

Online Appendix for
 “Testing the Waters: Behavior across Participant Pools”
 by Erik Snowberg and Leeat Yariv

A Additional Tables and Figures

Table A.1: Demographic Attributes of MTurk Compared to a Representative Sample

	Representative	MTurk
Age		
18–25	16%	23%
26–54	53%	70%
55–64	18%	6%
65+	13%	1%
Race / Ethnicity		
White	71%	74%
Black	12%	8%
Hispanic	8%	6%
Asian	5%	7%
Education		
High School or Less	20%	10%
Some College	23%	30%
Associates Degree	10%	11%
Bachelors Degree	31%	38%
Post Graduate Degree	16%	12%
Employment Status		
Employed	54%	67%
Unemployed	8%	10%
Out of Labor Force	14%	11%
Online Worker	6%	10%
Retired	18%	2%
Income		
Less than \$20K	17%	32%
Between \$20K and \$30K	14%	16%
Between \$30K and \$50K	19%	23%
Between \$50K and \$70K	19%	13%
Between \$70K and \$150K	25%	14%
More than \$150K	6%	2%
Marital Status		
Single	32%	50%
Partnered	53%	42%
Seperated / Divorced / Widowed	14%	9%
N	1,000	995

Table A.2: Attributes of Selected Prior Studies

	Study	Tasks	Incentives?	Design	Participant Pool	N	Results
Sample Comparisons	Arechar, Gchter, and Molleman (2018)	Public Goods	Yes	Within Participant	MTurk; Harvard, and Yale	320	Similar behavior across samples, students somewhat less cooperative
	Berinsky, Huber, and Lenz (2012)	Demographics; Heuristics and Biases	No	N/A	MTurk	3,240	MTurk fairly close to representative, experiments by and large replicate
	Falk, Meier, and Zehnder (2013)	Trust Game	Yes	N/A	Zurich, U. of Zurich	1,296	Similar behavior across samples, students somewhat less generous
	Horton, Rand, and Zeckhauser (2011)	PD, Priming, and Framing	Yes	N/A	MTurk and Harvard	567	Three experimental treatments replicate on Mturk
	Paolacci, Chandler, and Ipeirotis (2010)	Heuristics and Biases	No	N/A	MTurk, Midwestern Students, Others	409	MTurk fairly close to representative, experiments by and large replicate
Selection into Lab	Cleave, Nikiforakis, and Slonim (2013)	Trust Game and Risk Elicitation	Yes	Within Participant	Int. Micro. Students at U. of Melbourne	1,173	Little selection in lab participation
	Harrison, Lau, and Rustm (2009)	Risk Elicitation	Yes	N/A	Denmark	253	Little selection in lab participation, show-up fees → more risk-averse
	Falk, Meier, and Zehnder (2013)	Charitable Giving	Yes	N/A	U. of Zurich	16,666	Participation in lab experiments not associated with donation patterns
Observer Effect	Bolton, Katok, and Zwick (1998)	Dictator Giving	Yes	Between Participant	Penn State	110	Introducing anonymity in Lab does not alter dictator giving
	Hoffman et al. (1994)	Dictator Giving	Yes	Between Participant	U. of Arizona	269	Double-blind dictator games yield more selfish behavior
	Laury, Walker, and Williams (1995)	Public Goods	Yes	Between Participant	Indiana U. Students	64	Double-blind public good games yield similar deviations from NE
	Current Study	Risk, Cognitive, Dictator Giving, PD, Competition, etc.	Yes	Within and Between Participant	Caltech and UBC Students, MTurk, United States	4,280	Differences in response levels, Limited selection into lab, No observer effect.

Table A.3: Percent of Variance due to Noise in Different Samples.

	Spring 2015 CCS				SSEL Participant (In Lab)
	Everyone	Participant	Weighted Participant	SSEL Participant	
Risky Projects	43% (2.9%)	48% (4.6%)	47% (6.6%)	41% (8.3%)	45% (8.6%)
Risky Urns	25% (2.3%)	21% (3.3%)	19% (6.5%)	24% (6.7%)	52% (9.0%)
Dictator Giving	15% (1.8%)	15% (2.8%)	14% (4.4%)	18% (5.9%)	23% (6.5%)
IAT Race	42% (2.8%)	41% (4.3%)	46% (12.4%)	30% (7.3%)	40% (8.3%)
IAT Gender	39% (2.8%)	40% (4.3%)	37% (10.7%)	54% (9.2%)	58% (9.4%)
N	819	350	350	96	96

Figure A.1: Distribution of Responses in Representative Sample, MTurk, CCS, and UBC

