Appendix for: How Hard to Fight? Strategic Sophistication in an Asymmetric Contest Experiment

April 18, 2017

Further Analysis of Comparative Statics and Treatment Effects

The main manuscript specification used interaction terms to estimate treatment-condition-specific comparative statics. Pooling the data across treatment conditions yields similar results, namely strong support for the own value and getting deterred predictions, but weaker support for the doing the deterring predictions. Table A1 shows these results. There is a positive, significant coefficient for the Own Value effect. Also consistent with predictions, there is a negative and significant coefficient for the Getting Deterred variable. However, there is a negative coefficient for Doing the Deterring in the single valuation rounds, which is inconsistent with predictions; this coefficient is positive and insignificant for double valuations.

The main manuscript also estimated the effects of each treatment condition on the amount of over-effort, relative to the Nash prediction (Table 3). It included some controls for the Nash effort level, double valuations, experience, and zero value effort. Here, we also include an indicator for male subjects and measures of aggression and riskiness. Table A2 shows these results. For the main treatment effects, the results are very similar in sign, significance, and magnitude to those

Table A1: Comparative statics, pooled across treatment conditions

	Single Val. b/se	Double Val. b/se
Own Value	210.78**	295.60**
	(9.76)	(14.28)
Getting	-28.14**	-63.74**
	(6.96)	(8.46)
Doing	-22.82*	13.05
	(11.48)	(17.72)
Constant	109.80**	204.51**
	(7.95)	(9.36)
N	2400	2400
R^2	0.30	0.37

+ $p < .10$, * $p < .05$, ** $p < .01$

reported in the main text. Among the three variables, aggression and being male had negative effects on effort that were significant in one specification apiece. Apart from those two results, none reached conventional levels of significance.

Feedback and Learning/Confounding

Since the feedback treatment occurs in later rounds in Sessions 1-4, we need to rule out the possibility that the feedback treatment effects are confounded by learning and an increasing familiarity with the game as the rounds progress.[1] If subjects tended to decrease their amount of effort in later rounds, then this might mistakenly attribute a decrease in the distance from the Nash prediction, even without the feedback treatment effect.

In the main text, we showed analysis where we limited regressions to only the first part of each experimental session, as in columns 3-4 of Table 3. This allows us to compare behavior with and without feedback, across the same rounds and time periods. The feedback treatment still has a

[1] We again thank our reviewers for highlighting this issue.

Table A2: Treatment effects with additional controls

	All Data b/se	Part 1 only b/se	Part 2 after BN b/se
Feedback	-68.81**	-54.35*	-133.89**
	(17.94)	(20.85)	(45.98)
Calculator	-44.82+	-37.94	
	(23.30)	(23.02)	
Feed. X Calc.	56.35*	50.55+	
	(23.48)	(29.47)	
Nash effort	-0.13**	-0.20**	-0.05
	(0.04)	(0.04)	(0.05)
Double valuation	-10.94*	-14.21*	-12.32
	(4.91)	(6.23)	(10.25)
Experience	-1.01*	-3.61**	-3.63*
	(0.39)	(0.98)	(1.63)
Zero value effort	0.23**	0.32**	0.11*
	(0.05)	(0.06)	(0.04)
Male	-27.94+	-32.63+	-46.64
	(15.62)	(18.34)	(31.11)
Risk Scale	-2.79	-4.51	-35.84
	(25.74)	(30.58)	(42.66)
Aggr. Scale	-13.84	-17.95	-70.64+
	(21.30)	(26.18)	(38.67)
Constant	169.48**	199.16**	255.32**
	(26.47)	(27.98)	(63.02)
N	4,800	2,400	928
R^2	0.06	0.06	0.09

+ $p < .10$, * $p < .05$, ** $p < .01$

negative, large and significant effect on decreases the distance from the Nash prediction. We also limited our analysis only to (a) sessions which had no feedback or calculator in the first part, BN, and (b) the second part of those sessions. In other words, we can limit analysis the second part of sessions 1,4,6, and 8. This subset of the data allows us to look at the effect of feedback in later rounds, holding constant that every player has already played 16 rounds in the BN condition. This analysis was in in columns 5-6 of Table 3.

Additional analysis also confirms these results. Here, we add two additional checks.

