Appendix

(For Online Publication)

Appendix A: *Piece-rate* vs. *Tournament* Treatments

In this Appendix, we describe the results from our incentive scheme treatments and explain why we have not included tests of the hypotheses related to the incentive schemes in the main text. In short, our incentive structure in the *tournament* treatment seems not to have been salient enough to induce changes in effort compared to the *piece-rate* treatment.

Hypothesis A1 (Main Effect of *Tournament*): ambiguous

The effect of competition in the *tournament* treatment relative to the *piece-rate* treatment is theoretically ambiguous. In order to derive a comparative static prediction, we would need to know the optimal effort in the *piece-rate* and *tournament* treatment, but the optimal effort levels depend on the parameters (payoffs) as well as the unobserved costs of effort. As such, we cannot predict the effect of *tournament* treatment on learning relative to the *piece-rate* treatment.

Columns (1) and (2) in Pane A of Table A1 present the estimated treatment effects of competition on learning with the *piece-rate* treatment serving as the control treatment. There is no difference in learning between the two treatments even after controlling the individual covariates in Column (2).[1] Columns (3) and (4) present the estimated treatment effects by initial performance (rank). For both subjects in the top half and bottom half, these estimates are far from statistically significant at conventional levels and small in magnitude. Figure A1 confirms that the kernel densities of learning for the full sample, subjects in the bottom half, and subjects in the top half are almost identical in both treatments. The p-value for the two-sample Kolmogorov-Smirnov test for the equality of the distributions for the full sample (Panel A) is 0.618, while the p-value for subjects in bottom half (Panel B) is 0.982.

Hypothesis A2 (Interaction of *Tournament* and *Teaching*): negative

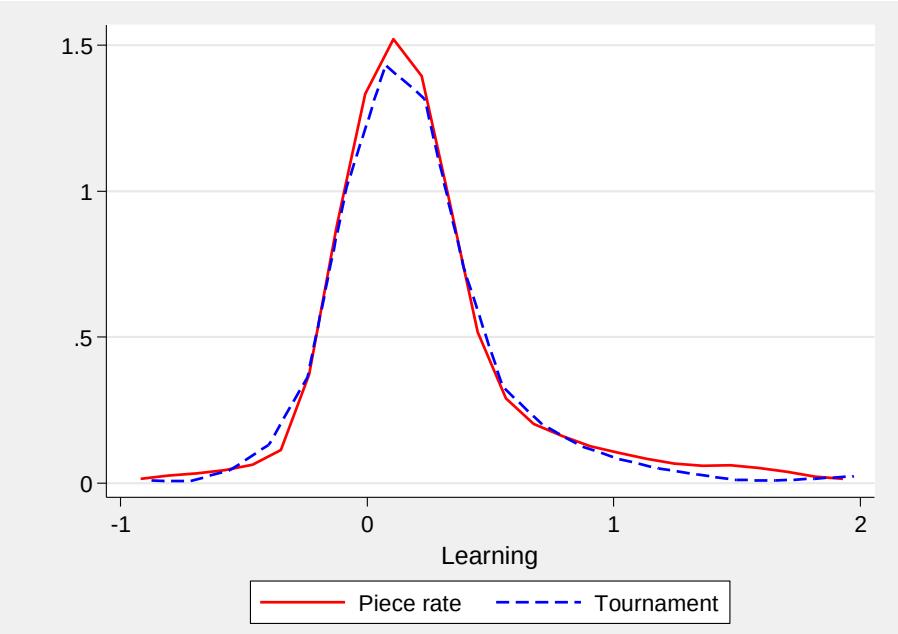
We also hypothesized that there would be interactions between the *tournament* and *teaching* treatments. Specifically, we expected that subjects would be teach each other less in the Practice Block when they anticipated competing against the other group members subsequently in the Evaluation Block—an effect which would be strongest when grouped with subjects of similar ability (i.e., in the *tracked* treatment). We found, however, no evidence of any such interaction effects. These and all of our estimates concerning the (non-)effects of the *tournament* treatment are available on request.

Overall, the *tournament* treatment does not seem to have affected learning. One possibility is that the payoff structure by rank in the *tournament* treatment was very nearly linear, and the payoffs were chosen so as to lead to similar earnings in the *piece-rate* treatment and the *tournament* treatment. As such, the tournament incentives may have been insufficient to induce significant changes in behavior. We find similar results in the Nonogram experiments presented in Panel B of Table A1.

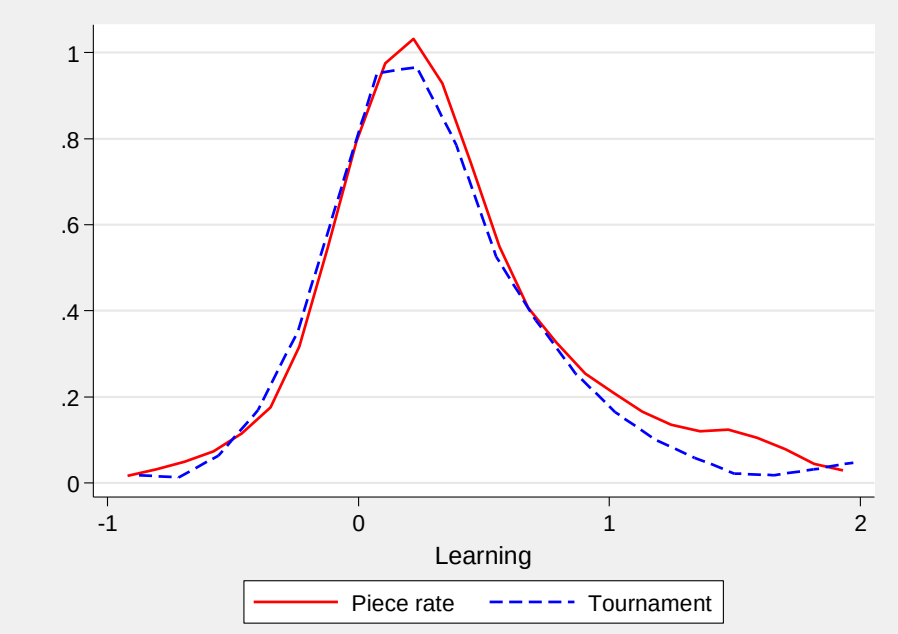
[1] The raw means of learning in the *piece-rate* and *tournament* treatments are very similar (24.4 vs. 25.5 seconds).

Figure A1: Kernel Density of Learning by *Piece-rate* vs. *Tournament* (Sudoku)

A. Full Sample



B. Among Subjects in Bottom Half



Notes: Kernel density plots of learning for the *piece-rate* and *tournament* treatments are displayed. Panel A uses the full sample, while Panel B is limited to subjects in bottom half who ranked 5–8 in the Ability Block (T=0). Learning is calculated by subtracting the average solving time (AST) in Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. For Panel A, the p-value for the two-sample Kolmogorov-Smirnov test for the equality of the distributions is 0.618, while that for Panel B is 0.982. There are 448 subjects.

Table A1: Effect of *Tournament* on Learning

A. Sudoku

	(1)	(2)	(3)	(4)
A. Overall				
Tournament	-0.006	-0.005		
	(0.056)	(0.054)		
B. Heterogeneity				
Tournament for Top half			-0.006	-0.005
			(0.056)	(0.054)
Tournament for Bottom half			-0.006	-0.005
			(0.056)	(0.054)
Controls	No	Yes	No	Yes

B. Nonograms

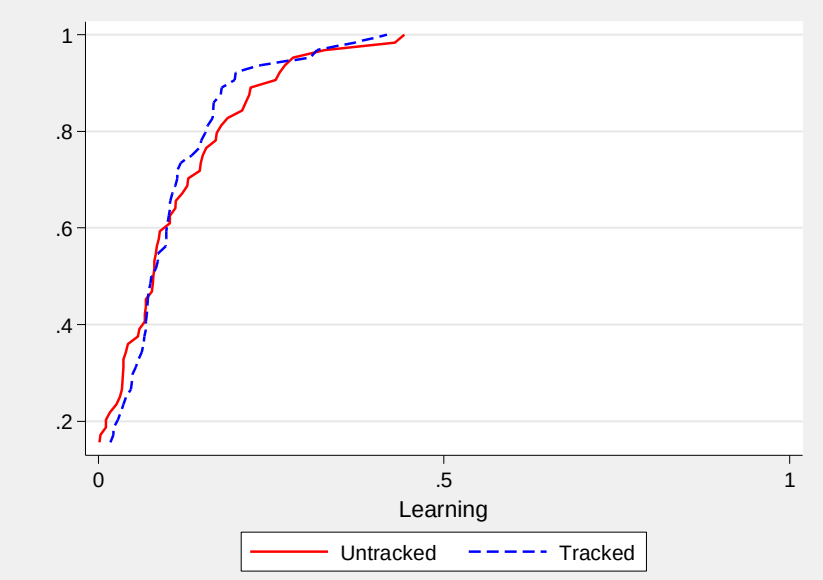
	(1)	(2)	(3)	(4)
A. Overall				
Tournament	0.006	0.003		
	(0.070)	(0.068)		
B. Heterogeneity				
Tournament for Top half			0.023	0.010
			(0.021)	(0.024)
Tournament for Bottom half			-0.012	-0.007
			(0.134)	(0.131)
Controls	No	Yes	No	Yes

Notes: Each column reports the estimated treatment effects from a different OLS regression. The estimates in Columns (1) and (2) are obtained by estimating equation [2] with the *teaching* dummy being replaced by a *tournament* dummy. The estimates in Columns (3) and (4) are obtained by estimating equation [3] with the *teaching* dummy being replaced by a *tournament* dummy (the original estimates available upon request). The control group is the *piece rate* treatment. The estimated treatment effects and their standard errors were computed using the *lincom* command in STATA. Standard errors clustered at the group level are reported in parentheses. The outcome is learning, which is calculated by subtracting the average solving time (AST) in Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. The subjects in bottom half are those subjects ranked 5–8 in the Ability Block (T=0) and the top half those ranked 1–4. The controls include a dummy for being male, a dummy for being experienced with Sudoku, risk attitudes (0–9), prosociality (0–5), and a dummy for the eight subjects who could not solve any Sudoku puzzles in the Ability Block (T=0). See Table 2 for definitions of each control variable. For Panel A (Sudoku), there were 28 sessions with 224 subjects (8 subjects per session) each for the *piece rate* and *tournament* treatments, respectively. For Panel B (Nonograms), there were 16 sessions with 128 subjects each for the *piece rate* and *tournament* treatments, respectively. Each session consisted of two groups (4 subjects per group). Significance levels: *** p<0.01, ** p<0.05, * p<0.10

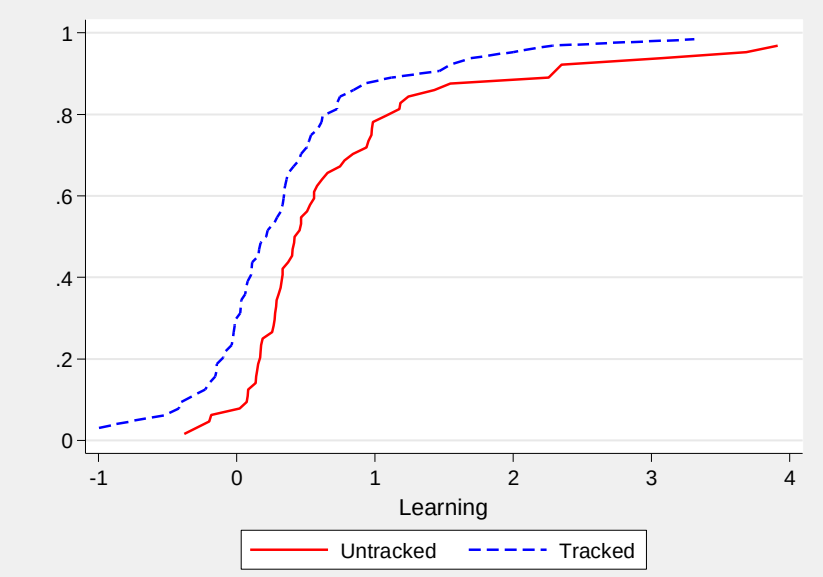
Appendix B: Additional Figures and Tables

