Online Appendix to

Are Two heads Better Than One: Team versus Individual Play in Signaling Games

David C. Cooper and John H. Kagel

This appendix contains a discussion of the robustness of the regression results reported in paper to alternative specifications as well as additional details about the codings of the dialogues.

A) Robustness of the Regression Results: The regression results reported in the published appendix of the paper incorporate three critical decisions about how to analyze the data: how to control for individual and team effects, how to define strategic play, and how to control for Es' behavior. For all three decisions, reasonable alternatives exist. In this appendix we show that our primary qualitative conclusions are not sensitive to our decisions on these three issues. We also discuss the reasoning behind our decisions.

We examine the robustness of the regression results in the context of the games with low cost Es. The low cost E games are the case of greatest interest for robustness checks because they are most likely to be sensitive to alternative specifications. In the high cost E games the 2x2 effects are quite weak, so that if an alternative specification yielded the conclusion that there were strong 2x2 effects, we'd be inclined to dismiss the alternative specification as flawed. Likewise, the 2x2 effects are so strong in the cross-over sessions that we would be dubious about any specification that failed to find strong 2x2 effects. We have also checked the robustness of our results for these treatments and find (unsurprisingly) that the results are robust, but do not report them here in the interest of brevity.

A.1) Controls for Individual and Team Effects: For all of the regressions reported in the published appendix we controlled for the individual and team effects by using the clustering approach of Moulton (1986) and Liang and Zeger (1986). This approach has the advantages of being conservative and easy to implement, but ignores a great deal of information contained in the data. It is therefore worthwhile to explore how the results are changed by using alternative controls for the individual and team effects that make greater use of the available information.

Specifically, we fit a variant on a random effects specification. For the 1x1 data, the random effects specification is completely standard. Let v_I be the variance of the random effects term in the 1x1 data. For the 2x2 data, the error term for each observation is the sum of a team error term, which plays the role of a standard random effects error term, as well as two individual error terms. Let y_{ijkt} be the latent dependent variable for individuals i and j in team k in round t, let X_{ijkt} be the vector of independent variables for individuals i and j in team k in round t, let μ_i and μ_j be individual specific error terms, let υ_k be a team specific error term, and let ϵ_{ijkt} be an i.i.d. error term for individuals i and j in team k in round t. Assume that $\mu_i \sim N(0,v_{T2})$, $\mu_j \sim N(0,v_{T2})$, $\upsilon_k \sim N(0,v_{T1})$, and that $\epsilon_{ijkt} \sim N(0,1)$. Then y_{ijkt} is given by the following equation where α is a scalar and β is a vector of parameters:

$$y_{ijkt} = \alpha + \beta X_{ijkt} + \mu_i + \mu_j + \upsilon_k + \epsilon_{ijkt}$$

The dependent variable is derived in the standard way for a probit given the latent variable. Note that setting $v_{T1} = v_I$ and $v_{T2} = 0$ gives us a standard random effects specification with the random effect at the team level for the 2x2 data. The error terms for individuals within a team, μ_i and μ_j allow us to capture any correlation between observations from two different teams that share a common individual. If $v_{T2} > 0$, there is positive correlation between such observations.

Fitting the full model with both team and individual effects for the 2x2 data is quite computationally intensive. A standard random effects model only requires integration over a single dimension making use of Gauss-Legendre quadrature simple. With the combination of individual and team effects, however, the dimensionality of the numerical integration problem equals the number of teams in the largest interlinked "chunk" as defined above. For our data set, that requires us to differentiate over 58 dimensions! If we used the Gauss-Legendre method with a quadrature of 18, this requires calculating the probit function $6.4 * 10^{72}$ times for a single calculation of the likelihood function. Even with a fast computer, making this many calculations isn't feasible. We therefore use simulated maximum likelihood methods to estimate the full model. Specifically, rather than numerically calculating the necessary integral for each chunk, we draw a sample of 25,000 random vectors of individual error terms from the appropriate distribution. We then maximize the likelihood for this sample, treating each vector of individual error terms as equally likely. Standard numerical integration techniques can still be used for the team effects and the individual effects in the 1x1 data.