Line figures

First, we looked for whether the treatment effect for Feedback was solely driven by later rounds, or if there were differences in behavior across feedback/no feedback early on. Figure A1 shows the percentage of overbidding by period, for the feedback and no feedback conditions. Even in the early rounds, there is a negative treatment effect, with over-effort levels being lower in the feedback treatments compared to those without feedback. The difference is large and apparent even in early rounds such as 7 and 8.[2]

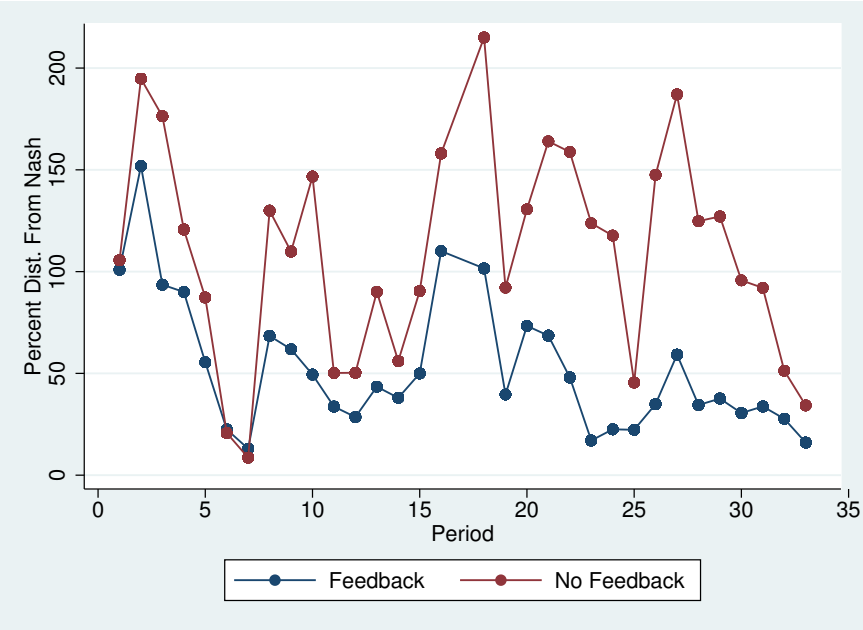
Discontinuity

To assess this, we look for a discontinuity in behavior before and after the feedback treatment. If learning explains the change in outcomes, then we should not see a discontinuity. The rate at which behavior converges towards Nash predictions should be steady before and after the treatment. If there is a jump, and behavior gets most closer to Nash predictions after the treatment, then this would suggest that the treatment effect is not an artifact of learning.

Figure A2 shows the percent distance from Nash predictions by period, with Lowess smoothers before and after the feedback treatment. Note that the treatment begins in Round 18, but since the

[2]There is a sequence of players' valuations that is set across sessions. That is why the lines move up and down.