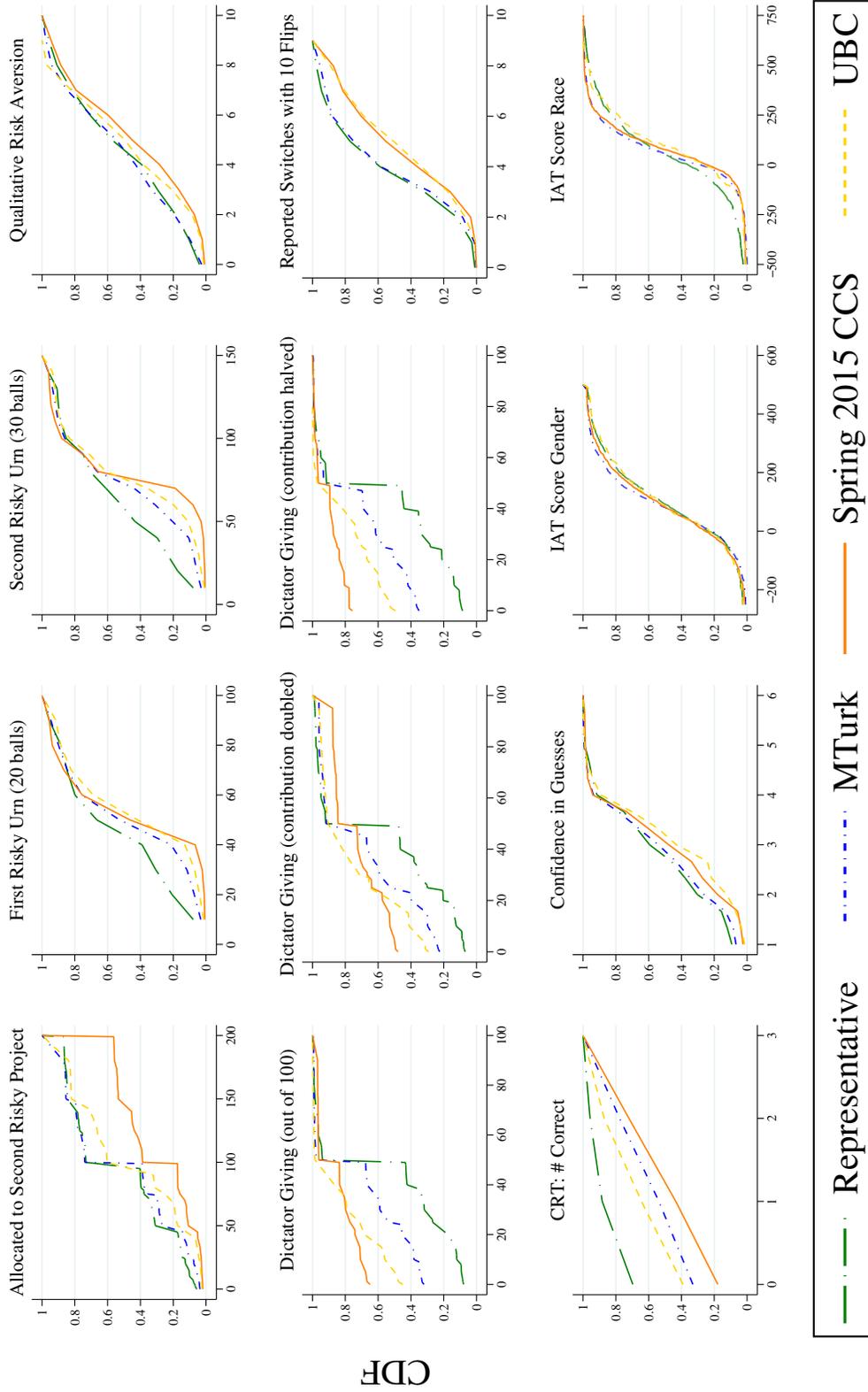


Figure A.2: Correlations across the Representative Sample, MTurk, and CCS (5% level).

	Risk Aversion	Discounting (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		0+-	+0+	0	-0-	00-	-	-	0	0	-
Discounting (δ)	0+-		0-0	0	0+0	+	0-0	0	0	00+	+00
Dictator	+0+	0-0		-	-	--0	0+0	00-	0	0	0-0
Prisoner's Dilemma	0	0	-		0++	+	00+	0++	0	0	00+
Lying	-0-	0+0	-	0++		0++	+0+	0++	0	0	0++
Cognitive	00-	+	--0	+	0++		--+	00+	00-	0	+
Confidence	-	0-0	0+0	00+	+0+	--+		+	0	00+	0++
Compete	-	0	00-	0++	0++	00+	+		0+0	0++	0++
IAT Race	0	0	0	0	0	00-	0	0+0		+	0
IAT Gender	0	00+	0	0	0	0	00+	00+	+		00+
Male	-	+00	0-0	00+	0++	+	0++	0++	0	00+	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.

Figure A.3: Correlations across the Representative Sample, MTurk, and CCS (1% level).

	Risk Aversion	Discounting (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		0+-	0	0	-0-	00-	-	0--	0	0	0--
Discounting (δ)	0+-		0-0	0	0	+	0	0	0	0	+00
Dictator	0	0-0		-	-	0-0	0+0	00-	0	0	0-0
Prisoner's Dilemma	0	0	-		0++	++0	0	0	0	0	00+
Lying	-0-	0	-	0++		0+0	+0+	0++	0	0	0++
Cognitive	00-	+	0-0	++0	0+0		0-0	00+	0	0	+
Confidence	-	0	0+0	0	+0+	0-0		0++	0	00+	0++
Compete	0--	0	00-	0	0++	00+	0++		0	00+	0++
IAT Race	0	0	0	0	0	0	0	0		+	0
IAT Gender	0	0	0	0	0	0	00+	00+	+		00+
Male	0--	+00	0-0	00+	0++	+	0++	0++	0	00+	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.