Figure B1: Cumulative Distribution of Learning by *Tracking* under *Teaching* separately for Top and Bottom half (Sudoku)

A. Among Subjects in Top Half

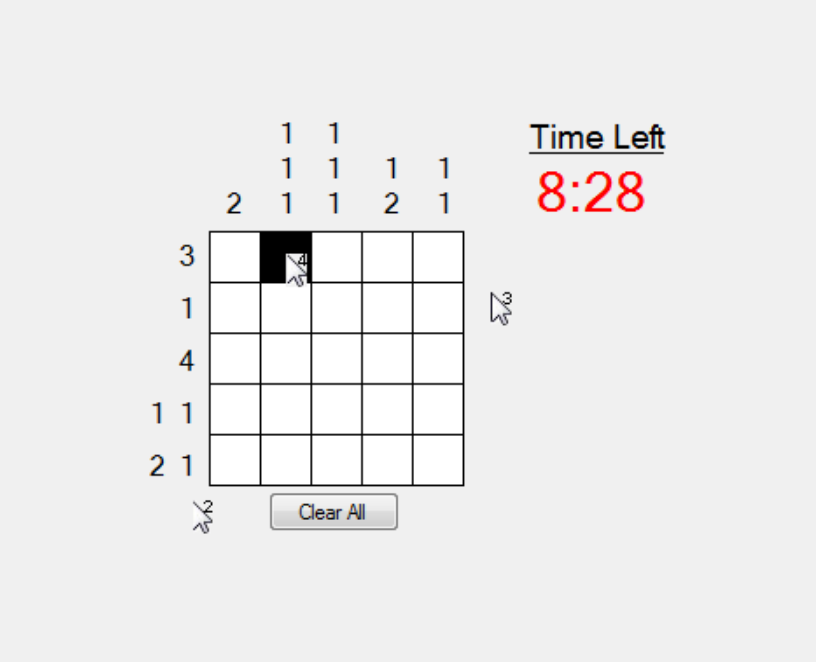


B. Among Subjects in Bottom Half



Notes: The cumulative distributions of learning reported in Figure 6-B are separately displayed for subjects in the top half (Panel A) and subjects in the bottom half (Panel B). Learning is calculated by subtracting the average solving time (AST) in the Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. Also note that the scales on the *x*-axes are different for the two graphs. The sample is limited to 32 *teaching* sessions with 256 subjects. For each session, there is one group for the top half and the bottom half; thus there are 32 groups each for the top half and bottom half. For Panel A, the p-value for the two-sample Kolmogorov-Smirnov test for the equality of the distributions in the *untracked* and *tracked* treatments is 0.843, while that for Panel B is 0.004.

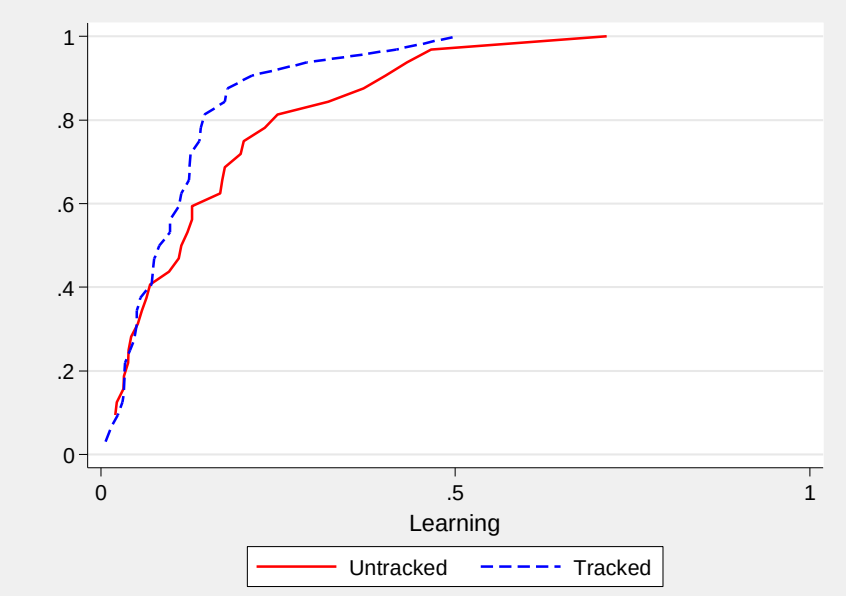
Figure B2: Screenshot from the *Teaching* Treatment during the Practice Block (Nonograms)



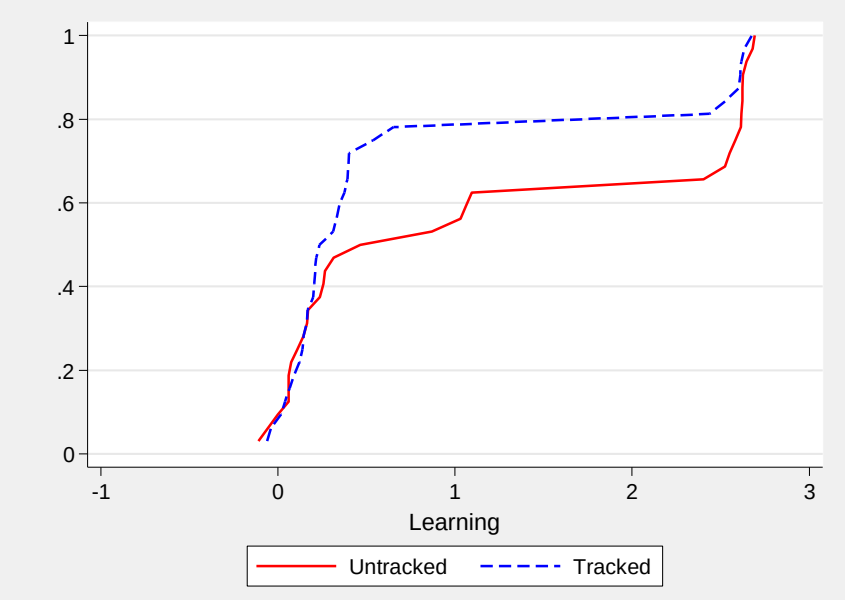
Notes: Subjects are able to simultaneously edit a common 5×5 Nonogram puzzle during the Practice Block. Each mouse arrow is labeled with the within-group performance rank of the person in the Ability Block (T=0). Performance is measured by the number of Nonogram puzzles solved with the average solving time serving as a tie-breaker. In the *no-teaching* treatment, the three arrows showing the within-group rank of the other subjects would not have been visible as each subject worked independently.

Figure B3: Cumulative Distributions of Learning by *Tracking* in the *Teaching* Treatment for Subjects in the Top and Bottom Half (Nonograms)

A. Among Subjects in the Top Half



B. Among Subjects in the Bottom Half



Notes: The cumulative distributions of learning reported in Figure 9-B are separately displayed for subjects in the top half (Panel A) and subjects in the bottom half (Panel B). Learning is calculated by subtracting the average solving time (AST) in Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. Also, note that the scales on the *x*-axes are different between the two graphs. The sample is limited to 16 *teaching* sessions with 144 subjects. For each session, there is one group of subjects in the top half and one group of subjects in the bottom half; thus there are 16 groups each for the top half and bottom half. For Panel A, the p-value for the two-sample Kolmogorov-Smirnov test for the equality of the distributions in the *untracked* and *tracked* treatments is 0.434, while that for Panel B is 0.160.

Table B1: Coefficient Estimates for Equations [2] and [3] to Estimate the Effect of *Teaching* on Learning (Sudoku)

Outcome: Learning

	A. Overall		B. Heterogeneity	
	(1)	(2)	(3)	(4)
Teaching	0.113**	0.119**	-0.018	-0.016
	(0.052)	(0.052)	(0.017)	(0.020)
Bottom Half			0.258***	0.256***
			(0.096)	(0.095)
Teaching × Bottom Half			0.185***	0.167***
			(0.053)	(0.052)
Male		0.041		0.007
		(0.059)		(0.057)
Experienced		-0.203***		-0.066
		(0.076)		(0.069)
Risk Attitudes (0–9)		0.010		0.015
		(0.020)		(0.019)
Prosociality (0–5)		-0.007		-0.024
		(0.030)		(0.031)
None Correct at T=0	3.528***	3.416***	3.385***	3.356***
	(0.386)	(0.395)	(0.386)	(0.387)
Constant	0.191***	0.286**	0.102***	0.142
	(0.028)	(0.128)	(0.010)	(0.112)
Controls	No	Yes	No	Yes
R-squared	0.38	0.40	0.44	0.44
# of Sessions	56	56	56	56
# of Groups	112	112	112	112
# of Subjects	448	448	448	448

Notes: Each column reports the results from a different OLS regression. Standard errors clustered at the group level are reported in parentheses. The outcome is learning, which is calculated by subtracting the average solving time (AST) in Evaluation Block (T=1) from that of Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. Experienced takes a value of one if a subject indicates experience with Sudoku puzzles before the experiment. Risk attitudes take on the values from 0 to 9 with higher number indicating higher risk-aversion. Prosociality takes on the values from 0 to 5 with higher number indicating higher prosociality. See Appendix C for details on the elicitation of risk attitude and prosociality. We also include a dummy for the eight subjects who could not solve any Sudoku puzzles in the Ability Block (T=0). There were 56 sessions with 448 subjects (8 subjects per session) in the experiment. The estimates were used to produce Table 3 in the main text. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B2: Coefficient Estimates for Equations [4] and [5] (Sudoku)

Outcome: Learning

	A. Overall	B. Heterogeneity
	(1)	(2)
Tracked	-0.039	0.024
	(0.061)	(0.026)
Teaching	0.171**	0.012
	(0.073)	(0.024)
Tracked × *Teaching*	-0.105	-0.056
	(0.101)	(0.044)
Bottom half		0.231***
		(0.085)
Tracked × Bottom half		-0.119
		(0.107)
Teaching × Bottom half		0.322**
		(0.127)
Tracked × *Teaching* × Bottom half		-0.132
		(0.183)
Controls	Yes	Yes
R-squared	0.41	0.45
# of Sessions	56	56
# of Groups	112	112
# of Subjects	448	448

Notes: Each column reports the results from a different OLS regression. The outcome is learning, which is calculated by subtracting the average solving time (AST) in Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. The controls include a dummy for being male, a dummy for being experienced with Sudoku, risk attitudes (0–9), prosociality (0–5), and a dummy for the subjects who could not solve any Sudoku puzzles at T=0. The bottom half consists of those subjects ranked 5–8. There were 24 *no-teaching* sessions with 192 subjects, and 32 *teaching* sessions with 256 subjects (8 subjects per session). Each session consisted of two groups (4 subjects per group). The estimates were used to produce Table 4 in the main text. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B3: Effect of *Teaching* on Logged Learning, Robustness Checks (Sudoku)

Outcome: Learning (logged)

	(1)	(2)	(3)	(4)
A. Overall				
Teaching	0.030	0.031		
	(0.026)	(0.027)		
B. Heterogeneity				
Teaching for Top half			-0.021	-0.024
			(0.021)	(0.022)
Teaching for Bottom half			0.080*	0.079*
			(0.046)	(0.046)
Controls	No	Yes	No	Yes

Notes: Each column reports the results from a different OLS regression. Columns (1) and (2) come from equation [2] with and without controls using the full sample. Columns (3) and (4) come from the equation [3] with and without controls using the full sample. Here, the control group is the *no-teaching* treatment. The estimated treatment effects and their standard errors reported in the table were computed using the *lincom* command in STATA. The coefficient estimates from equations [2] and [3] are not reported to save space (results available upon request). Standard errors clustered at the group level are reported in parentheses. The outcome is logged learning, which is defined as the difference between logged average solving time (AST) in the Ability Block (T=0) and the Evaluation Block (T=1). The bottom half consists of those subjects ranked 5–8 and the top half those subjects ranked 1–4 in T=0. All regressions—even those labeled as including "no" controls—include a dummy for the eight subjects who could not solve any Sudoku puzzles at T=0. The controls further include a dummy for being male, a dummy for being experienced with Sudoku, risk attitudes (0–9), and prosociality (0–5). See Table 2 for definitions of each control variable. There were 56 sessions with 448 subjects (8 subjects per session). Each session consisted of two groups (4 subjects per group), and thus there were 112 groups. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.10$

