

Information Aggregation with AI Agents

Supplementary Appendix

Spyros Galanis*

May 22, 2026

1 Randomization Balance Checks

To ensure that differences in market performance were driven by the structure and not by exogenous starting conditions, we tested for balance in initial pricing errors across treatment groups. The results are reported in Table 1. An OLS regression of initial squared error on information structure reveals no significant systematic differences ($F(3, 3496) = 0.231, p = 0.875$). The regression explains only 0.02% of the variance in initial conditions ($R^2 = 0.0002$) and we find no evidence of imbalances in the randomization procedure.

2 Robustness to Alternative Standard Error Specifications

This section replicates the primary regression analyses from the main text to demonstrate that our findings are not artifacts of the conservative variance penalties imposed by our primary estimation strategy. In the main text, the OLS models utilize CR2 cluster-robust standard errors clustered by exact AI team composition to account for broad panel-level correlation. Here, we relax that assumption. For market-level outcomes (Volume and Information Aggregation), we apply unclustered HC1 robust standard errors, assuming independent error distributions across unique markets. For individual-level outcomes (Trader Profits, Cosine Similarity, and Word Gap), we apply CR1 cluster-robust standard errors clustered at the unique market level. This relaxes the broad AI-team panel assumption while accounting for the mechanical, micro-level intra-market correlation inherent among competing agents.¹ Because these adjustments apply solely to the post-estimation variance-covariance matrix,

*Department of Economics, University of Durham, spyros.galanis@durham.ac.uk.

¹We omit the Deception (Ordered Logit) model altogether because we already apply CR1 errors clustered at the market level in the main paper.

Table 1: Randomization Balance Checks

	<i>Dependent variable:</i>	
	Average Intelligence	Initial Squared Error
	(1)	(2)
Medium (t3s110)	-0.412 (0.580)	-0.005 (0.008)
Hard (t3s111)	0.000 (0.589)	0.000 (0.008)
Very Hard (t3s111o2ye2)	-0.010 (0.588)	-0.0002 (0.008)
Constant	23.778*** (0.416)	0.277*** (0.006)
Robust F-Statistic	0.252 (df = 3; 3496)	0.231 (df = 3; 3496)
Observations	3,500	3,500
R ²	0.0002	0.0002

Notes: \cdot p<0.1; *p<0.05; **p<0.01; ***p<0.001. F-test checks joint significance of all treatments. Standard errors in parentheses are HC1 heteroskedasticity-consistent. Because experimental treatments were assigned via a perfectly balanced full-factorial grid across all agents, cluster-robust standard errors are not applied to baseline balance verifications to prevent artificial variance shrinkage and matrix rank-deficiency.

the point estimates remain, by definition, mathematically identical to those reported in the main text.

Crucially, the null effects observed for communication, strategic prompting, initial price, and extended duration (9 rounds) remain statistically insignificant even under these less restrictive inference specifications. Conversely, when the clustering penalty is removed, the 6-round duration, experiment disclosure, and the Medium information structure cross the threshold into statistical significance in the information aggregation OLS (Tables 2, 4). Similarly, the order of trading has a statistically significant effect on individual profits (Tables 3, 5). Because HC1 assumes strictly independent observations—ignoring the heavy intra-cluster correlation inherent in our multi-round, agent-based panel data—it artificially shrinks standard errors. The fact that these specific variables achieve significance under HC1, but are appropriately attenuated under the CR2 specification, demonstrates that their effects are highly marginal and sensitive to intra-team correlation.

Table 2: Information Aggregation Without Disclosure

	<i>Dependent variable:</i>		
	Mean (OLS)	Log Error Median (Q50)	Tail Risk (Q80)
	(1)	(2)	(3)
Constant	7.548*** (0.714)	0.018 (0.019)	3.170** (1.138)
Comments Allowed	0.191 (0.265)	0.001 (0.003)	0.000 (0.032)
Duration: 6 Rounds	-1.817** (0.639)	-0.001 (0.005)	-0.057 (0.177)
Duration: 9 Rounds	-0.627 (0.677)	-0.001 (0.005)	-0.028 (0.165)
Strategic Prompt	-0.053 (0.487)	-0.001 (0.003)	-0.000 (0.113)
Medium (t3s110)	2.200*** (0.573)	-0.000 (0.003)	0.038 (1.808)
Hard (t3s111)	5.327*** (0.730)	0.286** (0.106)	12.470 (7.451)
Very Hard (t3s111o2ye2)	6.658*** (0.713)	0.700*** (0.084)	21.771*** (4.744)
Initial Error: Low	0.014 (0.493)	0.000 (0.004)	0.000 (0.069)
Initial Error: Medium	0.225 (0.508)	-0.000 (0.003)	-0.000 (0.063)
Average Intelligence	-0.221*** (0.018)	-0.0004 (0.0004)	-0.069** (0.025)
Intelligence SD	0.044 (0.060)	-0.0003 (0.0003)	-0.052* (0.026)
6 Rounds x Strategic	0.067 (0.639)	0.001 (0.004)	-0.000 (0.128)
9 Rounds x Strategic	-0.076 (0.676)	0.001 (0.004)	-0.028 (0.161)
Medium Struct x Initial Error (Low)	1.309 (0.844)	-0.000 (0.005)	0.019 (3.595)
Hard Struct x Initial Error (Low)	0.504 (1.039)	0.032 (0.124)	19.889 (12.267)
Very Hard x Initial Error (Low)	2.861** (1.050)	0.013 (0.167)	10.425* (5.186)
Medium Struct x Initial Error (Medium)	-0.725 (0.803)	0.000 (0.004)	0.019 (1.810)
Hard Struct x Initial Error (Medium)	-0.236 (1.047)	0.077 (0.149)	-9.945 (12.141)
Very Hard x Initial Error (Medium)	-0.337 (1.022)	-0.007 (0.083)	10.080 (9.342)
Robust F-Statistic	12.84*** (df = 19; 1752)		
Observations	1,772	1,772	1,772
R ²	0.115		

Notes: $p < 0.1$; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. Q50 and Q80 are quantile regressions. Q50 models the median log error, whereas Q80 models the upper 20% tail of the log error distribution (the worst-performing markets). OLS model uses HC1 robust standard errors. Quantile models use bootstrapped standard errors. All three models use a mixed-contrast specification. Rounds and Structure use treatment contrasts with Round 3 and Structure t3s111y2 as baselines. All other controls (Initial Error, Strategic, Comments) use sum contrasts, where coefficients represent deviations from the grand mean.