Figure A1: Percent Distance from Nash by Round, Plus/Minus Five Round Window

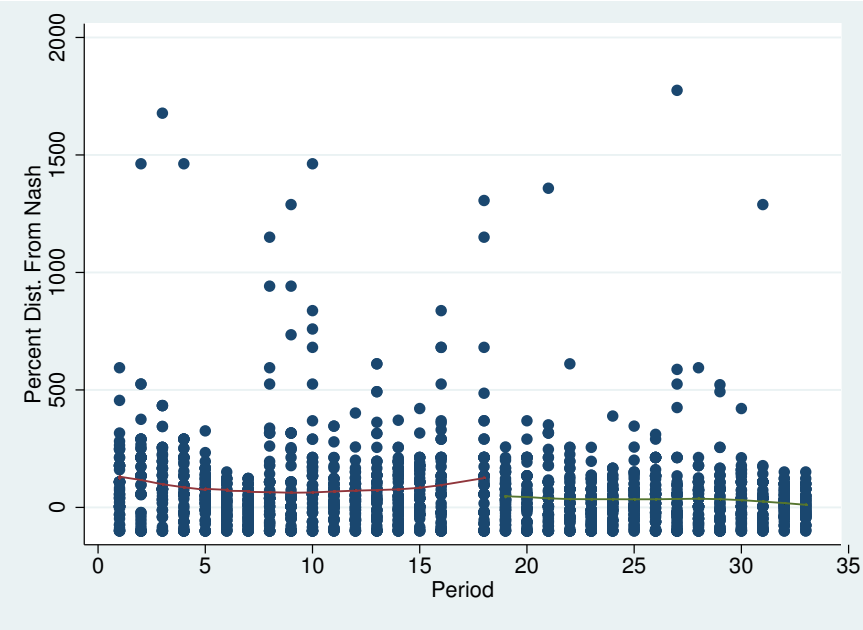


feedback is only provided after participants choose their effort levels, the treatment is administered *after* they make their Round 18 choice. That is why the left side Lowess line includes the efforts from Round 18. There is a slight decline in distance from Nash predictions over time, but there is a distinct jump downwards after the feedback treatment is administered. This jump is also apparent in Figure A3 which zooms in on the break point, only including Rounds 12-23. The distinct break supports the argument that the feedback treatment effect is not simply an artifact of learning over time.

Strategic Sophistication and the Level k Model

To demonstrate how our conception of strategic sophistication is different from that of the level k model, Figure A4 shows a each subject's level of sophistication as measured by the level k model, broken down by whether the subject's behavior tended to be consistent with all, none, or one of

Figure A2: Percent Distance from Nash by Round



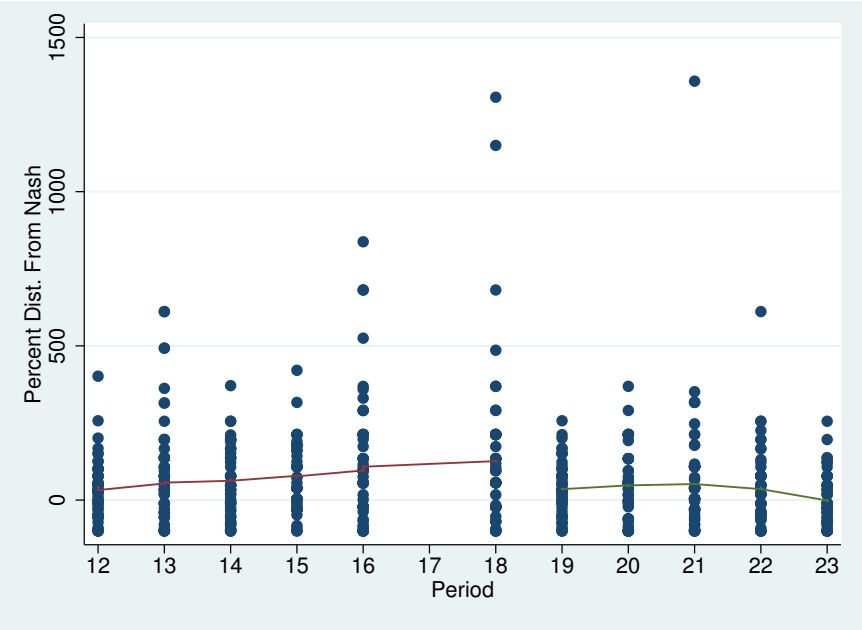
the comparative static predictions.[3] The subjects' levels are poorly correlated with the degree to which the subject displayed behavior consistent with comparative static predictions. The subjects whose behavior was consistent with both comparative static predictions are only estimated to play at a very slightly higher level of strategy according to the level k model. Among the subjects whose behavior was consistent with none of the comparative static predictions, the average level was 2.097. Among the subjects whose behavior was consistent with at least one of the comparative statics, their estimated level only very slightly higher, 2.103.

Variation in Search Quality

Motivated by these ideas, we construct several measures of search quality. For a set of minimal measures of search quality, we code whether each click or guess yields net positive expected utility,

[3] We calculated levels by assuming that a level zero player randomized between zero and her valuation. The results are similar if we assume that level zero players randomize over the interval zero to 1,000, the maximum tickets they can buy.

Figure A3: Percent Distance from Nash by Round, Plus/Minus Five Round Window



EU_i, relative to purchasing 0 tickets and ensuring a payoff of 1000 points. As described in the main text, Own Positive indicates whether $EU_i > 1000$ for subject i (the subject using the calculator), Opponent Positive indicates whether the guess yields positive expected utility $EU_j > 10000$ for i's opponent j, and Both Positive indicates searches where both $EU_i > 1000$ and $EU_j > 1000$.