Table B4: Group Mean, Group Standard Deviation, and Teaching Frequencies (Sudoku)

	A. Sudoku				B. Nonograms			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Group mean	0.36***		0.02	0.28	0.20		0.04	1.54
	(0.10)		(0.21)	(0.32)	(0.16)		(0.35)	(0.98)
Group SD		0.36***	0.34**	0.40**		0.22	0.18	-0.42
		(0.09)	(0.17)	(0.18)		(0.16)	(0.35)	(0.50)
Group mean × Group SD				-0.18				-1.34
				(0.17)				(0.84)
Vuong test of Zero-Inflated model vs. Standard Poisson	z=3.61	3.57	3.53	3.31	2.91	2.83	2.74	2.83
	p=0.0002	0.0002	0.0002	0.0005	0.0018	0.0023	0.0031	0.0023
# of Groups	64	64	64	64	30	30	30	30

Notes: Each column reports the results from a different zero-inflated Poisson regression in which the number of teaching statements recorded for a given group is the dependent variable. Group mean is the group average of *raw* average solving time (AST) in the Ability Block (T=0) so that higher values indicate that the group consisted of lower-ability subjects while lower values indicate that the group consisted of higher-ability subjects. Because this is a group-level analysis, the sample size for Sudoku in Panel A is 64, while that for Nonograms in Panel B is 30 as there was one Nonogram session (two groups) in which the audio recording did not work. The Vuong tests of the zero-inflated model against standard Poisson models are reported in each column with the z-scores and corresponding p-values. All of these tests support the use of the zero-inflated model.

Table B5: Coefficient Estimates for Equations [4] and [5] Using Logged Learning, Robustness Checks (Sudoku)

Outcome: Learning (Logged)

	A. Overall	B. Heterogeneity
	(1)	(2)
Tracked	-0.025	0.022
	(0.035)	(0.026)
Teaching	0.059	-0.002
	(0.039)	(0.026)
Tracked × *Teaching*	-0.056	-0.044
	(0.053)	(0.044)
Bottom half		0.137**
		(0.055)
Tracked × Bottom half		-0.091
		(0.062)
Teaching × Bottom half		0.123*
		(0.066)
Tracked × *Teaching* × Bottom half		-0.040
		(0.095)
Controls	Yes	Yes
R-squared	0.17	0.23
# of Sessions	56	56
# of Groups	112	112
# of Subjects	448	448

Notes: Each column reports the results from a different OLS regression. The outcome is logged learning, which is defined as the difference between the logged average solving time (AST) in the Ability Block (T=0) and the Evaluation Block (T=1). The controls include a dummy for being male, a dummy for being experienced with Sudoku, risk attitudes (0–9), prosociality (0–5), and a dummy for the subjects who could not solve any Sudoku puzzles at T=0. The bottom half consists of those subjects ranked 5–8. There were 24 *no-teaching* sessions with 192 subjects, and 32 *teaching* sessions with 256 subjects (8 subjects per session). Each session consisted of two groups (4 subjects per group). The estimates were used to produce Table 5 in the main text. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B6: Summary Statistics and Balance Tests (Nonograms)

	A. Overall			B. Heterogeneity		
		p-value of equality test		Bottom half	Top half	
	Mean					Dif
		2×2×2	2×2	(rank5-8)	(rank1-4)	(5)-(6)
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.45	*0.60*	*0.43*	0.41	0.49	-0.09
	[0.52]			[0.49]	[0.55]	(0.07)
Experienced	0.02	*0.33*	*0.57*	0.00	0.03	-0.03**
	[0.12]			[0.00]	[0.17]	(0.02)
Risk Attitude (0–9)	3.63	*0.12*	*0.14*	3.63	3.63	0.00
	[1.60]			[1.60]	[1.60]	(0.18)
Prosociality (0–5)	1.79	*0.28*	*0.13*	1.85	1.72	0.13
	[0.97]			[0.99]	[0.94]	(0.11)
Solved None at T=0	0.20	*0.47*	*0.18*	0.39	0.00	0.39***
	[0.40]			[0.49]	[0.00]	(0.04)
Solved None at T=1	0.07	-	-	0.13	0.00	0.13***
	[0.25]			[0.34]	[0.00]	(0.03)
Raw Average solve time at T=0 (*sec*)	186.47	*0.26*	*0.14*	307.31	65.64	241.66***
	[209.92]			[240.90]	[31.91]	(21.47)
Raw Average solve time at T=1 (*sec*)	85.38	-	-	132.37	38.40	93.97***
	[140.16]			[186.77]	[10.46]	(16.23)
Raw Learning (=AST0-AST1) (*sec*)	101.09	-	-	174.94	27.25	147.69***
	[170.62]			[216.42]	[25.00]	(19.02)
Standardized average solve time at T=0	0.00	*0.26*	*0.14*	0.58	-0.58	1.15***
	[1.00]			[1.15]	[0.15]	(0.10)
Standardized average solve time at T=1	-0.48	-	-	-0.26	-0.71	0.45***
	[0.67]			[0.89]	[0.05]	(0.08)
Learning	0.48	-	-	0.83	0.13	0.70***
	[0.81]			[1.03]	[0.12]	(0.09)
# of Sessions	32			32	32	
# of Groups	64			32	32	
# of Subjects	256			128	128	

Notes: Column (1) reports means for the full sample with standard deviations in brackets. Columns (2) and (3) report the p-values for each variable in the far-left column of the null hypotheses that the means are equal across 8 treatment combinations (Column (2)) and 4 treatment combinations pooling across the incentive treatments (Column (3)). Columns (4) and (5) report the means by ranks in the Ability Block (T=0). The bottom half consists of those subjects ranked 5–8, and the top half consists of those subjects ranked 1–4. Column (6) reports the difference in means between subjects in the top half and subjects in the bottom half with standard errors clustered at the group level in parentheses. See the notes for Table 2 for descriptions of the variables. The experienced variable equals one if a subject reported prior experience with Nonograms and zero otherwise. There were total of 32 sessions with 256 subjects (8 subjects per session). Each session consisted of two groups (4 subjects per group). Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B7: Coefficient Estimates for Equations [2] and [3] for the Effect of *Teaching* on Learning (Nonograms)

Outcome: Learning

	A. Overall		B. Heterogeneity	
	(1)	(2)	(3)	(4)
Teaching	0.203***	0.195***	0.021	0.022
	(0.064)	(0.064)	(0.022)	(0.021)
Bottom Half			-0.018	-0.006
			(0.077)	(0.081)
Teaching × Bottom Half			0.358***	0.337***
			(0.121)	(0.121)
Male		0.040		0.039
		(0.071)		(0.072)
Experienced		-0.279***		-0.166***
		(0.062)		(0.039)
Risk Attitudes (0–9)		0.003		0.002
		(0.025)		(0.026)
Prosociality (0–5)		-0.042		-0.038
		(0.041)		(0.041)
None Correct at T=0	0.091**	0.145	0.119***	0.163
	(0.036)	(0.152)	(0.011)	(0.148)
Constant	1.477***	1.481***	1.390***	1.394***
	(0.172)	(0.171)	(0.180)	(0.179)
Controls	No	Yes	No	Yes
R-squared	0.53	0.53	0.55	0.55
# of Sessions	32	32	32	32
# of Groups	64	64	64	64
# of Subjects	256	256	256	256

Notes: Each column reports the results from a different OLS regression. The outcome is learning, which is calculated by subtracting the average solving time (AST) in the Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. . The controls include a dummy for being male, a dummy for being experienced with Nonograms, risk attitudes (0–9), prosociality (0–5), and a dummy for the subjects who could not solve any Nonogram puzzles at T=0. The bottom half consists of those subjects ranked 5–8 in the Ability Block (T=0). There were 16 *no-teaching* and *teaching* sessions with 128 subjects, respectively (8 subjects per session). Each session consisted of two groups (4 subjects per group). Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B8: Coefficient Estimates for Equations [4] and [5] (Nonograms)

Outcome: Learning

	A. Overall	B. Heterogeneity
	(1)	(2)
Tracked	0.067	0.024
	(0.095)	(0.044)
Teaching	0.260***	0.066**
	(0.093)	(0.029)
Tracked × *Teaching*	-0.131	-0.087*
	(0.128)	(0.050)
Bottom half		-0.040
		(0.129)
Tracked × Bottom half		0.071
		(0.186)
Teaching × Bottom half		0.381**
		(0.187)
Tracked × *Teaching* × Bottom half		-0.089
		(0.243)
Controls	Yes	Yes
R-squared	0.53	0.55
# of Sessions	32	32
# of Groups	64	64
# of Subjects	256	256

Notes: Each column reports the results from a different OLS regression. The outcome is learning, which is calculated by subtracting the average solving time (AST) in the Evaluation Block (T=1) from that in the Ability Block (T=0) so that higher values indicate *improvement* in solving time. Note that AST in the Evaluation Block (T=1) and in the Ability Block (T=0) is standardized by the mean and standard deviation of raw AST at T=0 before taking the difference so that standardized AST at T=0 has a mean of zero and standard deviation of 1. The controls include a dummy for being male, a dummy for being experienced with Nonograms, risk attitudes (0–9), prosociality (0–5), and a dummy for the subjects who could not solve any Nonogram puzzles at T=0. The bottom half consists of those subjects ranked 5–8 in the Ability Block (T=0). There were 24 *no-teaching* sessions with 192 subjects, and 32 *teaching* sessions with 256 subjects (8 subjects per session). Each session consisted of two groups (4 subjects per group). Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Table B9: Frequency of Teaching in the *Tracked* vs. *Untracked* Treatments (Nonograms)

		Tracked		Difference (2)–(1)			Difference (3)–(1)		
	Untracked	Bottom Half	Top Half	OLS	Poisson	Zero-Inflated Poisson	OLS	Poisson	Zero-Inflated Poisson
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Teaching	4.14	5.75	4.00	1.61	0.33*	0.06	-0.14	-0.04	-0.31
	[5.26]	[7.13]	[2.98]	(2.65)	(0.20)	(0.20)	(2.04)	(0.22)	(0.22)
Vuong test of Zero-Inflated model vs. Standard Poisson					z-score = 2.64 p-value = 0.0041			z-score = 2.66 p-value = 0.0039	
# of Groups	14	8	8	22	22	22	22	22	22
# of Sessions	7	8		15	15	15	15	15	15

Notes: The unit of observation is a group. The sample is limited to the 15 *teaching* treatment sessions with 7 sessions for the *untracked* and 8 sessions for *tracked* treatments as there was one *untracked* session in which the audio recording did not work. For the *untracked* treatment, there are total of 14 groups (two groups for each session), while for the *tracked* treatment there are 8 groups each for subjects in bottom half (Group 2 in the *tracked* treatment in Figure 2) and for those in top half (Group 1 in the *tracked* treatment in Figure 2). Column (1) reports the mean number of teaching statements in the *untracked* treatment, and Columns (2) and (3) report the means for the *tracked* treatment for the bottom half group and the top half group, respectively. Standard deviations are reported in brackets. Columns (4)–(6) report the estimated difference between Columns (1) and (2) from OLS, Poisson and zero-inflated Poisson (where the inflation equation includes just a indicator for whether the group was *tracked*) models, respectively, with standard errors in parentheses. Columns (7)–(9) report the corresponding estimated differences between Columns (1) and (3). The bottom half consists of those subjects ranked 5–8 in the Ability Block (T=0) and the top half consists of those subjects ranked 1–4. A teaching statement is defined to be any utterance in which subjects are engaged in trying to teach each other how to do Nonograms such as "You can't have a five there; there is already one in that column." The Vuong tests of the zero-inflated Poisson models against the standard Poisson models are reported with the z-scores and corresponding p-values. All of the tests support the use of the zero-inflated model. Significance levels: *** p<0.01, ** p<0.05, * p<0.10

Appendix C: Experiment Instructions

Below we include the instructions for the experiment. Notes for the reader (not visible to subjects) indicate page breaks. These notes are contained in [brackets] and explain which instructions were viewed by subjects in each treatment. The instructions below are for the Sudoku experiment. The instructions for the Nonogram experiment were identical with "Nonogram" appearing everywhere "Sudoku" appears below; the only exception is that the portion of the instructions explaining the rules of the game varied accordingly, as indicated below.