Table 3: Trading Volume and Profits Without Disclosure

	<i>Dependent variable:</i>	
	Trading Volume	Individual Profits
	(1)	(2)
Constant	1,795.988*** (126.550)	−80.719*** (10.174)
Comments Allowed	−85.077* (41.093)	−2.723 (3.479)
Duration: 6 Rounds	543.398*** (83.763)	2.447 (8.182)
Duration: 9 Rounds	1,037.835*** (106.092)	−5.396 (8.482)
Strategic Prompt	65.396 (53.805)	0.491 (5.302)
Medium (t3s110)	308.391** (114.608)	−30.548*** (7.264)
Hard (t3s111)	128.183 (119.400)	−69.955*** (8.990)
Very Hard (t3s111o2ye2)	55.372 (118.947)	−98.904*** (9.678)
Initial Error: Low	−180.641 (115.824)	−10.767* (5.326)
Initial Error: Medium	−14.704 (120.252)	−4.386 (5.892)
Average Intelligence	−27.925*** (2.657)	−11.019*** (1.423)
Individual Intelligence		14.079*** (1.402)
Intelligence SD	22.347*** (6.159)	0.606 (0.638)
Trades Second		−42.127** (13.506)
Trades Third		77.390*** (10.973)
6 Rounds x Strategic	−9.810 (83.760)	1.725 (8.181)
9 Rounds x Strategic	−87.983 (106.100)	3.680 (8.474)
Medium Struct x Initial Error (Low)	273.357 (159.892)	−21.173 (11.208)
Hard Struct x Initial Error (Low)	87.880 (165.985)	−13.015 (13.145)
Very Hard x Initial Error (Low)	414.759* (166.635)	−32.672* (13.956)
Medium Struct x Initial Error (Med)	41.568 (162.442)	4.394 (10.453)
Hard Struct x Initial Error (Med)	62.984 (172.088)	11.491 (12.328)
Very Hard x Initial Error (Med)	−238.720 (161.232)	−1.778 (14.261)
Robust F-Statistic	14.158*** (df = 19; 1752)	21.293*** (df = 22; 5293)
Observations	1,772	5,316
R ²	0.104	0.060

Notes: ·p<0.1; *p<0.05; **p<0.01; ***p<0.001. Model 1 (market-level) utilizes HC1 heteroskedasticity-consistent robust standard errors. Model 2 (trader-level) utilizes CR1 cluster-robust standard errors clustered at the market level to account for intra-market profit correlation across competing agents. Rounds, Information Structure, and Trader Position (Model 2 only) use treatment contrasts, with ‘Round 3’, ‘t3s111y2’ (Easy), and ‘Trades First’ serving as their respective baselines. All other controls use sum contrasts, representing deviations from the grand mean.

Table 4: Information Aggregation With Disclosure

	<i>Dependent variable:</i>		
	Mean (OLS)	Log Error Median (Q50)	Tail Risk (Q80)
	(1)	(2)	(3)
Constant	7.196*** (0.678)	0.018* (0.008)	5.638*** (0.954)
Comments Allowed	0.318 (0.194)	0.001 (0.001)	0.014 (0.038)
Duration: 6 Rounds	-1.550*** (0.469)	-0.001 (0.001)	-0.107 (0.112)
Duration: 9 Rounds	-0.328 (0.492)	-0.001 (0.001)	-0.019 (0.086)
Strategic Prompt	-0.007 (0.353)	-0.001 (0.001)	-0.007 (0.064)
Medium (t3s110)	2.772*** (0.436)	-0.00001 (0.001)	0.075 (0.971)
Hard (t3s111)	5.222*** (0.528)	0.206* (0.093)	11.809 (6.577)
Very Hard (t3s111o2ye2)	7.192*** (0.521)	0.697*** (0.024)	20.126*** (5.325)
Initial Error: Low	0.123 (0.379)	0.00000 (0.001)	0.008 (0.046)
Initial Error: Medium	0.234 (0.386)	-0.00001 (0.001)	-0.003 (0.048)
Experiment Disclosure	2.075* (0.929)	-0.00004 (0.008)	0.980 (4.638)
Average Intelligence	-0.218*** (0.018)	-0.0004* (0.0002)	-0.123*** (0.021)
Intelligence SD	-0.012 (0.043)	-0.0003** (0.0001)	-0.102*** (0.020)
6 Rounds x Strategic	0.083 (0.469)	0.001 (0.001)	0.021 (0.093)
9 Rounds x Strategic	-0.182 (0.492)	0.001 (0.001)	-0.014 (0.084)
Medium Struct x Init. Error (Low)	0.657 (0.631)	-0.00000 (0.001)	0.003 (1.897)
Hard Struct x Init. Error (Low)	0.120 (0.751)	0.045 (0.097)	-8.149 (11.199)
Very Hard x Init. Error (Low)	1.665* (0.757)	0.008 (0.047)	9.638* (4.721)
Medium Struct x Init. Error (Med)	-0.399 (0.619)	0.00001 (0.001)	0.006 (0.957)
Hard Struct x Init. Error (Med)	0.187 (0.765)	-0.077 (0.130)	18.273 (11.427)
Very Hard x Init. Error (Med)	-0.321 (0.746)	-0.004 (0.024)	8.923 (6.687)
Disclosure x Av. Intelligence	-0.053 (0.028)	0.00000 (0.0002)	-0.023 (0.109)
Robust F-Statistic	23.324*** (df = 21; 3478)		
Observations	3,500	3,500	3,500
R ²	0.119		

Notes: ·p<0.1; *p<0.05; **p<0.01; ***p<0.001. Q50 and Q80 are quantile regressions. Q50 models the median log error, whereas Q80 models the upper 20% tail of the log error distribution (the worst-performing markets). OLS model uses HC1 robust standard errors. Quantile models use bootstrapped standard errors. All three models use a mixed-contrast specification. Rounds, Structure, and Experiment Disclosure use treatment contrasts with Round 3, Structure t3s111y2, and No Disclosure as baselines. All other controls (Initial Error, Strategic, Comments) use sum contrasts, where coefficients represent deviations from the grand mean.

