	0.00	1.00	2950
Black	0.07	0.26	0.00	0.00	1.00	2950
Hispanic	0.05	0.23	0.00	0.00	1.00	2950
Asian	0.07	0.26	0.00	0.00	1.00	2950
Full-time employment	0.47	0.50	0.00	0.00	1.00	2950
Part-time employment	0.12	0.32	0.00	0.00	1.00	2950
Unemployed	0.12	0.33	0.00	0.00	1.00	2950
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	2950
Conservative	0.26	0.44	0.00	0.00	1.00	2927
Number of HITs	5600.34	12081.52	1500.00	750.00	75000.00	2950

Note: This table summarizes the main covariates of all respondents in experiment 2.

Table D11—Summary Statistics: Experiment 3 (Effort Experiment: Strong demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.56	0.50	1.00	0.00	1.00	1691
Income	32877.00	17304.76	35000.00	5000.00	85000.00	1691
Age	37.19	12.32	36.00	21.00	70.00	1691
Household Size	3.77	1.39	4.00	2.00	12.00	1691
White	0.76	0.43	1.00	0.00	1.00	1691
Black	0.09	0.29	0.00	0.00	1.00	1691
Hispanic	0.04	0.20	0.00	0.00	1.00	1691
Asian	0.07	0.25	0.00	0.00	1.00	1691
Full-time employment	0.51	0.50	1.00	0.00	1.00	1691
Part-time employment	0.12	0.33	0.00	0.00	1.00	1691
Unemployed	0.11	0.31	0.00	0.00	1.00	1691
Bachelor Degree	0.38	0.48	0.00	0.00	1.00	1691
Republican	0.27	0.44	0.00	0.00	1.00	1691

Note: This table summarizes the main covariates of all respondents in experiment 3.

Table D12—Summary Statistics: Experiment 4 (Representative sample)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.48	0.50	0.00	0.00	1.00	2933
Income	66658.57	52862.72	62500.00	7500.00	225000.00	2882
Age	47.50	16.39	47.00	17.00	116.00	2933
Household Size	3.32	1.26	3.00	2.00	13.00	2926
White	0.78	0.41	1.00	0.00	1.00	2927
Black	0.07	0.25	0.00	0.00	1.00	2927
Hispanic	0.06	0.24	0.00	0.00	1.00	2927
Asian	0.06	0.23	0.00	0.00	1.00	2927
Full-time employment	0.49	0.50	0.00	0.00	1.00	2933
Part-time employment	0.08	0.28	0.00	0.00	1.00	2933
Unemployed	0.05	0.23	0.00	0.00	1.00	2933
Bachelor Degree	0.34	0.47	0.00	0.00	1.00	2933
Conservative	0.35	0.48	0.00	0.00	1.00	2797

Note: This table summarizes the main covariates of all respondents in experiment 4.

Table D13—Summary Statistics: Experiment 5 (Many task experiment)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.46	0.50	0.00	0.00	1.00	5045
Income	53034.84	26174.26	55000.00	5000.00	100000.00	4478
Age	37.27	11.72	34.00	17.00	88.00	5045
Household Size	3.68	1.44	3.00	2.00	13.00	5045
White	0.77	0.42	1.00	0.00	1.00	5045
Black	0.07	0.26	0.00	0.00	1.00	5045
Hispanic	0.05	0.21	0.00	0.00	1.00	5045
Asian	0.08	0.27	0.00	0.00	1.00	5045
Full-time employment	0.51	0.50	1.00	0.00	1.00	5045
Part-time employment	0.11	0.32	0.00	0.00	1.00	5045
Unemployed	0.13	0.34	0.00	0.00	1.00	5045
Bachelor Degree	0.37	0.48	0.00	0.00	1.00	5045
Conservative	0.26	0.44	0.00	0.00	1.00	5019
Number of HITs	8966.40	15468.91	2500.00	750.00	75000.00	5045

Note: This table summarizes the main covariates of all respondents in experiment 5.

Table D14—Summary Statistics: Experiment 6 (Effort Experiment: Weak demand)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.55	0.50	1.00	0.00	1.00	769
Income	32425.23	16975.09	35000.00	5000.00	85000.00	769
Age	37.44	11.73	35.00	21.00	70.00	769
Household Size	3.71	1.46	3.00	2.00	10.00	769
White	0.74	0.44	1.00	0.00	1.00	769
Black	0.08	0.28	0.00	0.00	1.00	769
Hispanic	0.06	0.24	0.00	0.00	1.00	769
Asian	0.08	0.27	0.00	0.00	1.00	769
Full-time employment	0.54	0.50	1.00	0.00	1.00	769
Part-time employment	0.11	0.31	0.00	0.00	1.00	769
Unemployed	0.12	0.33	0.00	0.00	1.00	769
Bachelor Degree	0.41	0.49	0.00	0.00	1.00	769
Conservative	0.30	0.46	0.00	0.00	1.00	764

Note: This table summarizes the main covariates of all respondents in experiment 6.

Table D15—Summary Statistics: Experiment 7 (Within Design)

	Mean	SD	Median	Min.	Max.	Obs.
Male	0.58	0.49	1.00	0.00	1.00	999
Income	54081.05	25778.96	55000.00	5000.00	100000.00	876
Age	34.45	10.73	31.00	19.00	83.00	999
Household Size	3.54	1.39	3.00	2.00	13.00	999
White	0.74	0.44	1.00	0.00	1.00	999
Black	0.08	0.27	0.00	0.00	1.00	999
Hispanic	0.06	0.23	0.00	0.00	1.00	999
Asian	0.10	0.30	0.00	0.00	1.00	999
Full-time employment	0.54	0.50	1.00	0.00	1.00	999
Part-time employment	0.11	0.32	0.00	0.00	1.00	999
Unemployed	0.13	0.34	0.00	0.00	1.00	999
Bachelor Degree	0.40	0.49	0.00	0.00	1.00	999
Conservative	0.24	0.43	0.00	0.00	1.00	994

Note: This table summarizes the main covariates of all respondents in experiment 7.