[Introduction – All treatments]

You are now participating in a decision-making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in CASH at the end of the experiment.

Today's experiment will involve multiple tasks, and your total payment will be the sum of your payments from each task. Please do not talk to other subjects during the experiment. This set of instructions is for your private use only. In case of questions, please raise your hand. Then we will come to you and answer your questions privately.

[Dictator Game – All treatments]

In Part I of this experiment, you are matched with one other person in this room.

Each person begins with 5 Dollars. Each person chooses how to allocate this money between him/herself and the person he/she is paired with. To specify an allocation, please type the amount you want to allocate to yourself and the amount you want to allocate to the other person and then click "Next." The two amounts must sum up to 5 Dollars.

After everyone has chosen their allocation, the computer will randomly choose one person from each pair whose decision is implemented. Your payment for Part I will be based on the randomly chosen person's decisions.

For example, if your decision is randomly chosen to be implemented, then you will be paid according to your allocation. If instead, the other person's decision is chosen to be implemented, you will be paid according to their allocation.

This is the end of the instructions for Part I. If you have any questions, please raise your hand and an experimenter will answer them privately.

[Risk Preference – All treatments]

In this task, you will make a series of choices between two uncertain options. For each decision, all you have to do is indicate whether you prefer Option A or Option B and click the appropriate button with your mouse.

At the end, the computer will randomly pick ONE decision and then randomly draw a whole number between 1 and 100 to determine your payoff. Your payoff then depends on your choice of A or B in the randomly chosen decision and the number drawn by the computer.

If you choose option A, then numbers between 1 and 50 will always pay $1 and numbers between 51 and 100 will always pay $3.

If you choose option B, then the relationship between the number drawn and the payment is different in each decision, so make sure you carefully compare options A and B to make your choices.

In the first decision, Option B will pay $0 on numbers between 1 and 90 and $3 on numbers between 91 and 100. As you proceed through the decisions, Option B will change. The chance of receiving $3 will increase, while the chance of receiving $0 will decrease.

Remember, after you make all of your choices, the computer will randomly pick one of them to count and will draw a random number to determine your payment.

This means that each decision could be the decision-that-counts so it is in your interest to treat each decision as if it could be the one that determines your payment.

We will reveal the outcome of this task at the end of today's experiment, just before you are paid.

[Risk Preference Decision – All treatments]

Choose whether you prefer Option A or Option B.

At the end of the experiment, the computer will choose one of these decisions for actual payment and will then draw a random number between 1 and 100, which together with your choices, determines how much you get paid.

Remember, this could be the decision that counts, so consider your choice carefully!

[Sudoku General – All treatments]

The next task will involve solving puzzles known as Sudokus.

	6	4	3	2	
5	3			6	4
4					1
3					2
6	4			1	3
	1	3	4	5	

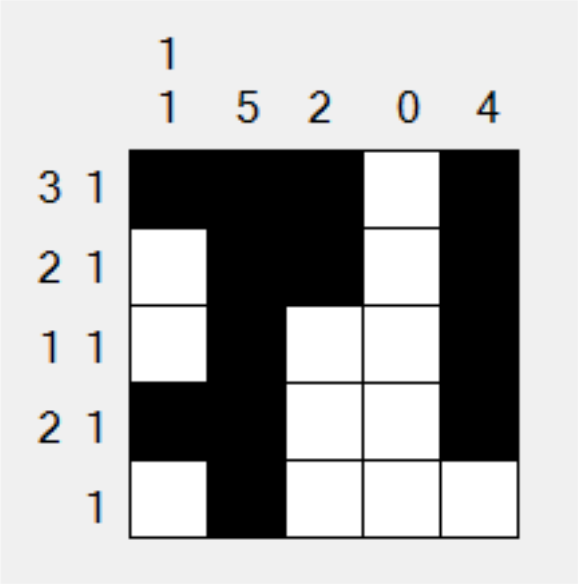
A 6×6 sudoku puzzle is a grid 6 squares wide and 6 squares deep. The lines of squares running horizontally are called rows, and the lines running vertically are called columns. The grid is further divided by the darker lines into six 2×3 rectangular 'boxes'.

Some of the squares already have numbers in them. Your task is to fill in the blank squares.

There's only one rule: Each row, column and box must end up containing all of the numbers from 1 to 6. This rule has an important side-effect, which is the basis of all solving techniques: Each number can only appear once in a row, column or box.

[Nonograms General – All treatments]

The next task will involve solving puzzles known as Nonograms.



A 5×5 Nonogram puzzle is a grid 5 squares wide and 5 squares deep. The lines of squares running horizontally are called rows, and the lines running vertically are called columns.

Each square must be either filled in black or left blank (or, equivalently, marked with an X).

Beside each row of the grid are listed the lengths of the runs of black squares on that row.

For example, the top row of the grid says 3,1. This means the row must contain a run of 3 consecutive black squares and a single black square.

Above each column are listed the lengths of the runs of black squares in that column.

For example, the leftmost column says 1, 1. This means the column must contain two (and only two) single black squares that are not consecutive.

Your aim is to find all black squares. Left click on a square to make it black. Left click a second time to mark it with an X. Left click a third time to turn it white again.

The puzzle is correct when all the rows and columns have the correct number of black squares in the proper positions.

[Sudoku Video – All treatments]

Before the experiment begins, we will watch the following video. Please pay attention to the video as you may find its content useful for the next portion of the experiment for which you will be paid.

This video describes strategies for solving 9×9 Sudokus, but they apply equally to 6×6 Sudokus.

[Ability Block – All treatments]

You will now be given 20 Sudoku puzzles to solve. You will have 10:00 minutes. You will be paid $0.5 for each puzzle that you correctly complete and $0 for each puzzle you do not complete or complete incorrectly. There is a timer in the upper right hand portion of the screen for your reference. Use the mouse to select cells. Numbers may be entered by clicking the buttons on screen or with the numbers on the keyboard. If you finish before the time limit, please wait quietly until the next puzzle begins.

[Review - *tracked* and *piece-rate*]

Now we will begin task 4. You will participate in two periods where you will attempt to solve more sudoku puzzles. The first of these periods will be for practice and the second will be for payment which will be added to your earnings at the end of the experiment.

Based on your performance in Part I, you have been sorted into groups. The top half performers are in one group and the bottom half performers are in the other group. The highest performers are those who correctly completed the most puzzles, with ties broken in favor of those who completed them the fastest

In the second period, you will be paid $0.5 for each puzzle that you correctly complete and $0 for each puzzle you do not complete or complete incorrectly.

[Review - *tracked* and *tournament*]

Now we will begin task 4. You will participate in two periods where you will attempt to solve more sudoku puzzles. The first of these periods will be for practice and the second will be for payment which will be added to your earnings at the end of the experiment.

Based on your performance in Part I, you have been sorted into groups. The top half performers are in one group and the bottom half performers are in the other group. The highest performers are those who correctly completed the most puzzles, with ties broken in favor of those who completed them the fastest

In the second period, your payment will be determined by your RELATIVE performance in your group. In other words, at the end of the second period we will count the number of puzzles completed correctly by each person, and pay will be based on your rank.

The participant in rank 1 will be paid $20
The participant in rank 2 will be paid $10
The participant in rank 3 will be paid $5
The participant in rank 4 will be paid $0

Ties will be broken in favor of those who completed the puzzles the fastest

[Review - *untracked* and *piece-rate*]

Now we will begin task 4. You will participate in two periods where you will attempt to solve more sudoku puzzles. The first of these periods will be for practice and the second will be for payment which will be added to your earnings at the end of the experiment.

Based on your performance in Part I, you have been sorted into groups. These groups have been balanced so that they contain both high and low performers. The highest performers are those who correctly completed the most puzzles, with ties broken in favor of those who completed them the fastest.

In the second period, you will be paid $0.5 for each puzzle that you correctly complete and $0 for each puzzle you do not complete or complete incorrectly.

[Review - *untracked* and *tournament*]

Now we will begin task 4. You will participate in two periods where you will attempt to solve more sudoku puzzles. The first of these periods will be for practice and the second will be for payment which will be added to your earnings at the end of the experiment.

Based on your performance in Part I, you have been sorted into groups. These groups have been balanced so that they contain both high and low performers. The highest performers are those who correctly completed the most puzzles, with ties broken in favor of those who completed them the fastest.

In the second period, your payment will be determined by your RELATIVE performance in your group. In other words, at the end of the second period we will count the number of puzzles completed correctly by each person, and pay will be based on your rank.

The participant in rank 1 will be paid $20
The participant in rank 2 will be paid $10
The participant in rank 3 will be paid $5
The participant in rank 4 will be paid $0

Ties will be broken in favor of those who completed the puzzles the fastest.

[Practice Block - *no-teaching*]

You will now be given a single Sudoku puzzle to solve. You will not be paid for this puzzle. You will have 10:00 minutes to solve the puzzle.

[Practice Block - *teaching*]

You will now be given a single Sudoku puzzle to solve. You will not be paid for this puzzle. You will have 10:00 minutes to solve the puzzle.

You can complete this puzzle working with the people in your group. During this period, your microphone will be enabled and a voice chat room will be available in which you can discuss the puzzle you are working on. You may discuss any aspects of the experiment in the chat room, but you may not reveal your identity, make threats, or use inappropriate language (including shorthand like WTF). Other

participants will be identified by a number next to their mouse cursor. This is their rank within the group. Please only speak English.

[Evaluation Block - *piece-rate*]

You will now be given 30 Sudoku puzzles to solve. You will have 15:00 minutes.

Remember, you will be paid $0.5 for each puzzle that you correctly complete, and you will be paid $0 for each puzzle you do not complete or complete incorrectly. At the end of the 15:00 minutes, you will learn the number of puzzles you correctly completed.

You will also learn your rank among the other subjects in terms of the number of correctly completed items in the previous 15:00 minutes.

[Evaluation Block - *tournament*]

You will now be given 30 Sudoku puzzles to solve. You will have 15:00 minutes.