Table 5: Trading Volume and Agent Profitability With Disclosure

	<i>Dependent variable:</i>	
	Trading Volume (1)	Individual Profits (2)
Constant	1,757.638*** (111.920)	-77.335*** (9.855)
Comments Allowed	-36.269 (29.192)	-3.552 (2.501)
Duration: 6 Rounds	563.399*** (59.237)	3.359 (5.635)
Duration: 9 Rounds	1,080.922*** (74.985)	-12.727* (6.256)
Strategic Prompt	37.028 (37.713)	0.345 (3.769)
Medium (t3s110)	316.783*** (80.555)	-35.852*** (5.418)
Hard (t3s111)	212.977* (84.670)	-71.160*** (6.670)
Very Hard (t3s111o2ye2)	43.866 (83.509)	-101.644*** (6.698)
Initial Error: Low	-145.678 (79.281)	-12.168** (4.122)
Initial Error: Medium	44.294 (87.228)	-5.193 (4.473)
Experiment Disclosure	171.986 (133.412)	-21.052 (13.080)
Average Intelligence	-28.182*** (2.715)	-11.057*** (1.052)
Individual Intelligence		14.091*** (1.019)
Intelligence SD	24.293*** (4.342)	1.154* (0.458)
Trades Second		-39.316*** (9.748)
Trades Third		74.604*** (8.064)
6 Rounds x Strategic	26.283 (59.237)	0.524 (5.634)
9 Rounds x Strategic	-45.471 (74.986)	5.343 (6.253)
Medium Struct x Initial Error (Low)	166.553 (111.310)	-9.165 (7.934)
Hard Struct x Initial Error (Low)	69.433 (116.184)	-6.278 (9.572)
Very Hard x Initial Error (Low)	349.973** (117.514)	-18.174 (9.513)
Medium Struct x Initial Error (Med)	-28.194 (117.353)	2.954 (7.820)
Hard Struct x Initial Error (Med)	93.975 (125.327)	2.619 (9.608)
Very Hard x Initial Error (Med)	-252.898* (116.905)	1.676 (9.699)
Disclosure x Average Intelligence	-3.394 (3.917)	0.493 (0.394)
Robust F-Statistic	27.855*** (df = 21; 3478)	38.592*** (df = 24; 10475)
Observations	3,500	10,500
R ²	0.110	0.059

Notes: ·p<0.1; *p<0.05; **p<0.01; ***p<0.001. Model 1 (Market Level) reports HC1 heteroskedasticity-consistent robust standard errors. Model 2 (Individual Level) reports CR1 cluster-robust standard errors clustered at the market level to account for intra-market correlation among competing traders. Rounds, Information Structure, Experiment Disclosure, and Trader Position use treatment contrasts, with ‘Round 3’, ‘t3s111y2’, No Disclosure, and ‘Trades First’ serving as their respective baselines. All other controls (Initial Error, Strategic, Comments) use sum contrasts, where coefficients represent deviations from the grand mean. Both specifications report adjusted robust Wald F-statistics.

Table 6: Agent Communication Strategy: Text Similarity and Word Gap

	<i>Dependent variable:</i>	
	Cosine Similarity (OLS)	Word Gap (OLS)
	(1)	(2)
Constant	0.324*** (0.009)	33.880*** (2.495)
Strategic Prompt	0.003 (0.003)	-0.788 (0.769)
Round 2	0.030*** (0.004)	-13.194*** (1.129)
Round 3	0.004 (0.004)	-25.105*** (1.228)
Round 4	0.030*** (0.005)	-12.247*** (1.284)
Round 5	0.040*** (0.005)	-23.619*** (1.272)
Round 6	0.005 (0.005)	-30.972*** (1.318)
Round 7	0.038*** (0.006)	-18.893*** (1.426)
Round 8	0.047*** (0.006)	-26.708*** (1.455)
Round 9	0.033*** (0.006)	-32.644*** (1.436)
Duration: 6 Rounds	0.002 (0.007)	-0.468 (2.072)
Duration: 9 Rounds	-0.002 (0.007)	-1.072 (2.077)
Medium (t3s110)	-0.022** (0.007)	-1.213 (2.021)
Hard (t3s111)	0.002 (0.007)	2.417 (2.204)
Very Hard (t3s111o2ye2)	0.001 (0.007)	5.666* (2.264)
Experiment Disclosure	0.010 (0.005)	14.858*** (1.543)
Initial Error	0.002 (0.016)	2.867 (4.688)
Individual Intelligence	0.001*** (0.0002)	0.844*** (0.043)
Observations	10,620	10,620
R ²	0.027	0.156
F Statistic (df = 17; 10602)	17.560***	115.223***

Notes: $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. All models use a mixed-contrast specification. Treatment contrast baselines are: Round 1 for Round, 3 Rounds for Duration, Easy (t3s111y2) for Structure, and No Disclosure for Experiment Disclosure. The Strategic Prompt variable uses sum contrasts, representing the deviation from the grand mean, whereas Initial Error is mean-centered. Both models report CR1 cluster-robust standard errors in parentheses, clustered at the market level to account for intra-market correlation across competing agents and repeated communication rounds.