Table A4 reprises Model 2 from Table A2 in the published appendix. The first column gives the same regression as is reported on Table A2. In other words, this is a standard probit with the error terms corrected for clustering at the "chunk" level. The second column fits the random effects model with v_{T2} = 0. In other words, this is a random effects model with individual effects for the 1x1 data and team effects (but no individual effects) for the 2x2 data. We do not restrict $v_I = v_{T1}$, although the null hypothesis that these parameters are equal cannot be rejected statistically. The results are even stronger than those reported in the initial regression, with the 2x2 interaction terms almost always statistically significant at the 1% level. The magnitude of the estimates indicates that the 2x2 effect is largest in the second cycle of the experienced data, consistent with the data but not with the regression using clustering. The results of this regression are probably too strong, as observations from different teams sharing a common individual are treated as independent. The third column therefore reports the results of fitting the full model via maximum simulated likelihood estimation. While the added parameter is statistically significant, the economically relevant results change little. To summarize, the estimated 2x2 effects in the low cost entrants games are strengthened with the use of less conservative controls for individual and team effects.

A.2) Alternative Definitions of Strategic Play: The analysis reported in the published appendix defines strategic play for MLs as the choice of output levels 5, 6, or 7. The inclusion of output level 7 is irrelevant, as no subject ever chose 7 in the low cost entrant games. On the other hand, output level 5 is chosen quite frequently and isn't as unambiguously strategic as output level 6. Playing devil's advocate, output level 5 is not part of a pure strategy separating equilibrium, leaves open the possibility of pooling by MHs, and is sufficiently close to the myopic maximum as to make random errors a plausible explanation for its choice. We have argued elsewhere that we believe choice of 5 *is* strategic play, but nonetheless it seems prudent to check that our conclusions do not depend on the inclusion of output level

5 as strategic play.[1]

Table A5 revisits Model 1 from Table A2 in the published appendix, the basic regression testing for the existence of 2x2 effects in the low cost entrant games. The first column reports the same regression as appears in Table A2. The second column runs the same specification with strategic play defined solely as play of output level 6. For this regression, choices of output level 5 are coded as zeros for non-strategic play. This regression only counts play as strategic in the most unambiguous case. The final column takes a balanced approach, reporting results from an ordered probit where each output level is coded as a separate category.[2] Given that most choices by MLs are of output levels 4, 5, and 6, this has the effect of giving some credit for a choice of output level 5 as strategic play, but less credit than a choice of output level 6 receives.

The impact of changing the definition of strategic play is small. Defining only play of 6 as strategic slightly weakens the 2x2 effects while the balanced approach of using an ordered probit has very little effect.[3] Regardless of which regression is used, the qualitative conclusions remain the same. There are significant 2x2 effects which grow stronger with experience.

A.3) Alternative Controls for Es' Behavior: Model 2 on Table A2 in the published appendix added a control for the entry rate differential between output levels 4 and 6. This new variable is statistically significant and impacts the estimates of the 2x2 effects. It therefore seems wise to check whether plausible alternative controls for Es' behavior affect the estimated 2x2 effects sufficiently to alter our conclusions.

Table A7 reexamines the results reported in Model 2 on Table A2 in the published appendix. The first column is identical to the regression reported on Table A2. The second column includes not only the

[1] For the games with high cost Es, we have run analogous robustness checks for the classification of output level 3 as strategic play by MHs.

[2] We have also run ordered probits in which output levels 1 - 4 are considered a single category. This has little impact on the results.

[3] The estimated breakpoints between categories are not reported. While these are statistically significance, they are of no direct economic relevance.

entry rate differential for the current cycle, but also the lagged entry rate from the previous cycle. This yields a large number of missing observations, as the lagged entry rate differential (by definition) isn't defined for the first cycle of play. We therefore demean the lagged entry rates for the cycles where it is defined and set the lagged entry equal to zero for the cycles where it is not defined.[4] The use of lagged entry rates captures any persistent effects of Es' behavior on Ms' choices. The third column calculates the entry rate differential based on *all* preceding periods.[5] Unlike the entry rate differential measure used in Table A2, this measure does not include any data from periods that have not yet occurred. It also has the advantage of responding immediately to changes in Es' behavior and allowing for persistent effects of Es' behavior. Once again there is a problem with missing observations. We therefore demean the entry rate differential for observations in which it is defined and set the variable equal to zero for observations with missing values. The final column includes separate parameter estimates for the entry rates on output levels 4, 5, and 6. We return to using entry rates defined on a cycle by cycle basis. This specification nests the entry rate differential specification used in Table A2. By allowing subjects to put different weights on these three entry rates, we allow for the possibility that entry rates which are based on more observations (e.g. output level 4) are given greater weight. We also allow the entry rate on output level 5 to play a role in MLs' choices, a reasonable precaution given the frequent choices of output level 5.