At the end of this task, you will learn your rank among the other subjects in terms of the number of correctly completed items. Remember, in this task, your payment will be determined by your RELATIVE performance in your group. In other words, at the end of this task we will count the number of puzzles completed correctly by each person, and pay will be based on your rank

The participant in rank 1 will be paid $20
The participant in rank 2 will be paid $10
The participant in rank 3 will be paid $5
The participant in rank 4 will be paid $0

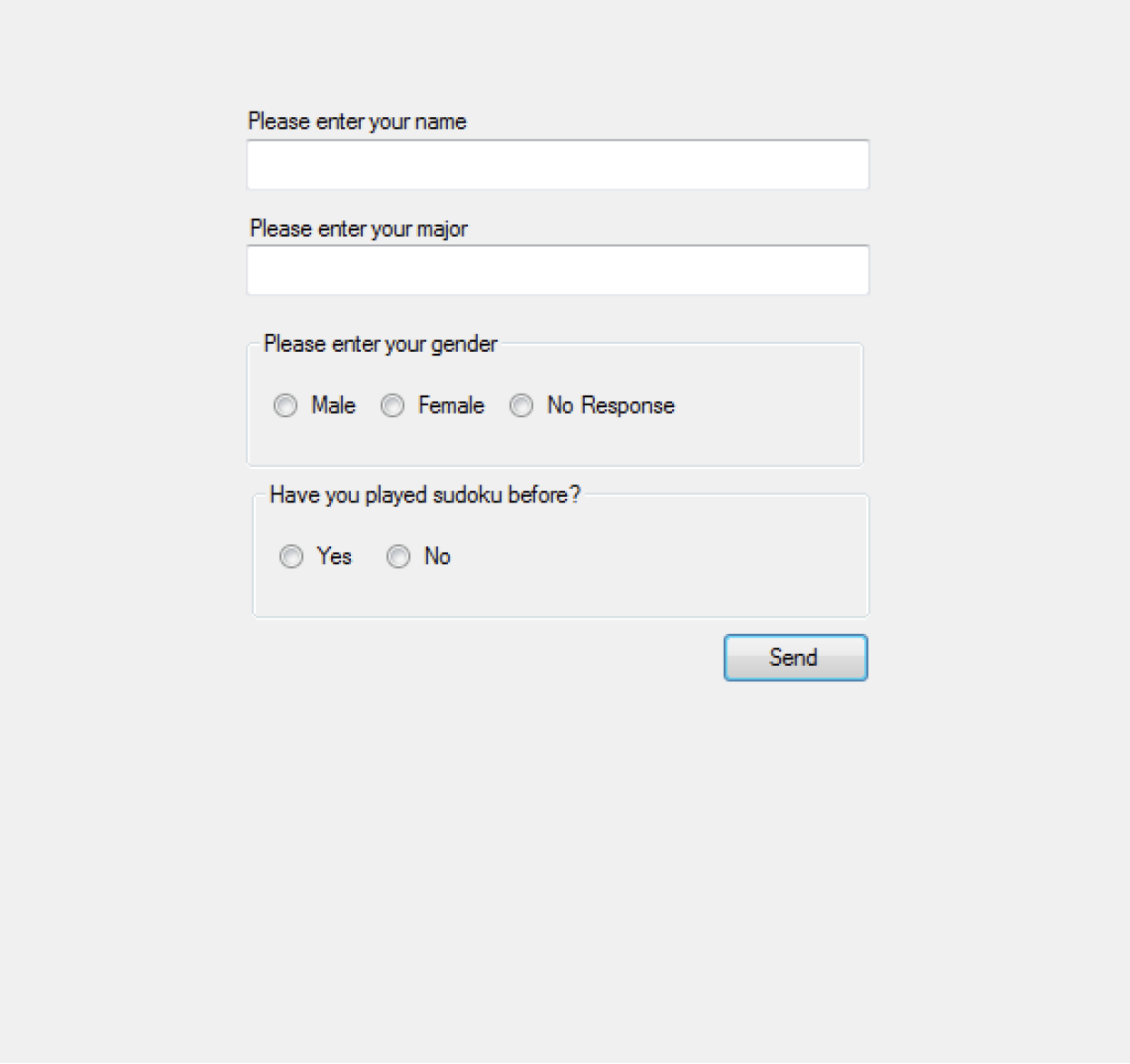
Ties will be broken in favor of those who completed the puzzles the fastest

[At the conclusion, subjects observe their payment from each task and are asked to wait quietly to be called over for payment]

Appendix D: Screenshots

Screenshots from each stage of the Experiment. Nonograms screenshots included only where they differ from Sudoku.

Page 1 – Demographics



Page 2 – General Instructions

You are now participating in a decision-making experiment. If you follow the instructions carefully, you can earn a considerable amount of money depending on your decisions and the decisions of the other participants. Your earnings will be paid to you in CASH at the end of the experiment.

Today's experiment will involve multiple tasks, and your total payment will be the sum of your payments from each task. Please do not talk to other subjects during the experiment. This set of instructions is for your private use only. In case of questions, please raise your hand. Then we will come to you and answer your questions privately.

Next

Page 3 – Dictator Allocation

In Part I of this experiment, you are matched with one other person in this room.

Each person begins with 5 Dollars. Each person chooses how to allocate this money between him/herself and the person he/she is paired with. To specify an allocation, please type the amount you want to allocate to yourself and the amount you want to allocate to the other person and then click "Next." The two amounts must sum up to 5 Dollars.

After everyone has chosen their allocation, the computer will randomly choose one person from each pair whose decision is implemented. Your payment for Part I will be based on the randomly chosen person's decisions.

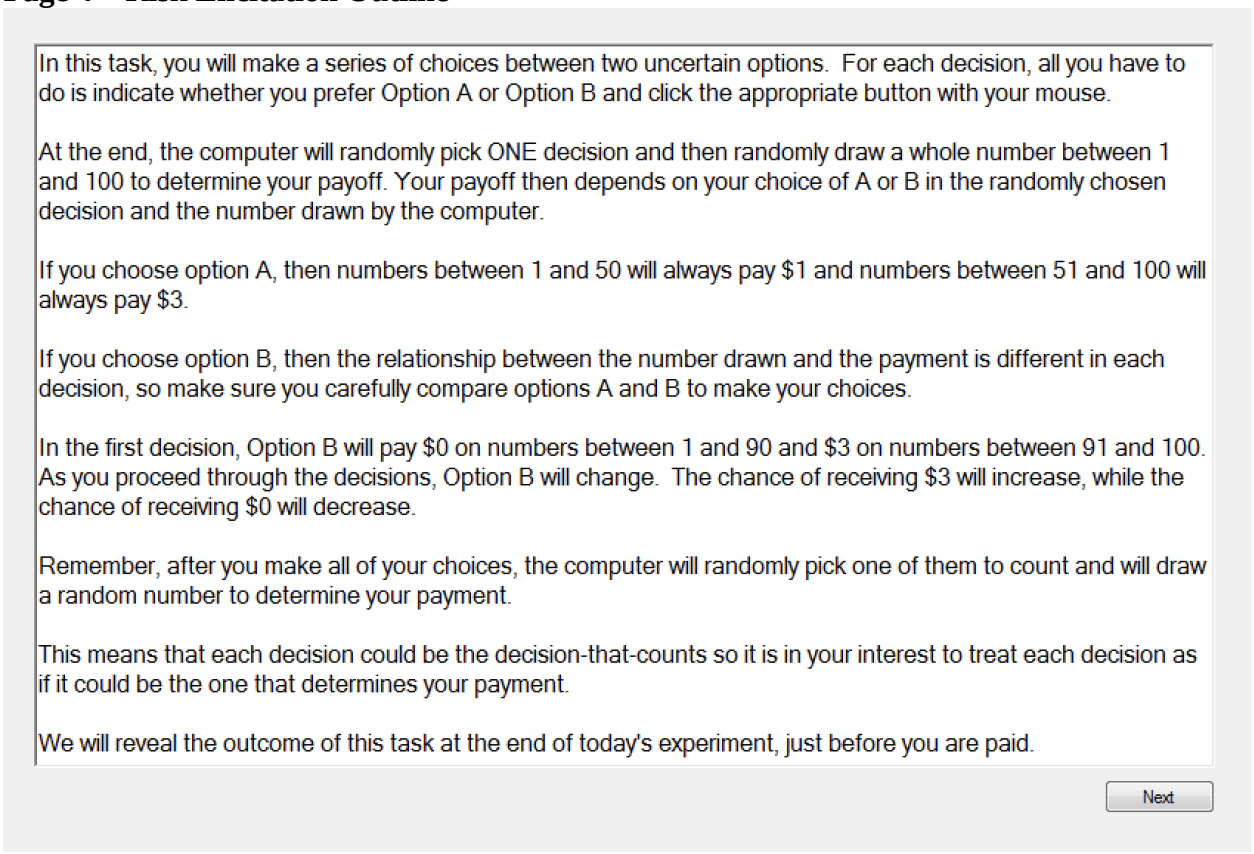
For example, if your decision is randomly chosen to be implemented, then you will be paid according to your allocation. If instead, the other person's decision is chosen to be implemented, you will be paid according to their allocation.

This is the end of the instructions for Part I. If you have any questions, please raise your hand and an experimenter will answer them privately.

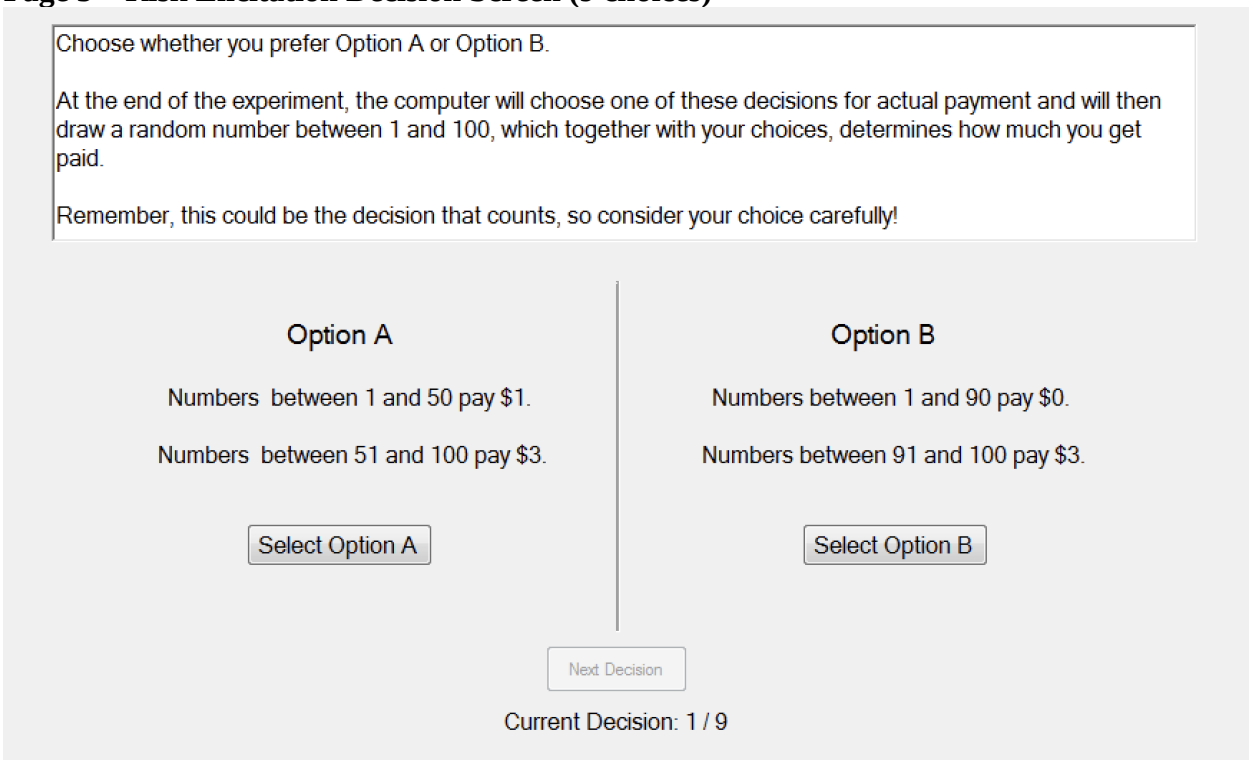
You Receive	They Receive
$	$

Next

Page 4 – Risk Elicitation Outline



Page 5 – Risk Elicitation Decision Screen (9 choices)



Page 6a – Basic Sudoku Instructions

The next task will involve solving puzzles known as Sudokus.

	6	4	3	2	
5	3			6	4
4					1
3					2
6	4			1	3
	1	3	4	5	

A 6x6 sudoku puzzle is a grid 6 squares wide and 6 squares deep. The lines of squares running horizontally are called rows, and the lines running vertically are called columns. The grid is further divided by the darker lines into six 2 X 3 rectangluar 'boxes'.

Some of the squares already have numbers in them. Your task is to fill in the blank squares. There's only one rule:

- Each row, column and box must end up containing all of the numbers from 1 to 6.

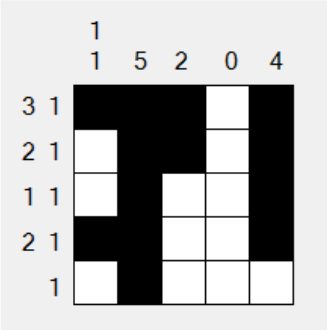
This rule has an important side-effect, which is the basis of all solving techniques:

- Each number can only appear once in a row, column or box.