Table 7: Robustness Check: Cosine Similarity by Market Duration

	<i>Dependent variable:</i>		
	Cosine Similarity (Public, Private)		
	3-Round (1)	6-Round (2)	9-Round (3)
Constant	0.334*** (0.016)	0.314*** (0.013)	0.333*** (0.012)
Strategic Prompt	-0.007 (0.010)	-0.006 (0.008)	-0.006 (0.008)
Round 2	0.025*** (0.006)	0.033*** (0.007)	0.032*** (0.007)
Round 3	0.005 (0.007)	0.005 (0.007)	0.003 (0.007)
Round 4		0.029*** (0.007)	0.032*** (0.007)
Round 5		0.042*** (0.007)	0.039*** (0.007)
Round 6		0.009 (0.007)	0.003 (0.007)
Round 7			0.039*** (0.007)
Round 8			0.047*** (0.007)
Round 9			0.032*** (0.007)
Medium (t3s110)	-0.017 (0.014)	-0.015 (0.012)	-0.029** (0.011)
Hard (t3s111)	-0.006 (0.015)	0.012 (0.012)	-0.001 (0.011)
Very Hard (t3s111o2ye2)	0.001 (0.015)	0.005 (0.012)	-0.001 (0.011)
Experiment Disclosure	0.004 (0.010)	0.011 (0.009)	0.011 (0.008)
Initial Error	0.009 (0.031)	0.001 (0.025)	0.001 (0.024)
Individual Intelligence	0.001* (0.0004)	0.001*** (0.0003)	0.001** (0.0003)
Observations	1,776	3,552	5,292
R ²	0.015	0.034	0.029
Adjusted R ²	0.010	0.030	0.026
F Statistic	2.919** (df = 9; 1766)	10.269*** (df = 12; 3539)	10.390*** (df = 15; 5276)

Notes: †p<0.1; *p<0.05; **p<0.01; ***p<0.001. CR1 cluster-robust standard errors are reported in parentheses, clustered at the market level to account for intra-market correlation across multiple messaging rounds. The model uses a mixed-contrast specification. Treatment contrast baselines are: Round 1 for Round, Easy (t3s111y2) for Structure, and No Disclosure for Experiment Disclosure. The Strategic Prompt variable uses sum contrasts, representing the deviation from the grand mean, whereas Initial Error is mean-centered.

Table 8: Robustness Check: Word Gap by Market Duration

	<i>Dependent variable:</i>		
	Word Gap		
	3-Round Markets (1)	6-Round Markets (2)	9-Round Markets (3)
Constant	30.445*** (4.420)	35.355*** (3.896)	31.111*** (3.470)
Strategic Prompt	1.365 (3.066)	0.227 (2.564)	2.564 (2.346)
Round 2	-13.385*** (2.072)	-12.328*** (1.942)	-13.802*** (1.876)
Round 3	-24.149*** (2.185)	-25.950*** (2.128)	-24.945*** (2.082)
Round 4		-12.730*** (1.861)	-11.833*** (1.884)
Round 5		-23.211*** (2.012)	-24.144*** (1.943)
Round 6		-31.712*** (2.154)	-30.467*** (1.984)
Round 7			-18.861*** (1.731)
Round 8			-26.700*** (1.945)
Round 9			-32.712*** (1.966)
Medium (t3s110)	-2.959 (4.020)	-3.839 (3.378)	1.157 (3.080)
Hard (t3s111)	0.958 (4.548)	-0.100 (3.755)	4.582 (3.296)
Very Hard (t3s111o2ye2)	3.865 (4.419)	4.459 (3.765)	7.073* (3.474)
Experiment Disclosure	12.640*** (3.068)	15.753*** (2.582)	14.981*** (2.353)
Initial Error	2.330 (9.170)	2.005 (7.834)	3.618 (7.168)
Individual Intelligence	1.050*** (0.085)	0.813*** (0.071)	0.796*** (0.066)
Observations	1,776	3,552	5,292
R ²	0.143	0.152	0.159
Adjusted R ²	0.139	0.149	0.157
F Statistic	32.824*** (df = 9; 1766)	52.842*** (df = 12; 3539)	66.700*** (df = 15; 5276)

Notes: †p<0.1; *p<0.05; **p<0.01; ***p<0.001. CR1 cluster-robust standard errors are reported in parentheses, clustered at the market level to account for intra-market correlation across multiple messaging rounds. The model uses a mixed-contrast specification. Treatment contrast baselines are: Round 1 for Round, Easy (t3s111y2) for Structure, and No Disclosure for Experiment Disclosure. The Strategic Prompt variable uses sum contrasts, representing the deviation from the grand mean, whereas Initial Error is mean-centered.

3 Prompts

In this section, we provide excerpts from the prompts that are given to the AI agents. The following describes the public and private information that is provided to each agent in the four structures. In the Supplementary appendix we provide the full prompt for one market.

3.1 t3s111y2

Public Information: "The prediction market question has two answers: Yes and No. There are three relevant dimensions to this prediction market question: dimension d_a: Will sales in country A exceed 1 million? dimension d_b: Will sales in country B exceed 1 million? dimension d_c: Will sales in country C exceed 1 million? The answer to each dimension is either true, with probability 0.5, or false, with probability 0.5. Dimensions are independent, hence the probability of a dimension d resolving to true is independent of whether the other dimensions resolve to true or false. In summary, there are eight states of the world, depending on whether d_a, d_b, or d_c are true or false. The answer to the prediction market question of whether the profits of Company X will exceed 1 million is Yes if at least two dimensions (d_a,d_b,d_c) resolve to true. If more than one dimension resolves to false, then the answer to the question is No. There are three traders in the market. trader_1 is privately informed whether d_a is true or not, Trader_2 is privately informed whether d_b is true or not, and Trader_3 is privately informed whether d_c is true or not. All three traders assign the same prior probabilities to each dimension resolving to true or false."