Another measure of search quality relates to the direction of search. Let $g_k = (e_{ik}, e_{jk})$ denote subject i's k-th guess in any given round. The direction of search refers to the angle of the difference vector $\Delta g = g_{k+1} - g_k$, which we measure in degrees (from 0° to 360°). If a subject searches the strategy space by holding the opponent's effort constant $e_{j,k} = e_{j,k+1}$ while varying her own effort $e_{i,k} \neq e_{i,k+1}$, the direction of search will be *horizontal*. Conversely, if a subject holds her own effort constant $e_{i,k} = e_{i,k+1}$ while varying her guesses about her opponent's effort $e_{j,k} \neq e_{j,k+1}$, the direction of search will be *vertical*. Horizontal searches reflect a subject's attention to her own payoffs, which is individually rational in the sense of maximizing one's own payoffs, while vertical searches reflect attention to her opponent's payoffs and reflect strategic rationality in the sense of forming rational expectations about opponent behavior. We allow for two levels of error tolerance

Table A3: Measures of search quality

	Total		Subject-period		Subject	
	Mean	N	Mean	N	Mean	N
Own Positive	53	12,010	.59	1,179	.58	57
Opponent Positive	.50	12,010	.53	1,179	.52	57
Both Positive	.32	12,010	.35	1,179	.33	57
Horizontal ($\pm 10°$)	.30	10,831	.30	1,034	.28	57
Horizontal ($\pm 22°$)	.40	10,831	.40	1,034	.37	57
Vertical ($\pm 10°$)	.23	10,831	.24	1,034	.25	57
Vertical ($\pm 22°$)	.32	10,831	.32	1,034	.32	57
Distance	121.7	10,831	172.9	1,034	185.4	57
Searches	–	–	10.2	1,179	20.7	57

in how we classify horizontal and vertical searches, with a relatively narrow tolerance of $\pm 10°$ and a wider tolerance of $\pm 22°$. We then code each guess after the first ($k > 1$) as horizontal, vertical, or diagonal (neither horizontal nor vertical).

We find that the quality of subjects' searches according to these measures tends to be fairly poor. Table A3 describes the averages for our measures of search quality along with the distance between each guess and the total number of guesses. We present the overall means, the subject-period level means, and the subject-level means. The results do not differ much by level of aggregation.

According to our positive expected payoff measures, at most half of subjects' searches in the *Calculator* treatment can be classified as minimally rational. 53.1% of guesses involve positive expected values for the subject's own payoffs and 50.8% of guesses involve positive expected values for their opponent's payoffs. However, fewer than one-third of guesses (32%) involve positive expected payoffs for both the subject and their opponent. While we would expect to see that initial searches within a period yield net negative expected payoffs, we also thought that minimally rational search behavior would move quickly towards areas of the strategy space where both players receive positive expected utility. The prevalence of negative expected payoff guesses suggests to us that most searches are of low quality.

We also find that horizontal and vertical searches comprise half of the guesses entered into the calculator. While we might expect some searches to be diagonal, systematic guesses along one of the dimensions to search for a player's best reply appear to be rare. Searches along one dimension also tend to be horizontal (31% of all searches using the 10° tolerance) rather than vertical (20%), which suggests that subjects tend to focus on their own payoffs rather than their opponents. This may reflect a failure of subjects to engage in any kind of meaningful strategic reasoning.

To assess whether the quality of search affects behavior, we estimate several regression models with our search measures as right-hand side variables. Table A4 to Table A6 show various specifications for these regressions. The first table uses all parts of all sessions that included a calculator. The second and third tables limit analysis to sessions without and with feedback, respectively. For each table, the first column uses the Own/Opponent/Both Positive variables. The second column uses the number of searches with each property, as opposed to the mean. The third column uses the variables describing the direction of the search. The fourth column uses variables describing the total amount of searching the player conducted as well as the distance she covered in her search. The final column uses the Own/Opponent/Both Positive variables and the search direction variable.

The variables indicating searches in the Both Positive region consistently have negative coefficients and are statistically significant in most specifications. This indicates that subjects searching in this region generally exerted less over-effort compared to subjects who searched in the regions where only one player (or neither player) received a positive payoff. This is consistent with the idea that better searching leads to better play.

The variables indicating vertical and horizontal searches have positive coefficients. Players who searched only in one dimension, as opposed to diagonal searches that varied both players' effort levels, tended to exert higher levels of over-effort. This is also consistent with the idea that better searching yields better play, although these results were not statistically significant.

More extensive searching, either in terms of distance or the number of clicks, did not generally improve play. Players searching a greater distance exerted higher degrees of over-effort. The total

number of clicks had an inconsistent effect on over-effort.