Next

Page 6b – Basic Nonograms Instructions

The next task will involve solving puzzles known as Nonograms.



A 5x5 Nonogram puzzle is a grid 5 squares wide and 5 squares deep. The lines of squares running horizontally are called rows, and the lines running vertically are called columns.

Each square must be either filled in black or left blank (or, equivalently, marked with an X).

Beside each row of the grid are listed the lengths of the runs of black squares on that row.

- For example, the top row of the grid says 3,1. This means the row must contain a run of 3 consecutive black squares and a single black square.

Above each column are listed the lengths of the runs of black squares in that column.

- For example, the leftmost column says 1, 1. This means the column must contain two (and only two) single black squares that are not consecutive.

Your aim is to find all black squares. Left click on a square to make it black. Left click a second time to mark it with an X. Left click a third time to turn it white again.

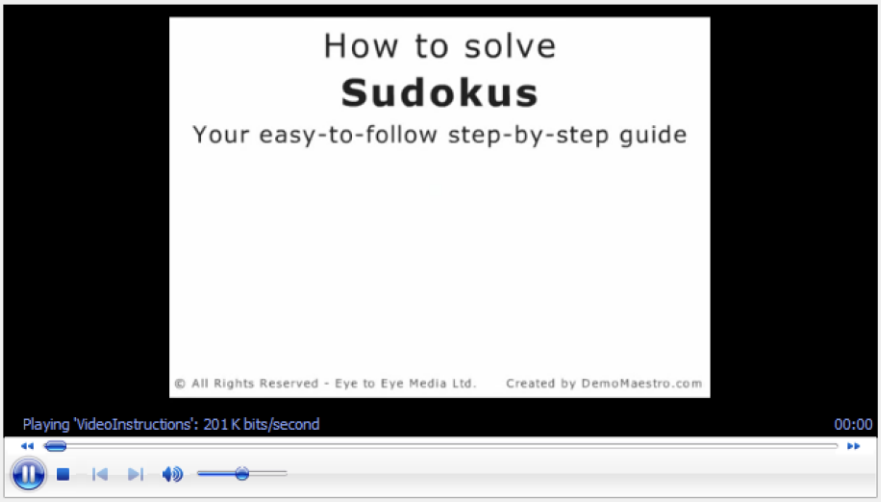
The puzzle is correct when all the rows and columns have the correct number of black squares in the proper positions.

Next

Page 7a – Sudoku Instructional Video

Before the experiment begins, we will watch the following video. Please pay attention to the video as you may find its content useful for the next portion of the experiment for which you will be paid .

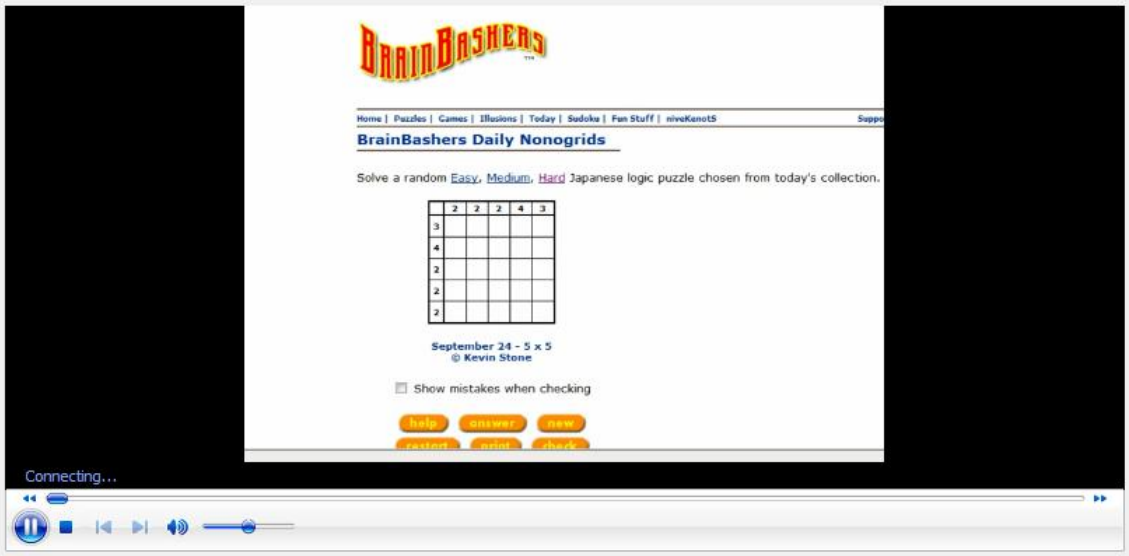
This video describes strategies for solving 9x9 sudokus, but they apply equally to 6x6 sudokus.



Please Watch The Video

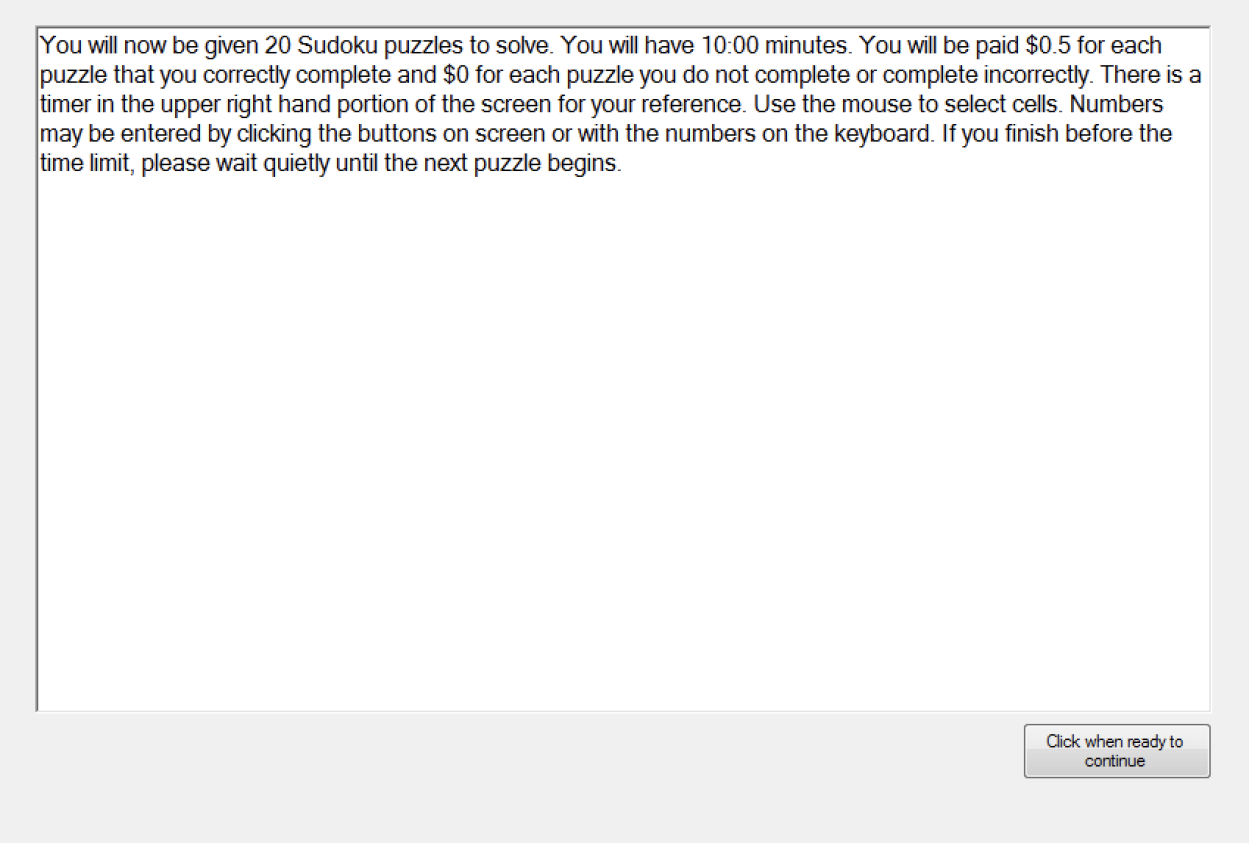
Page 7b – Nonogram Instructional Video

Before the experiment begins, we will watch the following video. Please pay attention to the video as you may find its content useful for the next portion of the experiment for which you will be paid .