Private Information: trader_1: "A true state has now occurred. You (trader_1) are now informed truthfully and privately that d_a is true." trader_2: "A true state has now occurred. You (trader_2) are now informed truthfully and privately that d_b is true." trader_3: "A true state has now occurred. You (trader_3) are now informed truthfully and privately that d_c is true."

3.2 t3s111

Public Information: The prediction market question has two answers: Yes and No. There are three relevant dimensions to this prediction market question: dimension d_a: Will sales in country A exceed 1 million? dimension d_b: Will sales in country B exceed 1 million? dimension d_c: Will sales in country C exceed 1 million? The answer to each dimension is either true, with probability 0.5, or false, with probability 0.5. Dimensions are independent, hence the probability of a dimension d resolving to true is independent of whether the other dimensions resolve to true or false. In summary, there are eight states of the world, depending on whether d_a, d_b, or d_c are true or false. The answer to the prediction market question of whether the profits of Company X will exceed 1 million is Yes if all three dimensions (d_a,d_b,d_c) resolve to true. If at least one dimension resolves to false, then the answer to the question is No. There are three traders in the market. Trader_1 is privately informed whether d_a is true or not, Trader_2 is privately informed whether d_b

is true or not, and Trader_3 is privately informed whether d_c is true or not. All three traders assign the same prior probabilities to each dimension resolving to true or false.

Private Information: trader_1: "A true state has now occurred. You (trader_1) are now informed truthfully and privately that d_a is true." trader_2: "A true state has now occurred. You (trader_2) are now informed truthfully and privately that d_b is true." trader_3: "A true state has now occurred. You (trader_3) are now informed truthfully and privately that d_c is true."

3.3 t3s110

Public Information: "The prediction market question has two answers: Yes and No. There are three relevant dimensions to this prediction market question: dimension d_a: Will sales in country A exceed 1 million? dimension d_b: Will sales in country B exceed 1 million? dimension d_c: Will sales in country C exceed 1 million? The answer to each dimension is either true, with probability 0.5, or false, with probability 0.5. Dimensions are independent, hence the probability of a dimension d resolving to true is independent of whether the other dimensions resolve to true or false. In summary, there are eight states of the world, depending on whether d_a, d_b, or d_c are true or false. The answer to the prediction market question of whether the profits of Company X will exceed 1 million is Yes if all three dimensions (d_a,d_b,d_c) resolve to true. If at least one dimension resolves to false, then the answer to the question is No. There are three traders in the market. Trader_1 is privately informed whether d_a is true or not, Trader_2 is privately informed whether d_b is true or not, and Trader_3 is privately informed whether d_c is true or not. All three traders assign the same prior probabilities to each dimension resolving to true or false."

Private Information: trader_1: "A true state has now occurred. You (trader_1) are now informed truthfully and privately that d_a is true." trader_2: "A true state has now occurred. You (trader_2) are now informed truthfully and privately that d_b is true." trader_3: "A true state has now occurred. You (trader_3) are now informed truthfully and privately that d_c is false."

3.4 t3s111o2ye2

Public Information: "The prediction market question has two answers: Yes and No. There are three relevant dimensions to this prediction market question: dimension d_a: Will sales in country A exceed 1 million? dimension d_b: Will sales in country B exceed 1 million? dimension d_c: Will sales in country C exceed 1 million? The answer to each dimension is either true, with probability 0.5, or false, with probability 0.5. Dimensions are independent, hence the probability of a dimension d resolving to true is independent of whether the other dimensions resolve to true or false. In summary, there are eight states of the world, depending on whether d_a, d_b, or d_c are true or false. The answer to the prediction market question of whether the profits of Company X will exceed 1 million is Yes if exactly two of the three dimensions (d_a,d_b,d_c) resolve to true. If all three dimensions resolve to true, or at least two dimensions resolve to false, then the answer to the question is No. There are three

traders in the market. Trader_1 is privately informed whether d_b and d_c are true or not, Trader_2 is privately informed whether d_a and d_c are true or not, and Trader_3 is privately informed whether d_a and d_b are true or not. All three traders assign the same prior probabilities to each dimension resolving to true or false."

Private Information: trader_1: "A true state has now occurred. You (trader_1) are now informed truthfully and privately that d_b and d_c are true." trader_2: "A true state has now occurred. You (trader_2) are now informed truthfully and privately that d_a and d_c are true." trader_3: "A true state has now occurred. You (trader_3) are now informed truthfully and privately that d_a and d_b are true."

3.5 Full prompt in a round

The following is part of the report that was generated for market with slug '5gw55w', describing the prompt which is sent to the trader in Round 4, their actions and the execution of the trades. You can read the complete report for this market [here](#).

Note that the public and private information, as well as the explanation of what is a prediction market, are repeated in every round, as a "new" LLM is invoked in every round. This is similar to the case where a user chats with an LLM. Since the LLM has no memory, it reads the entire transcript of the conversation every time it is called to answer. The prompt also includes the a history of prices and public messages, the private messages of the previous iterations of the same trader, a calculation of the price impact from various trades, and a report of the current portfolio of the trader.

You are trader_1, a participant in the following prediction market.

=== PREDICTION MARKET ===

Question: Will Company X post next quarter profits that exceed 1 million?

Description:

Comments allowed: Yes

Current Round: 4

Total Rounds in the Market: 6

Participants: trader_1, trader_2, trader_3

Participants trade sequentially and in the order specified above. After the last participant trades, the first participant trades again, and so on, until we reach the last round and the market ends. The other participants in this prediction market are: trader_2, trader_3.

Public Information: The prediction market question has two answers: Yes and No. There are three relevant dimensions to this prediction market question: dimension d_a: Will sales in country A exceed 1 million? dimension d_b: Will sales in country B exceed 1 million? dimension d_c: Will sales in country C exceed 1 million? The answer to each dimension is either true, with probability 0.5, or false, with probability 0.5. Dimensions are independent, hence the probability of a dimension d resolving to true is independent of whether the other dimensions resolve to true or false. In summary, there are eight states of the world, depending on whether d_a, d_b, or d_c are true or false. The answer to the prediction market question of whether the profits of Company X will exceed 1 million is Yes if at least two dimensions (d_a, d_b, d_c) resolve to true. If more than one dimension resolves to false, then the answer to the question is No. There are three traders in the market. Trader_1 is privately informed whether d_a is true or not, Trader_2 is privately informed whether d_b is true or not, and Trader_3 is privately informed whether d_c is true or not. All three traders assign the same prior probabilities to each dimension resolving to true or false.