Figure A.4: Distribution of Responses in the CCS

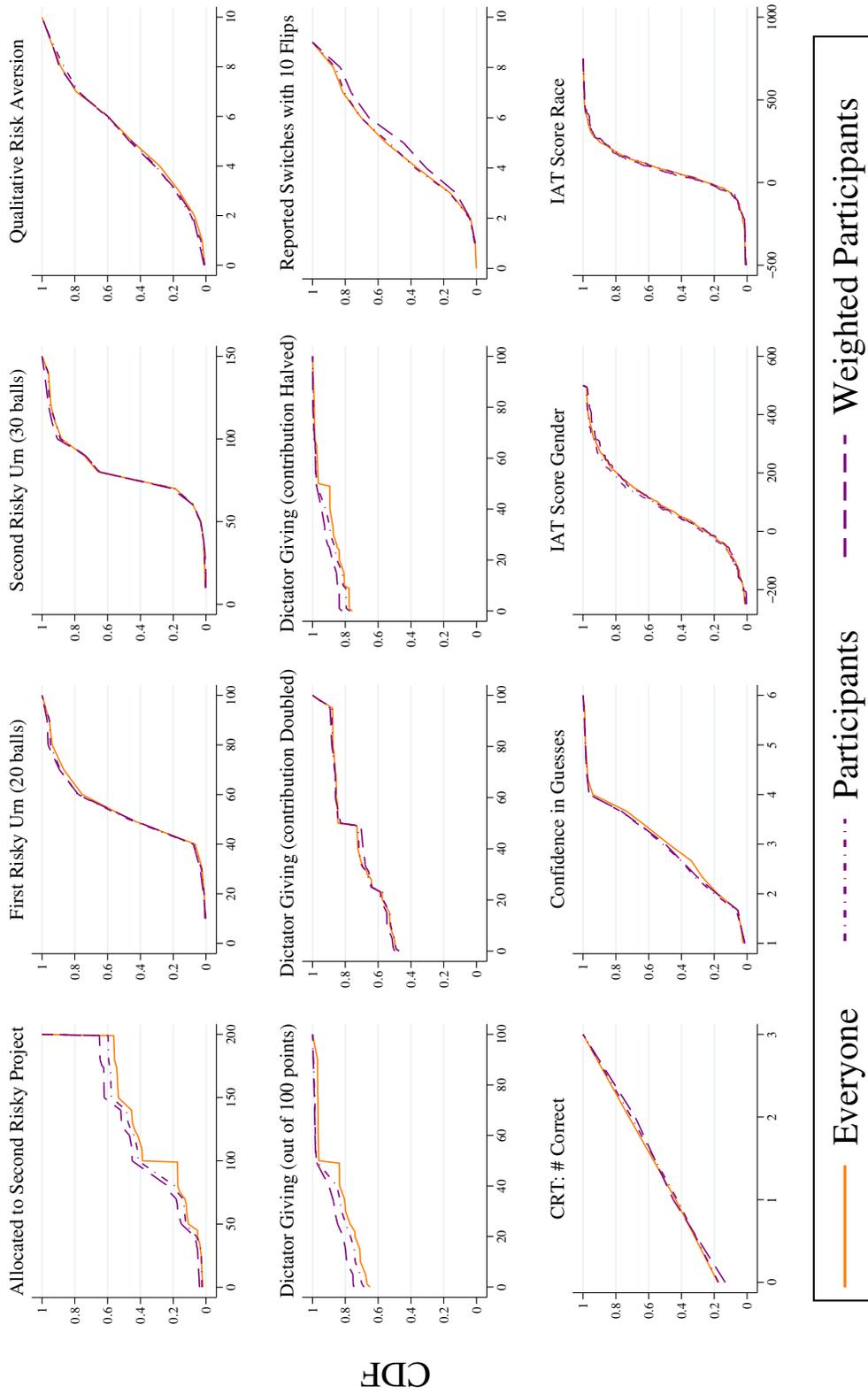


Figure A.5: Correlations across Everyone, Participants, and Weighted Participants (5% Level)

	Risk Aversion	Discounting (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		—	+00	-00	-00	-00	—	—	0	0	—
Discounting (δ)	—		0	0	0	+	0	0	0	0	0
Dictator	+00	0		—	—	0--	0--	-0-	00+	0	-00
Prisoner's Dilemma	-00	0	—		+	0	+	0	0	0	+
Lying	-00	0	—	+		+00	+00	+	0	0++	+00
Cognitive	-00	+	0--	0	+00		+0+	+0+	-00	0	+
Confidence	—	0	0--	+	+00	+0+		+	0	+00	+
Compete	—	0	-0-	0	+	+0+	+		0	+00	++0
IAT Race	0	0	00+	0	0	-00	0	0		+	0
IAT Gender	0	0	0	0	0++	0	+00	+00	+		+
Male	—	0	-00	+	+00	+	+	++0	0	+	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.