All of the alternative controls for Es' behavior lead to better fits of the data, as can be seen by the strong improvements in the log-likelihood. The impact on the estimated 2x2 effects of using different measures for the entry rate differential is mixed. The significant 2x2 effect we report for the second cycle of inexperienced play disappears with any of the alternative controls for entry rate differentials. On the other hand, the 2x2 effect for the second cycle of experienced play becomes statistically significant for two of the three new regressions. The 2x2 effect for the final cycle of experienced play is reduced in

---

[4] The lagged entry rate for the first cycle of experienced play is set equal to zero, not the entry rate for the final cycle of inexperienced play.

[5] For experienced subjects, this measure does not include data from their inexperienced subject session.

statistical significance, but this is due more to an increase in the standard errors than to a decrease in the magnitude of the parameter estimates.

Any controls for Es' behavior must be interpreted with caution. Missing observations are a frequent problem, entry rates are often based on very few observations, and the maintained assumption that individuals and teams are equally sensitive to entry rates is probably false. Nonetheless, the bottom line remains the same as before. Controlling for the behavior of Es' weakens the impact of the 2x2 treatment. The alternative measures of Es' behavior make it particularly clear that the 2x2 effect is strongest for experienced subjects.

B.4) Why Not Use the Alternative Specifications: While our qualitative results are not sensitive to the use of alternative specifications, the regressions reported in Tables A5 and A6 suggest that we could have fit the data better. Indeed, the use of alternative specifications would have strengthened our conclusions. Drawing on the results shown in Tables A5 and A6, consider a "best-case" regression. Specifically, we run an ordered probit with each output level as a category (as in the final column of Table A5). To control for Es' behavior we use the entry rates for output levels 4, 5, and 6 over *all preceding periods* (combining the controls used in the final two columns of Table A6). The resulting estimated 2x2 effects are shown in Table A7. These parameter estimates are very sensible. Controlling for Es' behavior, we see no 2x2 effect for inexperienced subjects and a consistent strong 2x2 effect among experienced subjects. This is probably the clearest set of results we could present. If our sole concern was generating the best possible picture of MLs' choices in the low cost entrant game, the specification shown in Table A8 would obviously be preferable to the specification reported in Table A2 of the published appendix.

However, in the broader scheme of things there are several good reasons we did not redo all of the regressions using this "best case" specification. First, the cumulative measures of entry rates are problematic in the crossover sessions. At the point where we are most interested in ascertaining whether teams differ from individuals, the crossover between the high cost entrant game and the low cost entrant game, all of the experience from preceding periods is coming from a different game than the one being currently played. Restarting the entry rates when the crossover occurs is problematic, especially since there is virtually no play of output level 6 in the individual sessions, and taking a weighted average of experience before and after the crossover is clumsy. Second, we feel it is important to confirm that the results are not sensitive to how we control for individual and team effects, as shown in Section A.1 of this appendix. The full random effects specification with individual and team effects is quite computationally intensive, taking days to run on a fast machine. The computational complexity increases exponentially in the number of parameters that must be fit. This argues in favor of using a simple specification with one entry rate control rather than three and a probit rather than an ordered probit. Finally, there is a lot to be said for simplicity. The specification we report in Table A2 is easy to interpret and captures the main features of the data.

B) Details of the Codings: In Section V we discuss the correlation between the two codings and, in Table 4, give a variety of data on the most common codings. Table A8 provides equivalent coding information for all categories. As with the text, data are for whether teams had *ever* been coded for a category prior to first strategic play. Cross-coder correlations are reported as well – these treat each period for each team as an observation. Calculating cross-coder correlations for whether teams had ever been coded for a category, either over all periods or prior to first strategic play, yields somewhat higher cross-coder correlations.