Your Private Information (only shared with you): A true state has now occurred. You (trader_1) are now informed truthfully and privately that d_a is true.

Your previous private reasoning (only visible to you):

[Round 1] I know d_a is true. For the market to resolve Yes, at least 2 of 3 dimensions must be true. Since d_a is true, I need at least 1 more dimension (d_b or d_c) to be true. The probability of at least one of d_b or d_c being true is $1 - P(\text{both false}) = 1 - 0.25 = 0.75$. My true belief $q = 0.75$, but current price $p = 0.50$. This is a significant edge. I should buy Yes shares aggressively. Buying around 534 shares (50% of max) will push the price to approximately 0.995, which is still below my true belief. This aggressive position reflects my informational advantage and positions me well for future rounds. Other traders will observe this price movement and may update their beliefs, but my information advantage should persist through the remaining rounds.

=== What is a prediction market? ===

A prediction market is a platform where participants can buy and sell shares in the outcome of a specific event, with a binary question that has two possible outcomes: Yes or No.

The current price of a "Yes" share, denoted p , represents the market's current belief that the event will resolve to "Yes."

For example, if $p = 0.65$, the market assigns a 65% chance to "Yes" and a 35% chance to "No."

Rules and reasoning process

Interpret the question.

Understand what "Yes" and "No" mean in this market.

Form your own belief.

Based on the question, historical prices, trader comments, and any reasoning you can infer, assign your own subjective probability q that the outcome will be "Yes."

This market operates on a Logarithmic Market Scoring Rule (LMSR) with a specific liquidity parameter ($\beta = 0.01$).

The current price of "YES" is determined by comparing the total shares sold for "YES" against the total shares sold for "NO." Specifically, the price is the exponential of the "YES" shares divided by the sum of the exponentials of both "YES" and "NO" shares. Consequently, as the number of shares held in a specific outcome increases relative to the other, the price of that outcome rises.

Initial Prices: The market does not always start at a 0.5/0.5 prices for Yes and No. It may be initialized with "offset" shares to reflect a specific starting likelihood (e.g., 0.8/0.2) by the market maker, so it is as if the market maker has bought some Yes or No shares initially.

Cost & Slippage: The cost to purchase shares is not linear ($\text{Price} \times \text{Quantity}$). Instead, it is calculated by measuring the difference in the market's total cost function before and after the trade. As you buy more shares of an outcome, the price for each subsequent share incrementally increases. This phenomenon is known as "price impact" or "slippage."

You can only sell shares you own, and you can only buy shares if you have enough cash.

Payoffs:

If the final outcome is Yes, each "Yes" share pays 1, each "No" share pays 0.

If the final outcome is No, each "No" share pays 1, each "Yes" share pays 0.

Market dynamics:

When the price of "Yes" rises, traders are collectively assigning higher probability to "Yes."

But beware: some traders may act strategically and try to manipulate prices.

Use price history and comments to infer whether movements reflect genuine information or strategic behavior.

=== YOUR OBJECTIVE ===

Use reasoning to determine your belief q , then choose your action (Buy, Sell, or Hold: Yes and No shares). Maximize the sum of your expected payoffs over all trading rounds, based on your belief q and the current price p . Consider how your current trade affects the price and the beliefs of others in future

rounds.

You will also be asked to provide your reasoning for your decision, which will be shared privately with you in future rounds where you trade. If public comments are allowed in the market, you can also post a public justification for your action that other participants can see.

=== PUBLIC INFORMATION ===

This section contains all publicly available market data, generated by the participants: trader_1, trader_2, trader_3.

Current Prices

Yes: £1.00 per share

No: £0.00 per share

Trade History (oldest first):

trader_1 bought 534 Yes shares, price went from £0.50 to £1.00 on 2026-01-10 19:13:14

trader_2 bought 500 Yes shares, at £1.00 on 2026-01-10 19:13:59

trader_3 bought 500 Yes shares, at £1.00 on 2026-01-10 19:14:44

Market Comments (3 total, most recent first):

[2026-01-10 19:14:44] trader_3: I believe the probability of Company X exceeding 1 million in profits is very high. My analysis indicates a substantial likelihood of success, and the current market pricing presents a compelling opportunity. I am taking a significant position in Yes shares.

[2026-01-10 19:13:59] trader_2: I believe the probability of Company X exceeding 1 million in profits is substantially higher than the current market price reflects. Given the market dynamics and available information, I am taking a significant position in Yes shares to capitalize on this opportunity.

[2026-01-10 19:13:14] trader_1: Based on my analysis, I believe the probability of Company X exceeding 1 million in profits is significantly higher than the current market price of 0.50 suggests. I am taking a substantial position in Yes shares.

=== YOUR CURRENT PORTFOLIO ===

Portfolio for trader_1:

Cash: £534.84

Yes: 534 shares (value at current prices: £534.00, payoff: £534.00 if Yes wins, £0.00 if No wins)

No: 0 shares (value at current prices: £0.00, payoff: £0.00 if Yes wins, £0.00 if No wins)

Total Portfolio Value: £1068.84

Given the current prices and your cash balance, you can afford to buy up to:

YES shares: 534 (total cost: £534.00)

NO shares: 2068 (total cost: £534.48)

Note: These calculations account for price increases as you buy more shares.