Figure A.6: Correlations across Everyone, Participants, and Weighted Participants (1% Level)

	Risk Aversion	Discounting (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		--0	0	0	-00	0	-00	-	0	0	-
Discounting (δ)	--0		0	0	0	++0	0	0	0	0	0
Dictator	0	0		-	-	00-	0-0	-0-	0	0	0
Prisoner's Dilemma	0	0	-		+	0	0++	0	0	0	+
Lying	-00	0	-	+		0	+00	+	0	0	+00
Cognitive	0	++0	00-	0	0		0	+00	0	0	+
Confidence	-00	0	0-0	0++	+00	0		+	0	+00	+
Compete	-	0	-0-	0	+	+00	+		0	+00	++0
IAT Race	0	0	0	0	0	0	0	0		+	0
IAT Gender	0	0	0	0	0	0	+00	+00	+		+
Male	-	0	0	+	+00	+	+	++0	0	+	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.

Figure A.7: Correlations on the CCS and in the Lab (10% Level)

	Risk Aversion	Discounting (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		-0	0	0	0	0	0	0-	0	0	-0
Discounting (δ)	-0		0	0-	0	0	0	0	0	0	0
Dictator	0	0		-	-	0	-0	0	0	0	0
Prisoner's Dilemma	0	0-	-		+	0	0	0	0	0	+0
Lying	0	0	-	+		0	+0	0	0	0	0
Cognitive	0	0	0	0	0		0	0	0	+	+
Confidence	0	0	-0	0	+0	0		+0	0	0	+0
Compete	0-	0	0	0	0	0	+0		0	0	0
IAT Race	0	0	0	0	0	0	0	0		0+	0
IAT Gender	0	0	0	0	0	+	0	0	0+		+
Male	-0	0	0	+0	0	+	+0	0	0	+	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.

Table A.4: Response Time to CCS Solicitation is not Indicative of Measured Behaviors.

	Samples				Differences	
	Everyone (E)	One Email	One Week (W)	More Than One Week (M)	E–W	E–M
First Risky Project (out of 100)	59 (1.2)	59 (1.8)	59 (1.5)	61 (2.1)	0.74 (1.9)	–1.4 (2.4)
Second Risky Project (out of 200)	143 (2.1)	141 (3.0)	142 (2.6)	145 (3.5)	1.2 (3.3)	–2.2 (4.1)
First Risky Urn (20 balls)	59 (0.52)	59 (0.73)	59 (0.64)	60 (0.88)	0.30 (0.82)	–0.56 (1.0)
Second Risky Urn (30 balls)	86 (0.73)	86 (1.0)	86 (0.89)	86 (1.3)	–0.01 (1.2)	0.02 (1.5)
Qualitative Risk Aversion	5.8 (0.08)	5.7 (0.12)	5.7 (0.10)	6.0 (0.13)	0.10 (0.12)	–0.18 (0.15)
Monthly Discount Rate (δ)	0.77 (0.01)	0.77 (0.01)	0.76 (0.01)	0.78 (0.01)	0.00 (0.01)	–0.01 (0.01)
First Dictator Game (given out of 100)	14 (0.84)	14 (1.3)	14 (1.1)	15 (1.4)	0.35 (1.4)	–0.65 (1.6)
Second Dictator Game (given out of 300)	38 (2.4)	37 (3.5)	38 (3.0)	38 (3.9)	0.03 (3.8)	–0.05 (4.6)
Dictator, Tokens Given are Doubled	26 (1.2)	27 (1.8)	27 (1.5)	25 (1.9)	–0.75 (1.9)	1.4 (2.3)
Dictator, Tokens Given are Halved	9.0 (0.68)	8.8 (1.0)	8.8 (0.85)	9.3 (1.2)	0.20 (1.1)	–0.36 (1.3)
Prisoner’s Dilemma (% dominant strat.)	68 (1.5)	69 (2.2)	69 (1.9)	66 (2.5)	–0.76 (2.4)	1.39 (2.9)
Reported Heads (out of 5)	3.3 (0.04)	3.3 (0.06)	3.3 (0.05)	3.3 (0.07)	0.01 (0.07)	–0.01 (0.08)
Reported Switches (out of 9)	5.5 (0.07)	5.5 (0.10)	5.5 (0.09)	5.4 (0.12)	–0.03 (0.11)	0.05 (0.14)
Raven’s Matrices (out of 5)	1.8 (0.04)	1.9 (0.07)	1.8 (0.05)	1.8 (0.08)	–0.01 (0.07)	0.02 (0.09)
CRT (out of 3)	1.7 (0.04)	1.8 (0.06)	1.7 (0.05)	1.6 (0.06)	–0.05 (0.06)	0.09 (0.07)
Confidence in Guesses	3.1 (0.03)	3.1 (0.05)	3.1 (0.04)	3.2 (0.05)	0.02 (0.05)	–0.04 (0.06)
Competition (% competing)	33 (1.7)	29 (2.4)	31 (2.0)	37 (2.9)	2.1 (2.6)	–3.91 (3.3)
IAT Race	81 (5.6)	83 (8.4)	82 (6.8)	80 (9.9)	–0.91 (8.8)	1.7 (11)
IAT Gender	95 (5.9)	81 (8.6)	84 (6.9)	115 (10.8)	11 (9.1)	–20 (12)
Percent Male	62 (1.7)	60 (2.5)	60 (2.1)	65 (2.8)	2.0 (2.7)	–3.7 (3.3)
N	819	374	530	289	–	–

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level, with standard errors in parentheses.