Table A4
Alternative Controls for Individual and Team Effects, Low Cost Entrant Sessions
Probit Regressions on Strategic Choice by MLs

(1375 obs, 176 teams)

Definition of Strategic Play	Clustering	Team Effects Only	Team and Individual Effects
Constant	-.451*** (.169)	-.944*** (.107)	-.956*** (.105)
Inexperienced Cycle 2	.270 (.177)	.460*** (.116)	.471*** (.116)
Inexperienced Cycle 3	.149 (.311)	.848*** (.155)	.865*** (.155)
Experienced Cycle 1	.190 (.238)	.691*** (.142)	.702*** (.142)
Experienced Cycle 2	.605** (.236)	.876*** (.137)	.881*** (.136)
Experienced Cycle 3	.638** (.248)	1.210*** (.146)	1.221*** (.145)
Experienced Cycle 4	.628** (.275)	1.331*** (.154)	1.345*** (.153)
2x2 * Inexperienced Cycle 1	-.046 (.195)	.324* (.174)	.344* (.180)
2x2 * Inexperienced Cycle 2	.463** (.205)	.733*** (.189)	.714*** (.179)
2x2 * Experienced Cycle 1	.709*** (.225)	.781*** (.232)	.797*** (.224)
2x2 * Experienced Cycle 2	.480 (.309)	1.242*** (.311)	1.242*** (.303)
2x2 * Experienced Cycle 3	Dropped (No variation)	Dropped (No variation)	Dropped (No variation)
2x2 * Experienced Cycle 4	.934*** (.399)	1.040*** (.330)	1.050*** (.327)
Entry Rate Differential	1.791*** (.443)	.392** (.170)	.359** (.168)
Individual Effect 1x1 Treatment		1.999*** (.185)	2.009*** (.184)
Individual Effect 2x2 Treatment			.780*** (.220)
Team Effect 2x2 Treatment		1.608*** (.260)	1.216*** (.284)
Log Likelihood	-703.53	-491.42	-489.06

* statistically significant at the 10% level
** statistically significant at the 5% level
*** statistically significant at the 1% level

Table A5
Alternative Measures of Strategic Play, Low Cost Entrant Sessions
Standard Errors Corrected for Clustering at the "Chunk" Level

(1374 obs, 176 teams)

Definition of Strategic Play	Standard Definition	Play of 6 Only	Ordered Choice
Constant	-1.032*** (.150)	-1.383*** (.200)	
Inexperienced Cycle 2	.795*** (.157)	.589*** (.182)	.464*** (.127)
Inexperienced Cycle 3	1.006*** (.244)	.641** (.258)	.624*** (.192)
Experienced Cycle 1	.809*** (.221)	.875*** (.243)	.630*** (.178)
Experienced Cycle 2	.915*** (.235)	1.178*** (.258)	.814*** (.216)
Experienced Cycle 3	1.230*** (.227)	1.338*** (.247)	1.069*** (.199)
Experienced Cycle 4	1.463*** (.236)	1.473*** (.264)	1.228*** (.202)
2x2 * Inexperienced Cycle 1	.424** (.189)	.488 (.301)	.291** (.142)
2x2 * Inexperienced Cycle 2	.553*** (.197)	.684*** (.258)	.590*** (.180)
2x2 * Experienced Cycle 1	.909*** (.260)	.635 (.392)	.702** (.291)
2x2 * Experienced Cycle 2	1.312*** (.336)	.986*** (.305)	1.135*** (.294)
2x2 * Experienced Cycle 3	Dropped (No variation)	1.094*** (.311)	1.300*** (.279)
2x2 * Experienced Cycle 4	1.320*** (.307)	1.155*** (.353)	1.243*** (.319)
Log Likelihood	-751.45	-728.64	-1506.43

* statistically significant at the 10% level
** statistically significant at the 5% level
*** statistically significant at the 1% level

Table A6
Alternative Measures of Entry Rates, Low Cost Entrant Sessions
Standard Errors Corrected for Clustering at the "Chunk" Level

(1375 obs, 176 teams)