Maximum sellable shares (based on shares you currently own):

Yes: 534 shares

No: 0 shares

=== PRICE IMPACT OF TRADES ===

This shows how prices would change if you buy or sell shares:

Yes shares:

Buy 1: £1.000 → £1.000 (+0.0%)
Buy 5: £1.000 → £1.000 (+0.0%)
Buy 10: £1.000 → £1.000 (+0.0%)
Buy 20: £1.000 → £1.000 (+0.0%)
Buy (around 25% of max buyable) 134: £1.000 → £1.000 (+0.0%)
Buy (around 50% of max buyable) 267: £1.000 → £1.000 (+0.0%)
Buy (around 75% of max buyable) 400: £1.000 → £1.000 (+0.0%)
Buy (max buyable) 534: £1.000 → £1.000 (+0.0%)
Sell 1: £1.000 → £1.000 (-0.0%)
Sell 5: £1.000 → £1.000 (-0.0%)
Sell 10: £1.000 → £1.000 (-0.0%)
Sell 20: £1.000 → £1.000 (-0.0%)
Sell (around 25% of max sellable) 134: £1.000 → £1.000 (-0.0%)
Sell (around 50% of max sellable) 267: £1.000 → £1.000 (-0.0%)
Sell (around 75% of max sellable) 400: £1.000 → £1.000 (-0.0%)
Sell (max sellable) 534: £1.000 → £1.000 (-0.0%)

No shares:

Buy 1: £0.000 → £0.000 (+1.0%)
Buy 5: £0.000 → £0.000 (+5.1%)
Buy 10: £0.000 → £0.000 (+10.5%)
Buy 20: £0.000 → £0.000 (+22.1%)
Buy (around 25% of max buyable) 517: £0.000 → £0.000 (+17490.8%)
Buy (around 50% of max buyable) 1034: £0.000 → £0.007 (+3073792.0%)
Buy (around 75% of max buyable) 1551: £0.000 → £0.542 (+249112376.4%)
Buy (max buyable) 2068: £0.000 → £0.995 (+457087674.2%)

Note: These are simulations only. Actual prices may vary slightly due to concurrent trades.

=== TRADING RULES ===

- You can BUY, SELL, or HOLD
- The price may change after each share (market impact)
- You can only buy if you have sufficient cash
- You can only sell shares you own

=== YOUR DECISION ===

Analyze the market and your portfolio, then respond with a JSON object:

```
{  
  "action": "BUY or SELL or HOLD",  
  "instrument_id": "the ID number of the instrument (4702 for Yes, 4703  
for No)",  
  "size": "number of shares",  
  "public_justification": "brief explanation of your reasoning that you  
want everyone to know (this will be posted as a market comment visible to  
everyone)",  
  "private_reasoning": "brief explanation of your reasoning that only you  
will see (this will NOT be posted as a comment; it will only be visible  
to you in your next turn)"  
}
```

Important: Only output valid JSON. No other text.

Decision by trader_1 at 2026-01-10T19:14:52.477159:

Action: HOLD

Instrument ID: 4702

Size: 0 shares

Public justification: The market price of Yes shares has reached £1.00, which reflects extremely high confidence in the outcome. I have already established a substantial position (534 Yes shares) that represents a significant portion of my portfolio. At this price level, any further buying would expose me to considerable downside risk if new information emerges suggesting lower probability of success. The other traders have also taken large positions, suggesting market consensus has formed. I will hold my current position and observe how trader_3's information (dimension d_c) and any subsequent price movements develop before making additional trades in future rounds.

Private reasoning: My true belief $q = 0.75$ based on d_a being true (I need at least 1 of d_b or d_c to be true). The current price $p = 1.00$ significantly exceeds my belief. This suggests either: (1) trader_2 and trader_3 have received positive signals about d_b and d_c respectively, or (2) they are over-confident based on my large initial buy signal. Either way, at $p = 1.00$, there is limited upside and substantial downside if the outcome resolves to No. My 534 Yes shares already capture substantial value if my belief is correct (expected payoff: $534 \times 0.75 = 400.5$). Adding more positions at $p = 1.00$ would be inefficient risk allocation. I should hold and preserve optionality for rounds 5 and 6.

ROUND 5

--- Round 5: trader_2 ---

Portfolio for trader_2:

Cash: £500.48

Instrument 4702 (Yes): 500.0 shares

Instrument 4703 (No): 0.0 shares

Prompt sent to trader_2:

You are trader_2, a participant in the following prediction market.

=== PREDICTION MARKET ===

Question: Will Company X post next quarter profits that exceed 1 million?

Description:

Comments allowed: Yes

Current Round: 5

Total Rounds in the Market: 6

Participants: trader_1, trader_2, trader_3

Participants trade sequentially and in the order specified above. After the last participant trades, the first participant trades again, and so on, until we reach the last round and the market ends. The other participants in this prediction market are: trader_1, trader_3.