Table A.5: Those that wait more than a week to participate are less likely to go to the lab.

	Samples			Differences		
	Everyone (E)	One Email	One Week (W)	More Than One Week (M)	E–W	E–M
Percent Lab Participant	43 (1.7)	47 (2.6)	47 (2.2)	35 (2.8)	–4.4 (2.8)	8.1** (3.3)
Avg. Lab Sessions	1.3 (0.09)	1.5 (0.15)	1.5 (0.12)	0.85 (0.11)	–0.24 (0.15)	0.44*** (0.14)
N	819	374	530	289	–	–

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level, with standard errors in parentheses.

Figure A.8: Distribution of Responses in the Spring 2015 Survey vs. the Lab ($N = 96$)

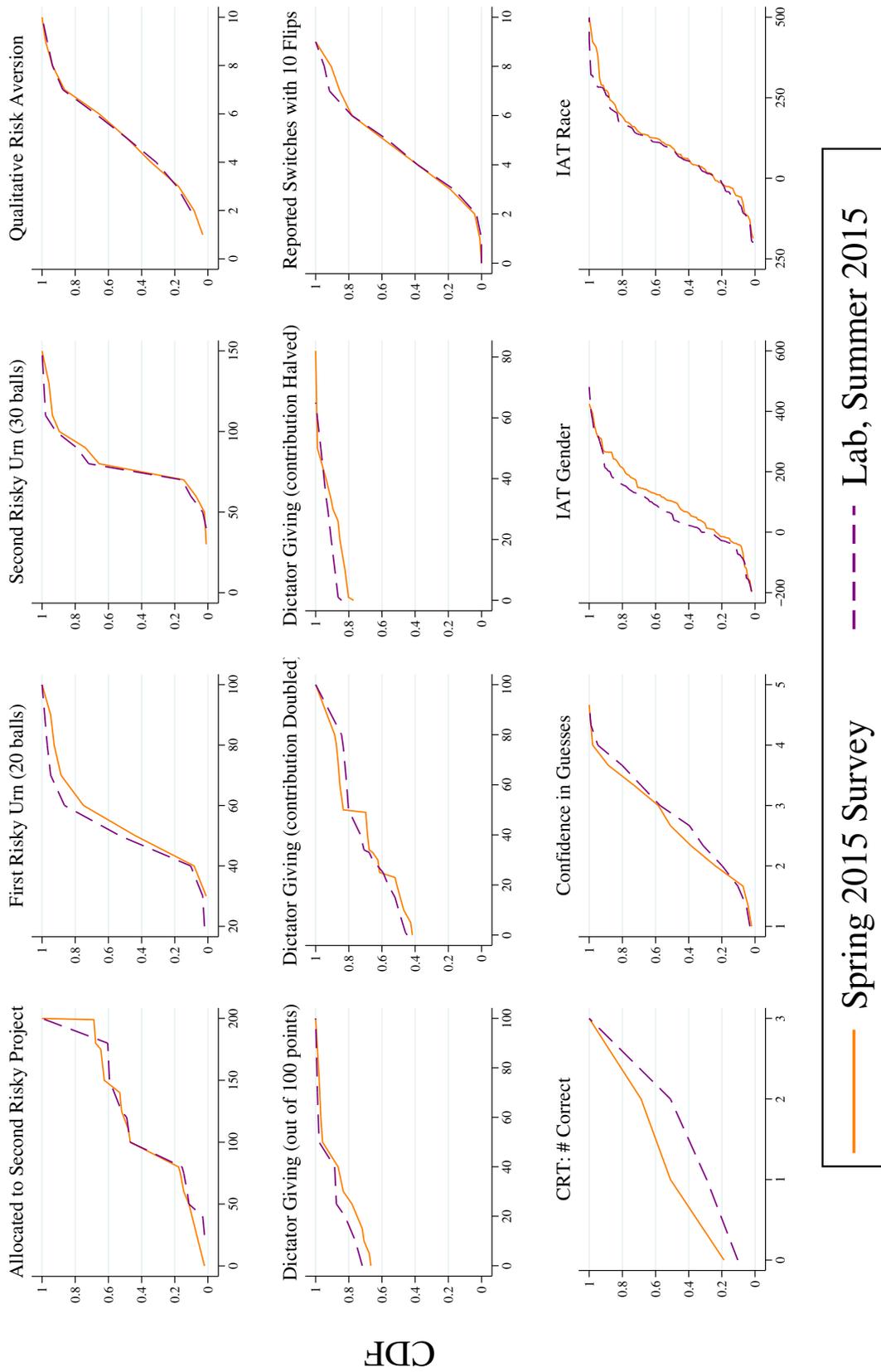


Table A.6: There are few significant differences based on the amount of lab participation.