Table D16—Attrition overview by task in the strong demand experiments

	Finished: Time	Finished: Risk	Finished: Ambiguity Aversion	Finished: Effort 0 cent bonus	Finished: Effort 1 cent bonus	Finished: Lying	Finished: Dictator Game	Finished: Ult. Game 1	Finished: Ult. Game 2	Finished: Trust Game 1	Finished: Trust Game 2
Panel A: Unconditional Means											
Positive demand	1.000	1.000	0.995	0.972	0.968	0.995	1.000	1.000	0.995	0.990	1.000
	(0.000)	(0.000)	(0.005)	(0.010)	(0.011)	(0.005)	(0.000)	(0.000)	(0.005)	(0.007)	(0.000)
No demand	0.996	1.000		0.941	0.980		1.000				
	(0.004)	(0.000)		(0.015)	(0.009)		(0.000)				
Negative demand	0.992	0.996	1.000	0.984	0.970	1.000	0.996	1.000	0.990	1.000	1.000
	(0.006)	(0.004)	(0.000)	(0.008)	(0.011)	(0.000)	(0.004)	(0.000)	(0.007)	(0.000)	(0.000)
Panel B: Differential attrition											
Positive - Negative	0.008	0.004	-0.005	-0.012	-0.002	-0.005	0.004		0.005	-0.010	0.000
	(0.006)	(0.004)	(0.005)	(0.013)	(0.016)	(0.005)	(0.004)		(0.008)	(0.007)	(0.000)
Positive - Neutral	0.004			0.031	-0.012						
	(0.004)			(0.018)	(0.014)						
Negative - Neutral	-0.004	-0.004		0.043	-0.009		-0.004				
	(0.007)	(0.004)		(0.017)	(0.014)		(0.004)				
Observations	730	729	405	757	734	366	771	409	424	384	371

Note: In Panel A we present the proportion of respondents who completed the experiment in the positive, negative and no-demand treatment arms respectively. In Panel B we assess whether there was differential attrition across treatment arms by examining differences in completion rates across demand treatment arms.

Table D17—Attrition overview by task in the weak demand experiments

	Finished: Time	Finished: Risk	Finished: Ambiguity Aversion	Finished: Effort 0 cent bonus	Finished: Effort 1 cent bonus	Finished: Lying	Finished: Dictator Game	Finished: Ult. Game 1	Finished: Ult. Game 2	Finished: Trust Game 1	Finished: Trust Game 2
Panel A: Unconditional Means											
Positive demand	0.995	1.000	0.990	0.955	0.941	0.995	0.996	1.000	1.000	1.000	1.000
	(0.005)	(0.000)	(0.007)	(0.015)	(0.017)	(0.005)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
No demand		0.993					0.992				
		(0.005)					(0.006)				
Negative demand	0.991	0.992	0.995	0.965	0.960	1.000	0.993	0.995	1.000	1.000	0.985
	(0.007)	(0.006)	(0.005)	(0.013)	(0.014)	(0.000)	(0.005)	(0.005)	(0.000)	(0.000)	(0.009)
Panel B: Differential attrition											
Positive - Negative	0.004	0.008	-0.005	-0.010	-0.019	-0.005	0.004	0.005		0.000	0.015
	(0.008)	(0.006)	(0.009)	(0.019)	(0.022)	(0.005)	(0.007)	(0.005)		(0.000)	(0.009)
Positive - Neutral		0.007					0.004				
		(0.005)					(0.007)				
Negative - Neutral		-0.001					0.001				
		(0.008)					(0.008)				
Observations	425	743	393	404	401	413	763	361	411	352	349

Note: In Panel A we present the proportion of respondents who completed the experiment in the positive, negative and no-demand treatment arms respectively. In Panel B we assess whether there was differential attrition across treatment arms by examining differences in completion rates across demand treatment arms.

E. Citations for Experimental Tasks

Our respondents complete one of the following tasks: a dictator game (Kahneman, Knetsch and Thaler, 1986); a risky investment game (Gneezy and Potters, 1997), without or with ambiguity; a convex time budget task (Andreoni and Sprenger, 2012); a trust game (first or second mover, Berg, Dickhaut and McCabe, 1995); an ultimatum game (first or second mover, Güth, Schmittberger and Schwarze, 1982); a lying game (Fischbacher and Föllmi-Heusi, 2013); and a real effort task with or without performance pay (DellaVigna and Pope, 2017, DellaVigna and Pope, 2018).

*

REFERENCES

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *The American Economic Review*, 102(7): 3333–56.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10(1): 122–142.

DellaVigna, Stefano, and Devin Pope. 2017. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy, forthcoming.*

DellaVigna, Stefano, and Devin Pope. 2018. "What Motivates Effort? Evidence and Expert Forecasts." *Review of Economic Studies*, 85(2): 1029–1069.

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2013. "Lies in Disguise—an Experimental Study on Cheating." *Journal of the European Economic Association*, 11(3): 525–547.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics*, 631–645.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization*, 3(4): 367–388.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler. 1986. "Fairness and the Assumptions of Economics." *Journal of Business*, S285–S300.

Additional citation

Stoye, Jörg. 2009. "More on Confidence Intervals for Partially Identified Parameters." *Econometrica*, 77(4): 1299–1315.