Type of Entry Rate Controls	Standard	Standard and Lagged	Cumulative	Separated by Output Level
Constant	-.451*** (.169)	-.593*** (.166)	-.864*** (.152)	-.227 (.208)
Inexperienced Cycle 2	.270 (.177)	.834*** (.204)	1.078*** (.180)	.209 (.198)
Inexperienced Cycle 3	.149 (.311)	.489 (.314)	.963*** (.243)	.185 (.302)
Experienced Cycle 1	.190 (.238)	.344 (.224)	.414* (.233)	-.042 (.263)
Experienced Cycle 2	.605** (.236)	.570** (.234)	.683*** (.233)	.535** (.242)
Experienced Cycle 3	.638** (.248)	.906*** (.247)	1.009*** (.229)	.449* (.263)
Experienced Cycle 4	.628** (.275)	.778*** (.273)	1.177*** (.256)	.467 (.300)
2x2 * Inexperienced Cycle 1	-.046 (.195)	.071 (.186)	2.83 (.210)	-.280 (.212)
2x2 * Inexperienced Cycle 2	.463** (.205)	.097 (.230)	.119 (.213)	.092 (.281)
2x2 * Experienced Cycle 1	.709*** (.225)	.769*** (.232)	.911*** (.235)	.916*** (.255)
2x2 * Experienced Cycle 2	.480 (.309)	.696** (.281)	1.005*** (.283)	.135 (.324)
2x2 * Experienced Cycle 3	Dropped (No variation)	Dropped (No variation)	Dropped (No variation)	Dropped (No variation)
2x2 * Experienced Cycle 4	.934*** (.399)	.502 (.453)	.788** (.357)	.852* (.511)
Entry Rate Differential	1.791*** (.443)	1.231*** (.362)	1.656*** (.369)	
Entry Rate Differential Lagged		1.211*** (.387)		
Entry Rate 4				2.599*** (.550)
Entry Rate 5				.772*** (.250)
Entry Rate 6				-.484 (.411)
Log Likelihood	-703.53	-681.81	-691.96	-675.45

* statistically significant at the 10% level
** statistically significant at the 5% level
*** statistically significant at the 1% level

Table A7
“Best Case” Specification, Low Cost Entrant Sessions
Standard Errors Corrected for Clustering at the “Chunk” Level

Variable Name	Parameter Estimate
2x2 * Inexperienced Cycle 1	-.075 (.169)
2x2 * Inexperienced Cycle 2	.069 (.231)
2x2 * Experienced Cycle 1	.592** (.281)
2x2 * Experienced Cycle 2	.753*** (.257)
2x2 * Experienced Cycle 3	.667** (.277)
2x2 * Experienced Cycle 4	.652* (.345)
Log Likelihood	-1437.53

Table A8

Proportion of Teams Coded for Category prior to First Strategic Play & Cross-coder correlations

Category	Inexperienced High Cost E Game	Inexperienced Low Cost E Game	Crossover Treatment (Low Cost E Game)[1]	Correlation Between Coders
1	0.418	0.298	0.056	0.548
3i	0.025	0.012	0.042	0.377
3ii	0.123	0.060	0.083	0.389
3iii	0.369	0.286	0.444	0.549
3iv	0.049	0.024	0.056	0.173
4i	0.000	0.024	0.000	0.199
4ii	0.000	0.143	0.181	0.607
4iii	0.016	0.310	0.486	0.516
4iv	0.000	0.071	0.083	0.225
5	0.016	0.107	0.153	0.132
5i	0.008	0.119	0.069	0.247
6	0.000	0.000	0.806	0.444
7	0.000	0.024	0.361	0.600
8	0.082	0.036	0.042	0.225
9	0.057	0.036	0.069	0.384
10	0.557	0.595	0.861	0.636
11	0.000	0.012	0.000	0.816
12i	0.025	0.000	0.056	0.435
12ii	0.049	0.000	0.125	0.665
13	0.107	0.083	0.153	0.183
14	0.008	0.012	0.056	-0.002
15	0.295	0.107	0.139	0.410
16	0.033	0.095	0.153	0.175
Number of Teams	61	42	36	

Note: Data are for first strategic play as an MH in the game with high cost Es and as an ML for the game with low cost Es.

[1] This column does not include the two teams that first played strategically as an ML prior to the crossover.