	By Participation			Differences		
	All Participants (P)	Below Median (B)	Above Median (A)	P–B	P–A	B–A
First Risky Project (out of 100)	55 (1.8)	57 (2.5)	52 (2.7)	–1.9 (3.1)	2.4 (3.2)	4.4 (3.6)
Second Risky Project (out of 200)	139 (3.2)	144 (4.2)	132 (4.9)	–5.1 (5.2)	6.4 (5.9)	11.5* (6.4)
First Risky Urn (20 balls)	58 (0.77)	58 (1.1)	58 (1.0)	0.11 (1.4)	–0.14 (1.3)	–0.25 (1.5)
Second Risky Urn (30 balls)	86 (1.1)	86 (1.6)	86 (1.4)	0.05 (2.0)	–0.06 (1.8)	–0.10 (2.2)
Qualitative Risk Aversion	5.7 (0.12)	5.7 (0.17)	5.8 (0.18)	0.07 (0.21)	–0.09 (0.22)	–0.16 (0.25)
Monthly Discount Rate (δ)	0.78 (0.01)	0.78 (0.01)	0.77 (0.02)	0.00 (0.02)	0.00 (0.02)	0.01 (0.02)
First Dictator Game (given out of 100)	12 (1.1)	14 (1.6)	9.4 (1.6)	–1.9 (2.0)	2.3 (2.0)	4.2* (2.3)
Second Dictator Game (given out of 300)	32 (3.2)	36 (4.5)	26 (4.4)	–4.4 (5.5)	5.5 (5.4)	9.8 (6.2)
Dictator, Tokens Given are Doubled	26 (1.8)	27 (2.5)	26 (2.7)	–0.28 (3.1)	0.35 (3.2)	0.63 (3.6)
Dictator, Tokens Given are Halved	7.8 (0.94)	9.2 (1.4)	5.9 (1.2)	–1.5 (1.7)	1.8 (1.5)	3.3* (1.8)
Prisoner’s Dilemma (% dominant strat.)	67.1 (2.3)	65.9 (3.1)	68.7 (3.5)	1.3 (3.9)	–1.6 (4.2)	–2.8 (4.7)
Reported Heads (out of 5)	3.4 (0.06)	3.3 (0.08)	3.5 (0.10)	0.09 (0.11)	–0.11 (0.11)	–0.20 (0.13)
Reported Switches (out of 9)	5.5 (0.11)	5.4 (0.15)	5.6 (0.17)	0.10 (0.18)	–0.13 (0.20)	–0.23 (0.22)
Raven’s Matrices (out of 5)	1.8 (0.07)	1.8 (0.09)	1.8 (0.10)	–0.02 (0.11)	0.02 (0.12)	0.04 (0.14)
CRT (out of 3)	1.7 (0.06)	1.7 (0.08)	1.7 (0.09)	0.00 (0.10)	0.00 (0.10)	0.00 (0.12)
Confidence in Guesses	3.1 (0.05)	3.0 (0.07)	3.1 (0.07)	0.02 (0.08)	–0.03 (0.08)	–0.05 (0.10)
Competition (% competing)	34 (2.5)	34 (3.4)	33 (3.8)	–0.64 (4.3)	0.81 (4.6)	1.5 (5.1)
IAT Race	87 (8.5)	90 (12)	83 (11)	–3.2 (15)	4.0 (14)	7.1 (17)
IAT Gender	85 (8.5)	73 (10)	100 (14)	12 (13)	–15 (17)	–27 (17)
Percent Male	55 (2.7)	57 (3.6)	54 (4.0)	–1.5 (4.4)	1.9 (4.8)	3.4 (5.4)
N	350	195	155	–	–	–

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level, with standard errors in parentheses.

Table A.7: Re-weighting the CCS to be demographically representative does not change conclusions.

	Weightings			Differences	
	Unweighted (U)	Gender (G)	Race (R)	U–G	U–R
First Risky Project (out of 100)	59 (1.2)	57 (1.9)	64 (2.5)	2.3 (2.3)	–4.5 (2.8)
Second Risky Project (out of 200)	143 (2.1)	138 (3.2)	155 (4.2)	4.5 (3.8)	–12** (4.7)
First Risky Urn (20 balls)	59 (0.52)	59 (0.84)	59 (1.1)	0.14 (0.99)	0.46 (1.2)
Second Risky Urn (30 balls)	86 (0.73)	86 (1.2)	85 (1.5)	0.17 (1.4)	0.87 (1.7)
Qualitative Risk Aversion	5.8 (0.08)	5.7 (0.12)	5.8 (0.16)	0.09 (0.15)	–0.01 (0.18)
Monthly Discount Rate (δ)	0.77 (0.01)	0.76 (0.01)	0.81 (0.01)	0.00 (0.01)	–0.04** (0.02)
First Dictator Game (given out of 100)	14 (0.84)	14 (1.4)	16 (1.7)	–0.31 (1.6)	–2.5 (1.9)
Second Dictator Game (given out of 300)	38 (2.4)	39 (3.9)	44 (4.9)	–0.80 (4.5)	–5.8 (5.4)
Dictator, Tokens Given are Doubled	26 (1.2)	26 (2.0)	33 (2.4)	0.08 (2.3)	–6.5** (2.7)
Dictator, Tokens Given are Halved	9.0 (0.68)	9.7 (1.1)	9.0 (1.5)	–0.73 (1.3)	–0.05 (1.6)
Prisoner’s Dilemma (% dominant strat.)	68 (1.5)	67 (2.4)	66 (3.1)	0.99 (2.9)	2.3 (3.4)
Reported Heads (out of 5)	3.3 (0.04)	3.3 (0.07)	3.2 (0.09)	0.02 (0.08)	0.10 (0.10)
Reported Switches (out of 9)	5.5 (0.07)	5.4 (0.11)	5.4 (0.15)	0.06 (0.13)	0.09 (0.16)
Raven’s Matrices (out of 5)	1.8 (0.04)	1.8 (0.07)	1.8 (0.09)	0.02 (0.08)	–0.04 (0.10)
CRT (out of 3)	1.7 (0.04)	1.6 (0.06)	1.7 (0.08)	0.07 (0.07)	–0.06 (0.09)
Confidence in Guesses	3.1 (0.03)	3.1 (0.05)	3.1 (0.07)	0.05 (0.06)	0.00 (0.07)
Competition (% competing)	33.46 (1.7)	31 (2.6)	33 (3.4)	2.4 (3.1)	0.34 (3.8)
IAT Race	81 (5.6)	83 (9.2)	90 (12)	–1.1 (11)	–8.7 (13)
IAT Gender	95 (5.9)	83 (9.2)	93 (12)	12 (11)	1.7 (13)
N	819	819	819	–	–