Table A4: Effect of search quality on effort (all)

	(1) b/se	(2) b/se	(3) b/se	(4) b/se	(5) b/se
Feedback	-10.83	-12.37	-11.02	-10.83	-7.13
	(15.54)	(14.77)	(16.10)	(14.71)	(14.74)
Nash effort	-0.06	-0.09+	-0.13*	-0.13*	-0.05
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Double val.	-8.18	-5.50	-3.98	-5.63	-9.60
	(8.17)	(8.44)	(8.92)	(9.06)	(8.21)
Experience	-0.71	-0.76+	-0.88+	-0.69	-0.57
	(0.45)	(0.45)	(0.50)	(0.46)	(0.44)
My Pos. Search	31.13*				1.29
	(12.61)				(25.45)
Opp. Pos. Search	35.57				15.27
	(22.05)				(27.72)
Both Pos. Search	-123.69**				-108.85**
	(24.26)				(31.69)
My Pos. Search (num)		0.67			
		(0.47)			
Opp. Pos. Search (num)		0.58			
		(0.92)			
Both Pos. Search (num)		-4.83**			
		(1.46)			
Horiz. (10 deg.)			7.47		43.43
			(20.59)		(32.24)
Vert. (10 deg.)			21.01		45.90*
			(18.09)		(22.11)
Distance				0.11*	0.11*
				(0.05)	(0.05)
Total clicks				-0.27	-0.33
				(0.38)	(0.38)
Constant	93.31**	100.89**	98.73**	91.83**	79.16**
	(29.40)	(26.76)	(28.34)	(24.90)	(25.28)
N	1,856	1,856	1,711	1,711	1,711
R^2	0.05	0.03	0.02	0.03	0.07

+ $p < .10$, * $p < .05$, ** $p < .01$

Table A5: Effect of search quality on effort (no feedback)

	(1)	(2)	(3)	(4)	(5)
	b/se	b/se	b/se	b/se	b/se
Nash effort	-0.14	-0.16	-0.19	-0.17+	-0.10
	(0.08)	(0.10)	(0.11)	(0.10)	(0.08)
Double val.	-12.96	-12.83	-13.79	-18.69	-18.96
	(12.21)	(12.62)	(10.93)	(13.97)	(12.21)
Experience	-5.01+	-4.69+	-4.79+	-3.83+	-4.06+
	(2.46)	(2.33)	(2.56)	(2.12)	(2.20)
My Pos. Search	15.69				-8.91
	(29.42)				(51.35)
Opp. Pos. Search	-17.57				-20.90
	(28.33)				(39.36)
Both Pos. Search	-77.67*				-76.23
	(33.24)				(46.36)
My Pos. Search (num)		0.80			
		(1.02)			
Opp. Pos. Search (num)		-1.28			
		(2.21)			
Both Pos. Search (num)		-3.92			
		(3.17)			
Horiz. (10 deg.)			25.28		68.43
			(46.01)		(70.23)
Vert. (10 deg.)			24.34		49.57
			(32.42)		(36.23)
Distance				0.16	0.16
				(0.14)	(0.14)
Total clicks				-1.79	-1.80
				(1.31)	(1.37)
Constant	159.39**	151.35**	138.44**	135.72**	129.04**
	(56.27)	(47.53)	(40.22)	(29.86)	(30.99)
N	448	448	408	408	408
R^2	0.08	0.06	0.04	0.07	0.12

+ $p < .10$, * $p < .05$, ** $p < .01$

Table A6: Effect of search quality on effort (feedback)

	(1)	(2)	(3)	(4)	(5)
	b/se	b/se	b/se	b/se	b/se
Nash effort	-0.03	-0.07	-0.11+	-0.11+	-0.03
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Double val.	-8.04	-4.41	-2.96	-4.09	-9.24
	(9.39)	(10.19)	(10.99)	(10.83)	(9.77)
Experience	-0.40	-0.48	-0.62	-0.50	-0.37
	(0.47)	(0.48)	(0.55)	(0.51)	(0.50)
My Pos. Search	31.38**				5.87
	(11.43)				(19.63)
Opp. Pos. Search	54.81+				37.55
	(29.12)				(36.95)
Both Pos. Search	-138.87**				-127.82**
	(31.76)				(41.57)
My Pos. Search (num)		0.64			
		(0.49)			
Opp. Pos. Search (num)		0.99			
		(1.16)			
Both Pos. Search (num)		-4.90**			
		(1.73)			
Horiz. (10 deg.)			0.16		28.72
			(18.67)		(27.00)
Vert. (10 deg.)			20.92		38.09
			(22.13)		(24.78)
Distance				0.09*	0.09**
				(0.03)	(0.03)
Total clicks				0.03	-0.03
				(0.31)	(0.27)
Constant	71.68**	80.34**	81.82**	75.05**	65.82**
	(17.71)	(17.63)	(18.54)	(18.07)	(17.54)
N	1,408	1,408	1,303	1,303	1,303
R^2	0.04	0.02	0.01	0.02	0.06