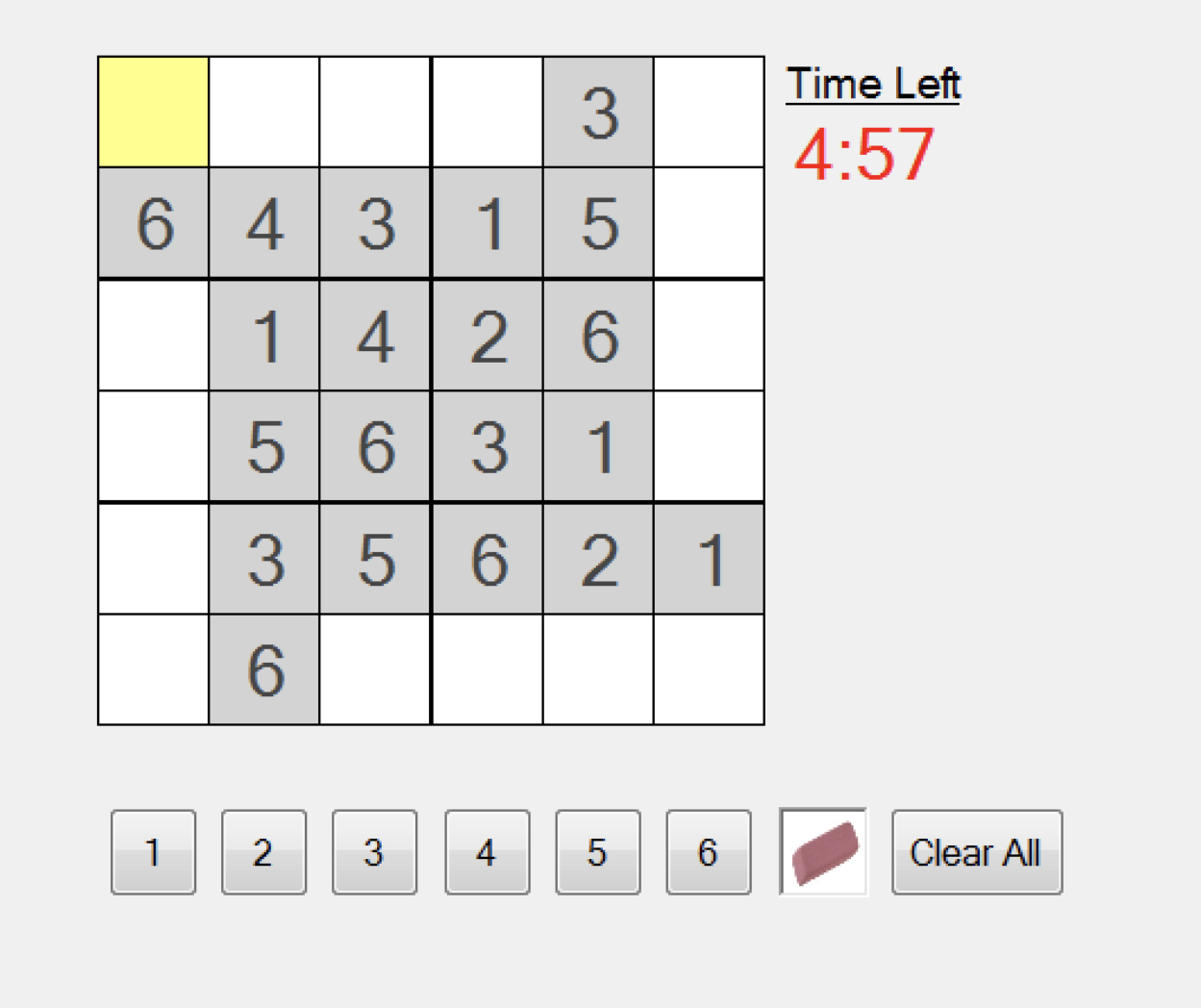


Please Watch The Video

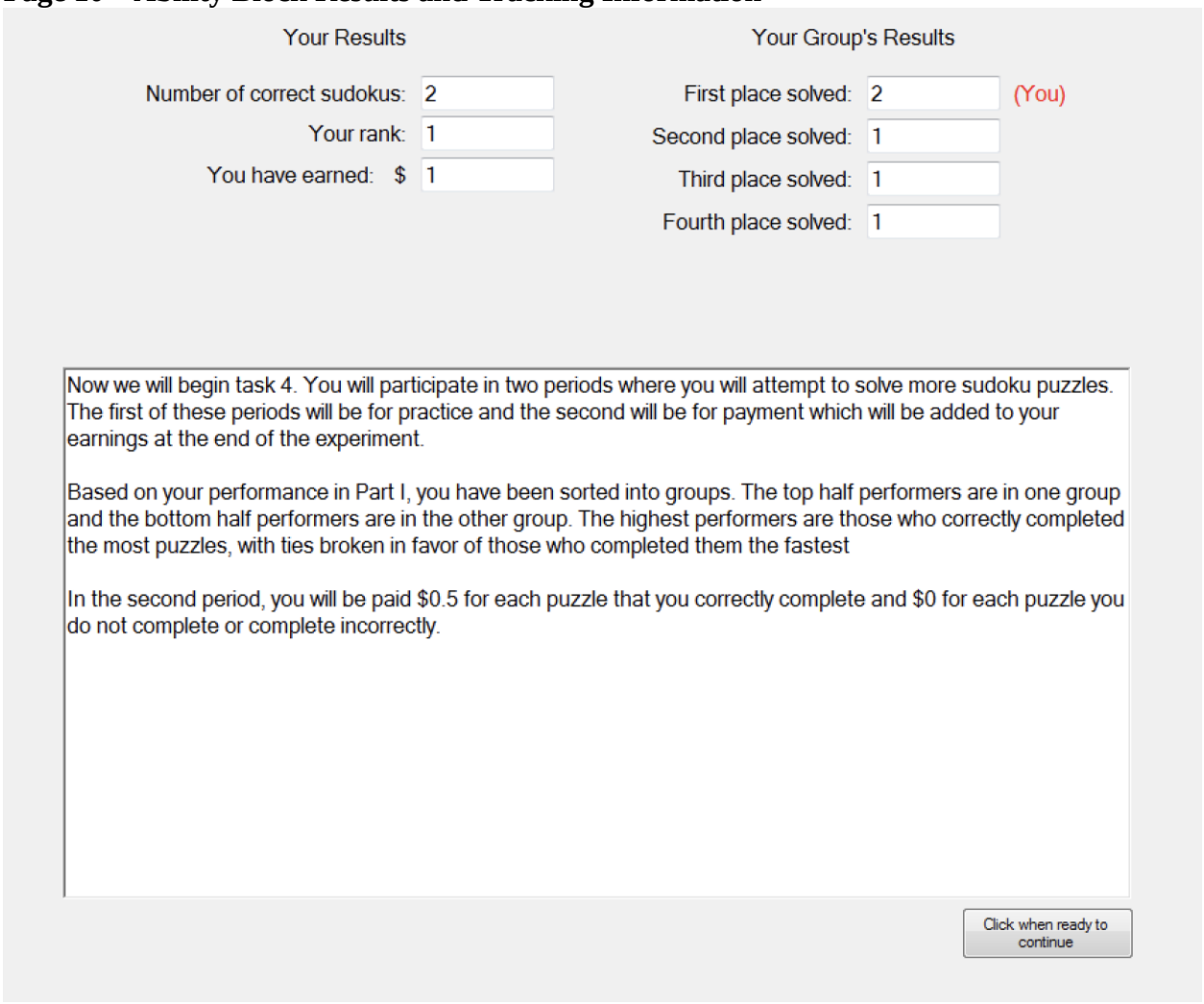
Page 8 – Ability Block Instructions



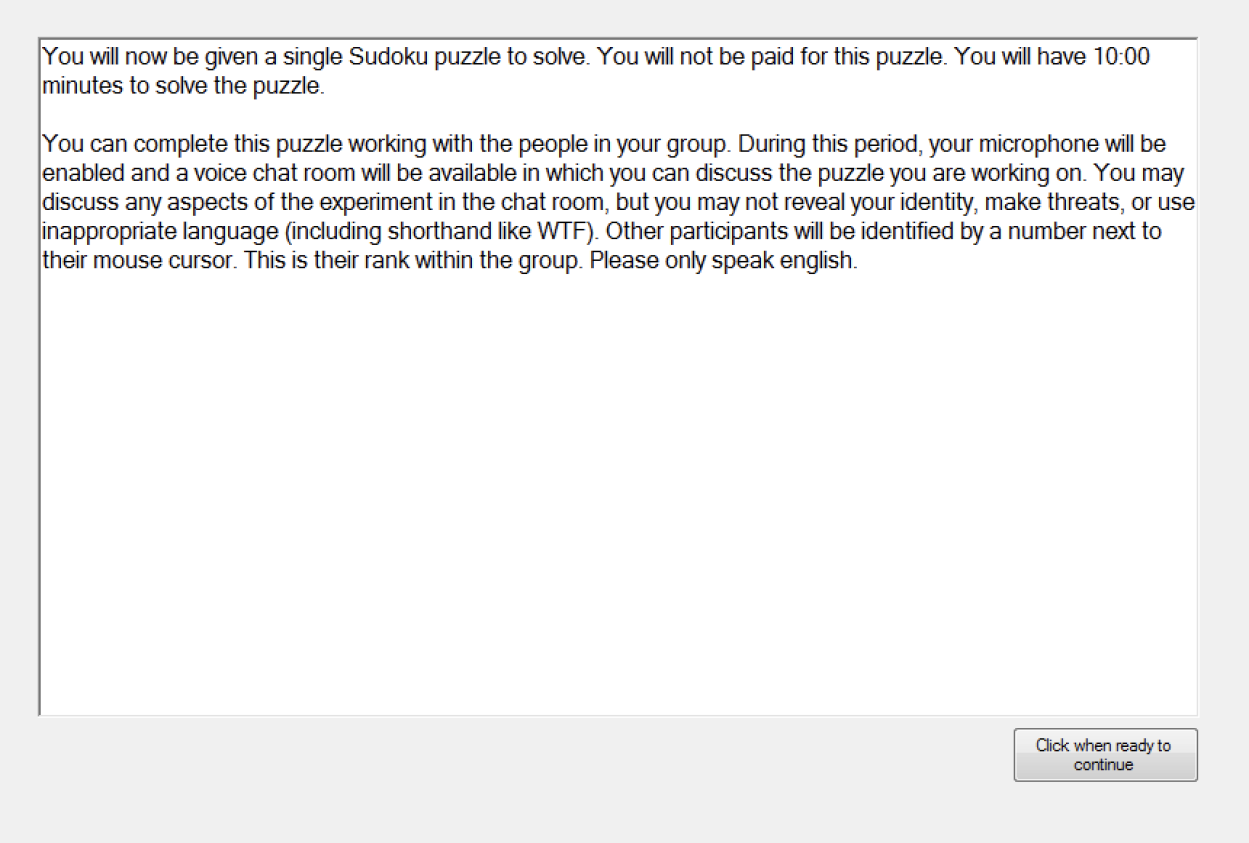
Page 9 – Ability Block Screen



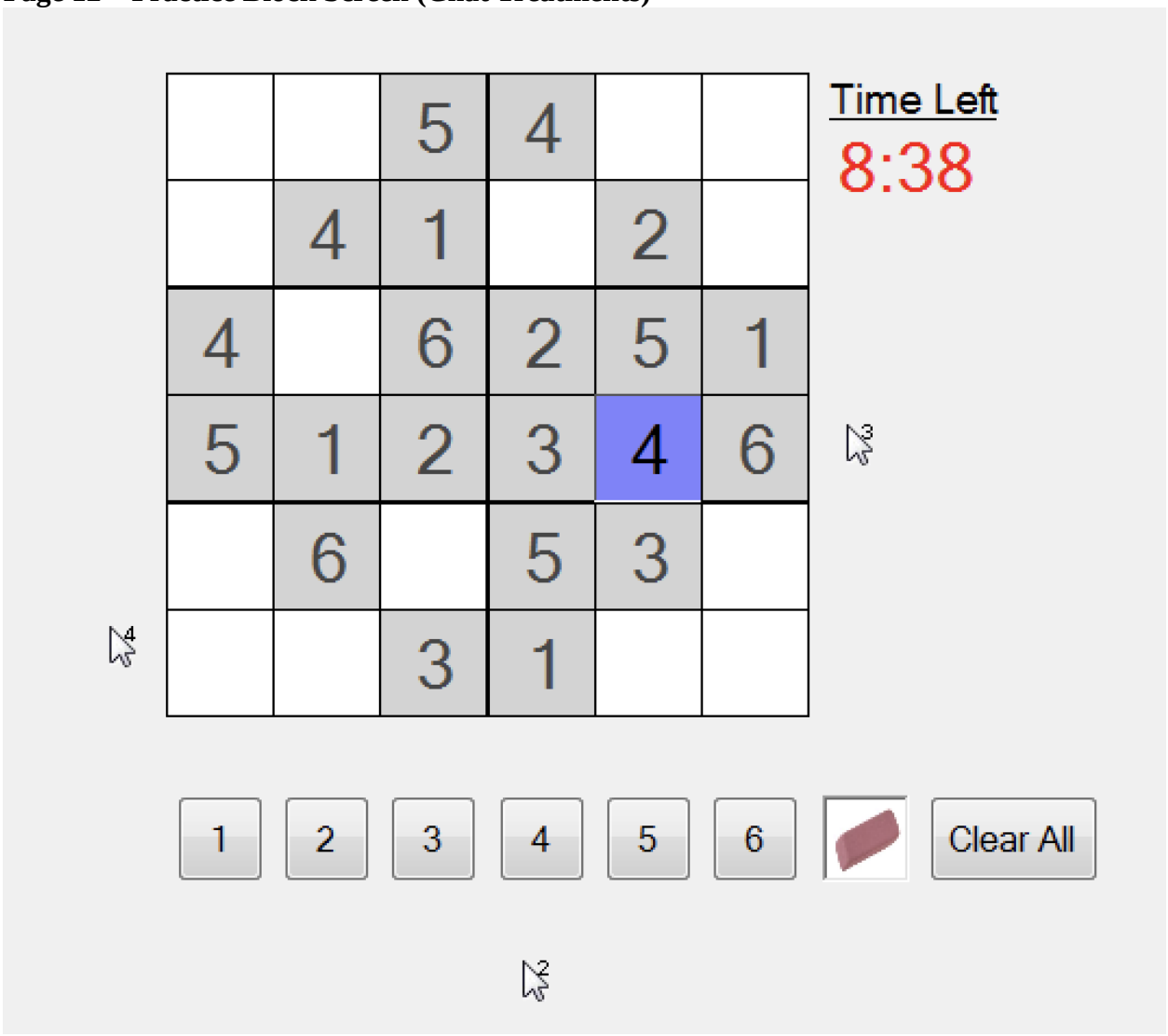
Page 10 – Ability Block Results and Tracking Information