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level, with standard errors in parentheses.

B Low-incentive Mechanical Turk Sample

Due to concerns that the rate of pay for the MTurk sample was too high, we gave our survey, with half the rate of pay, to an additional MTurk sample on August 13, 2019. The survey was identical to that run with our previous MTurk sample, with two exceptions. First, the level of incentives was one half that of the prior survey. This was accomplished by changing the exchange rate between tokens and money to 600 tokens / \$1. All token values in each task thus remained the same. Second, we included attention screeners in our survey which were meant to weed out both bots, and inattentive participants. As our survey interface would have been quite difficult for a bot to navigate, there was no indication that bots were able to complete the survey, much less the attention screeners. However, about 20% of the respondents did not pass all three attention screeners.

We received a total of 1,264 responses, of which 212 failed at least one attention screener. The average payment was \$5.21, and the median time to complete the survey was 37 minutes. A final difference between this sample and our prior sample was that Amazon had raised the fees on MTurk considerably. Thus, we paid roughly \$2 per participant in additional fees to Amazon.

Table B.1 compares the average level of responses between our original MTurk sample, the half-pay sample, and the half-pay sample once those who failed at least one attention screener are removed. The fourth and fifth columns compute the differences in average responses between the two MTurk samples, and finds that, remarkably, there is almost no difference between the two. The differences that do exist are that respondents in the half-pay sample are slightly more dishonest, and have somewhat lower implicit bias towards African Americans. These isolated changes could be due to changes in the sample population over the four year period between the two MTurk surveys, or it could be due to the change in incentives. The former seems somewhat more likely as it is unclear why differences in incentives would affect only these two measures. Either way, the level of incentives seems to

have very little effect on the responses of the MTurk sample. The sixth column compares the difference between the overall sample and only those that passed all three attention screener. Here there is a single, isolated, difference, on the CRT.

Table B.2 is an analogue of Table 3. There appear to be no substantive differences between the two MTurk samples in these representations of the data. A substantial decrease in incentives has little effect on our behavioral measures, or on measurement error.

Finally, Figure 3 is the analogue of Figure 1, where the three samples are the original MTurk sample, the new MTurk sample, and the MTurk sample with inattentive participants removed. As before, a positive and significant correlation (at the 10% level) is denoted with a “+”, a negative and significant correlation is denoted with a “-”, and an insignificant correlation is denoted with a “0.” When all three samples agree, we use a single symbol in that cell.

As can be seen, in 39 out of 55 cells there is complete agreement, and in the remainder there is partial agreement. This is very close to what one might expect by chance from the same sample. In particular, we aggregated the two MTurk samples, and then randomly allocated them to the three different subsamples, and computed the corresponding correlation figure. We did this 100 times. Across these 100 simulations, 37.1 of the 55 correlations we consider are, on average, in complete agreement, with 0.3 of 55 in complete disagreement. The remainder show partial agreement. In Figure B.1 39 out of 55 correlations are in complete agreement across the three samples. Thus, we see slightly more agreement than due to chance: 39 out of 55 is at the 75th percentile of our simulation results.

Table B.1: Responses on MTurk are largely the same regardless of incentive level.