+ $p < .10$, * $p < .05$, ** $p < .01$

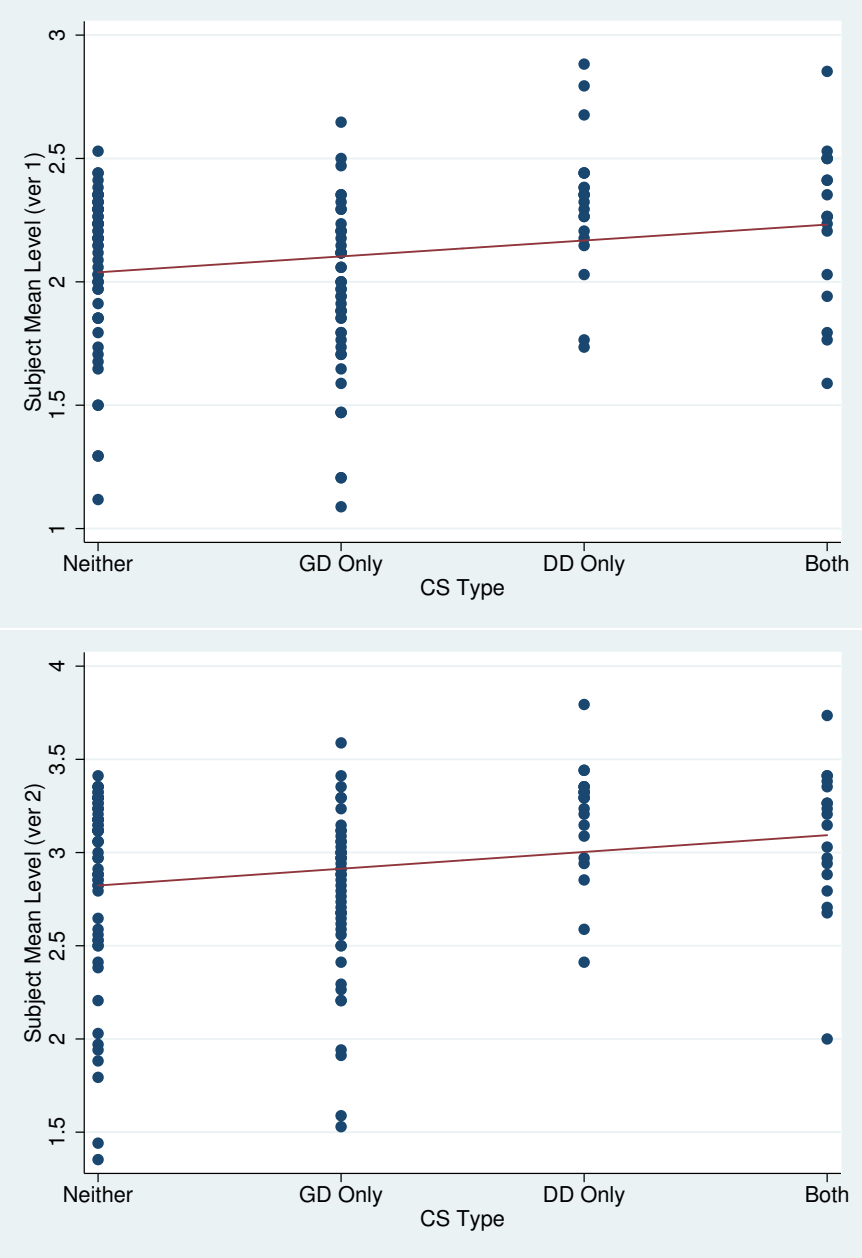



Figure A4: Comparative Statics vs. Level K