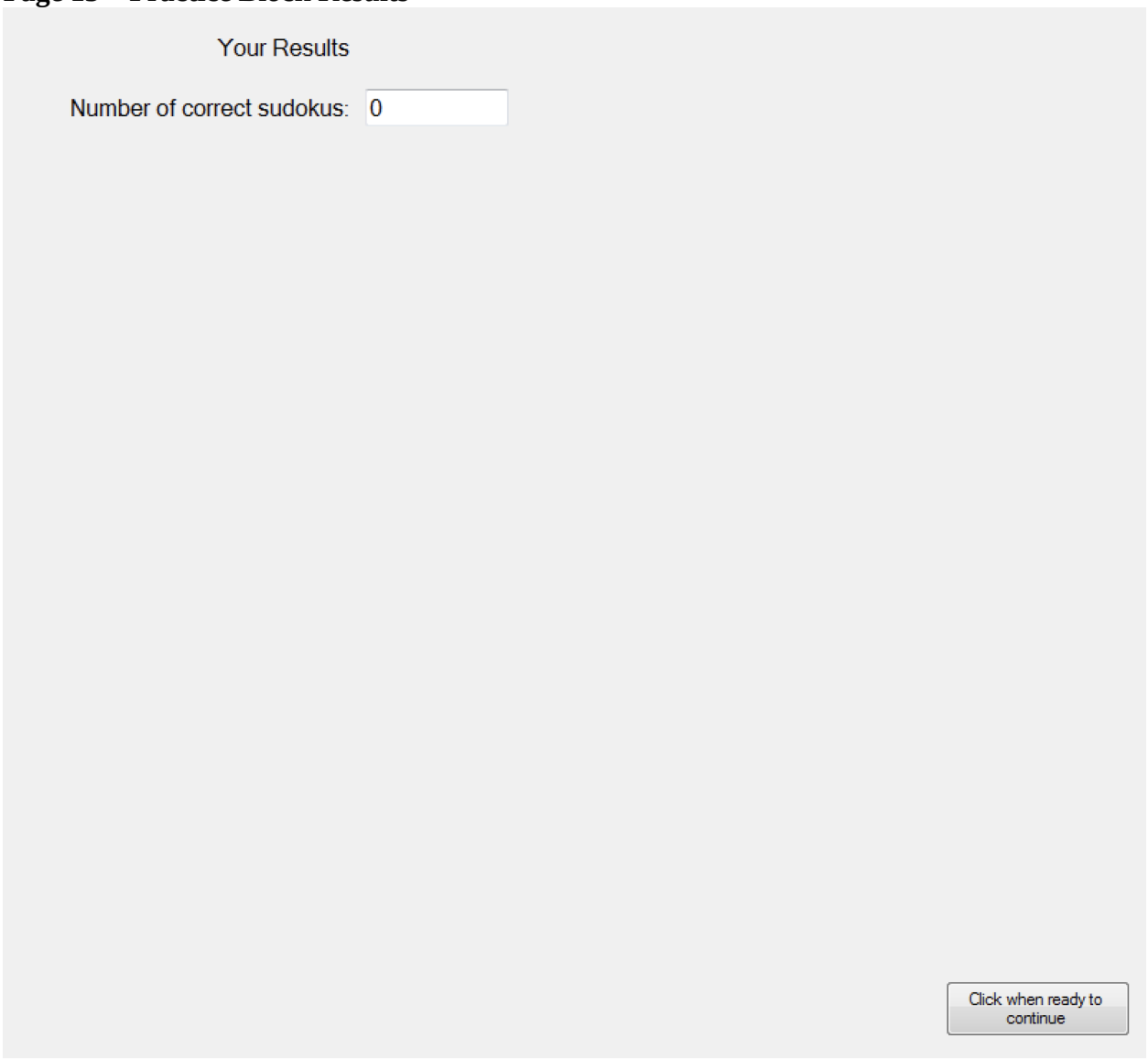
Page 11 – Practice Block Instructions (Chat Treatments)



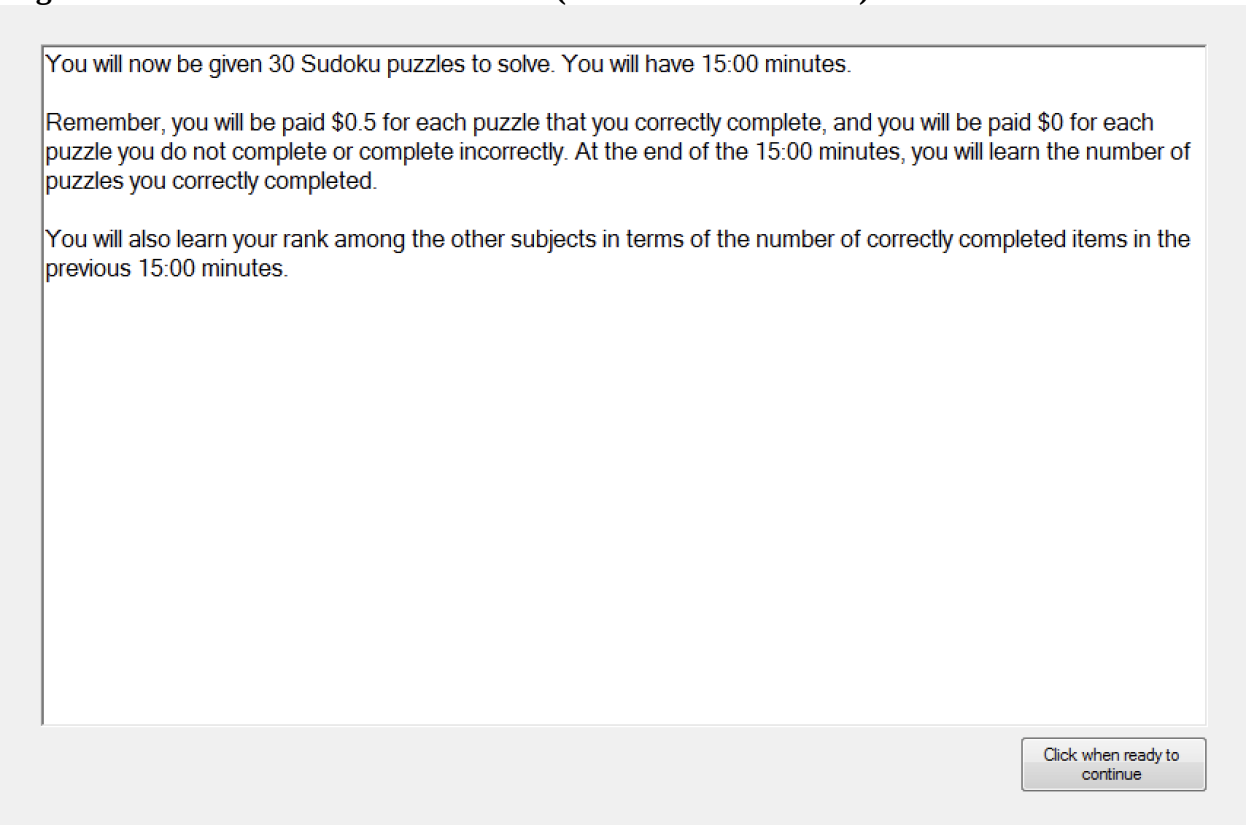
Page 12 – Practice Block Screen (Chat Treatments)



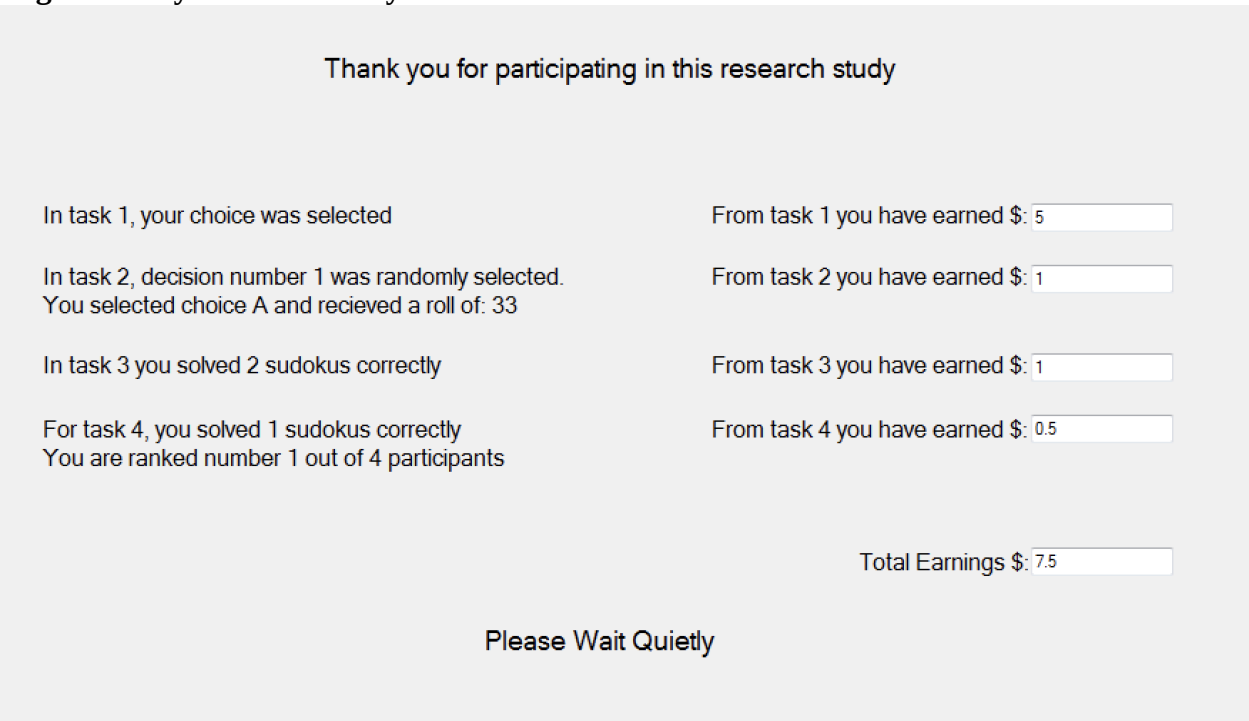
Page 13 – Practice Block Results



Page 14 – Evaluation Block Instructions (Piece Rate Treatment)



Page 15 – Payment Summary Screen



Appendix E: Methods Used to Count Instances of Teaching in Audio Data

Two research assistants were given audio files for each group in each session of the *teaching* treatments. We explained that subjects were working together on a single Sudoku puzzle and that the chat transcript provided a record of this interaction. We instructed them to come up with a count of the number of instances of peer-to-peer teaching in the text. To define teaching, they adapted a scheme developed by Kline (2016) that identifies teaching with behavior related to the relevant task in which the following events occur:

- **+/- verbal feedback**- Actor gives positive or negative verbal appraisal of F's behavior with respect to, in F's presence.
- **+/- consequences**- Actor creates positive or negative consequences for F. Includes physical punishment or reward, or verbal description of promised punishment or reward.
- **teasing**- Actor threatens or punishes but in a joking manner, as indicated by smiling and/or laughing. Verbal or gestural.
- **warning (of danger)**- Verbal warning of danger, a separately coded subset of negative verbal appraisal.
- **command to stop**- Verbal command to focal to stop, a separately coded subset of negative verbal appraisal.
- **command to say/do [x]**- Verbal command to say a given phrase, or repeat a particular phrase or gesture, a separately coded subset of positive verbal appraisal.
- **direct attention to object-** Actor directs F's attention toward an object or location, verbally or through gesture
- **direct attention to person-** Actor directs F's attention toward another person, verbally or through gesture.
- **command to watch**- Actor directs F's attention toward watching the actor.
- **other-prompted behavior-** F undertaking an action or behavior after being commanded to do so by another actor.
- **other-assisted behavior**- F undertaking an action or behavior made possible by another actor's help.
- **abstract communication-** Actor gives a verbal explanation or states abstract information to F – including the statement of rules and what is "taboo" behavior.
- **demonstration**- Actor performs behavior or action conditioned on F's attention; may follow a request from F, or the Actor may first manipulate F's attention.

After coding each session independently, our research assistants met to resolve any inconsistencies and produced a count of teaching instances for each group in each session.

A few other notes:

1) A single instance of teaching would include both a question asked by a subject and the answer received.

2) If no answer was received to a question, this was *not* counted as an instance of teaching.
3) We had hoped to get individual level data on teaching, but unfortunately it was not consistently possible to identify specific individuals by their voices and match them to the data.

References:

Kline, M.A. 2016. “TEACH: An Ethogram-based Method to Observe and Record Teaching Behavior.” *Field Methods.* doi:10.1177/1525822X16669282