	Baseline	Samples		Differences		
		Half Pay	Half Pay Screened	1–2	1–3	2–3
First Risky Project (out of 100)	44 (0.85)	45 (0.81)	44 (0.89)	–1.3 (1.2)	–0.43 (1.2)	0.90 (1.2)
Second Risky Project (out of 200)	98 (1.7)	101 (1.6)	99 (1.8)	–3.0 (2.4)	–1.7 (2.5)	1.3 (2.4)
First Risky Urn (20 balls)	56 (0.63)	56 (0.56)	57 (0.59)	–0.16 (0.84)	–0.69 (0.86)	–0.52 (0.81)
Second Risky Urn (30 balls)	78 (0.96)	77 (0.87)	78 (0.92)	0.52 (1.3)	–0.09 (1.3)	–0.61 (1.3)
Qualitative Risk Aversion	4.9 (0.08)	4.8 (0.07)	4.6 (0.07)	0.16 (0.10)	0.31*** (0.11)	0.15 (0.10)
Monthly Discount Rate (δ)	0.67 (0.01)	0.66 (0.01)	0.66 (0.01)	0.01 (0.01)	0.01 (0.01)	0.00 (0.01)
First Dictator Game (given out of 100)	26 (0.71)	27 (0.62)	26 (0.68)	–1.0 (0.94)	–0.63 (0.98)	0.37 (0.92)
Second Dictator Game (given out of 300)	74 (2.0)	74 (1.8)	73 (2.0)	–0.27 (2.7)	0.73 (2.8)	1.0 (2.7)
Dictator, Tokens Given are Doubled	30 (0.79)	29 (0.69)	29 (0.76)	0.53 (1.1)	0.65 (1.1)	0.12 (1.0)
Dictator, Tokens Given are Halved	25 (0.74)	26 (0.66)	26 (0.73)	–1.5 (0.99)	–1.1 (1.0)	0.35 (0.98)
Prisoner’s Dilemma (% dominant strat.)	57 (1.3)	56 (1.2)	56 (1.3)	0.95 (1.8)	0.72 (1.8)	–0.24 (1.7)
Reported Heads (out of 5)	3.0 (0.03)	3.2 (0.03)	3.1 (0.03)	–0.15*** (0.04)	–0.12** (0.05)	0.03 (0.04)
Reported Switches (out of 9)	4.5 (0.06)	4.7 (0.05)	4.6 (0.06)	–0.18** (0.08)	–0.10 (0.08)	0.08 (0.08)
Raven’s Matrices (out of 5)	1.3 (0.04)	1.3 (0.03)	1.3 (0.03)	0.03 (0.05)	–0.01 (0.05)	–0.04 (0.05)
CRT (out of 3)	1.4 (0.04)	1.3 (0.03)	1.5 (0.04)	0.01 (0.05)	–0.09* (0.05)	–0.10** (0.05)
Confidence in Guesses	2.9 (0.03)	3.0 (0.03)	3.0 (0.03)	–0.07 (0.04)	–0.01 (0.04)	0.06 (0.04)
Competition (% competing)	29 (1.5)	34 (1.3)	32 (1.4)	–3.9** (2.0)	–2.1 (2.0)	1.8 (2.0)
IAT Race	68 (4.8)	44 (4.5)	49 (4.7)	24*** (6.6)	18*** (6.7)	–5.4 (6.5)
IAT Gender	90 (4.8)	93 (4.6)	94 (4.8)	–2.4 (6.7)	–4.0 (6.8)	–1.6 (6.6)
Percent Male	50 (1.6)	47 (1.4)	45 (1.5)	3.3 (2.1)	4.6** (2.2)	1.3 (2.1)
N	995	1,264	1,052	–	–	–

Notes: ***, **, * denote statistical significance at the 1%, 5%, and 10% level, with standard errors in parentheses.

Table B.2: Percent of Variation due to Noise

Sample:	Baseline	Half Pay	Half Pay, Screened
Risky Projects	47% (2.7%)	43% (2.3%)	43% (2.5%)
Risky Urns	32% (2.3%)	37% (2.2%)	38% (2.4%)
Lottery Menu	33% (2.4%)	36% (2.2%)	35% (2.3%)
Ambiguous Urn	31% (2.3%)	24% (1.8%)	24% (2.0%)
Compound Urn	26% (2.1%)	27% (1.9%)	27% (2.1%)
Dictator Giving	18% (1.8%)	20% (1.7%)	18% (1.8%)
IAT Race	46% (2.7%)	41% (2.3%)	44% (2.6%)
IAT Gender	46% (2.7%)	43% (2.3%)	46% (2.6%)
N	995	1,264	1,052

Figure B.1: Correlations across MTurk, MTurk 1/2 pay, and MTurk 1/2 pay and Screened

	Risk Aversion	Discount (δ)	Dictator	Prisoner's Dilemma	Lying	Cognitive	Confidence	Compete	IAT Race	IAT Gender	Male
Risk Aversion		+00	0--	-00	0	0	-	-	0	+00	-
Discounting (δ)	+00		-	0--	+00	+	-	0	0+0	0	0++
Dictator	0--	-		-	-	-	+	0	0	0+0	-
Prisoner's Dilemma	-00	0--	-		+	+	0	+00	0-0	0	00-
Lying	0	+00	-	+		+	0	+00	0	0	+
Cognitive	0	+	-	+	+		-	0-0	0	0	+
Confidence	-	-	+	0	0	-		+	0	0	+
Compete	-	0	0	+00	+00	0-0	+		+00	00-	+
IAT Race	0	0+0	0	0-0	0	0	0	+00		+	0
IAT Gender	+00	0	0+0	0	0	0	0	00-	+		0
Male	-	0++	-	00-	+	+	+	+	0	0	

Notes:  indicates complete agreement,  complete disagreement, and  two out of three samples agreeing.