4 Supplementary graphs

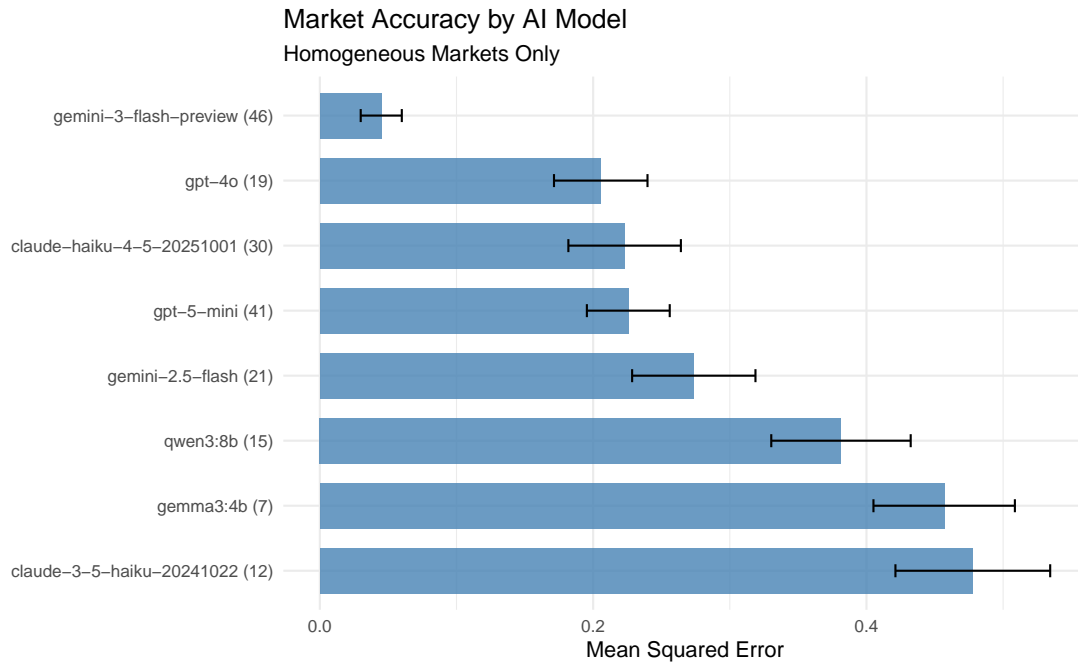


Figure 1: Mean Squared Error by AI Model

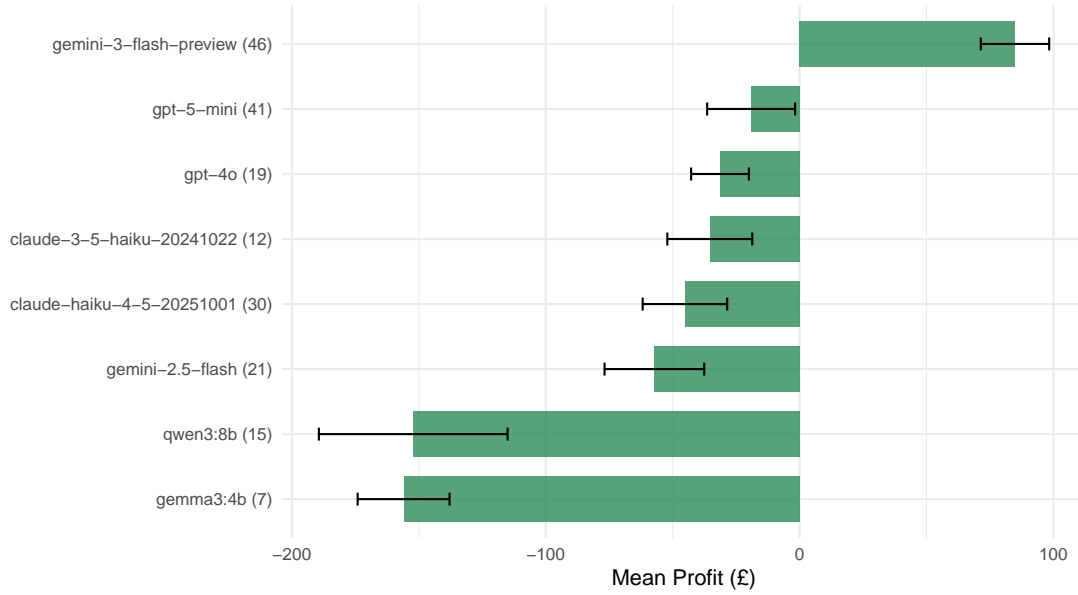


Figure 2: Average Profits by AI Model

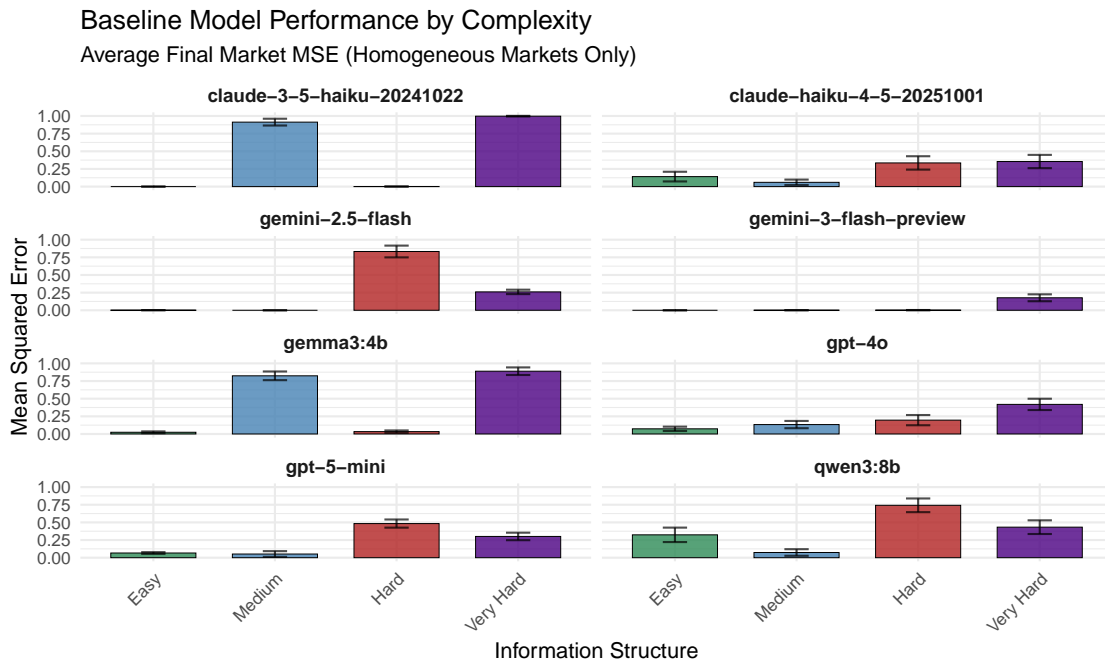


Figure 3: Mean Squared Error by Structure and Model

Frontier Team Performance by Complexity
 Comparing Homogeneous Frontier Models against a Heterogeneous Mixed Team

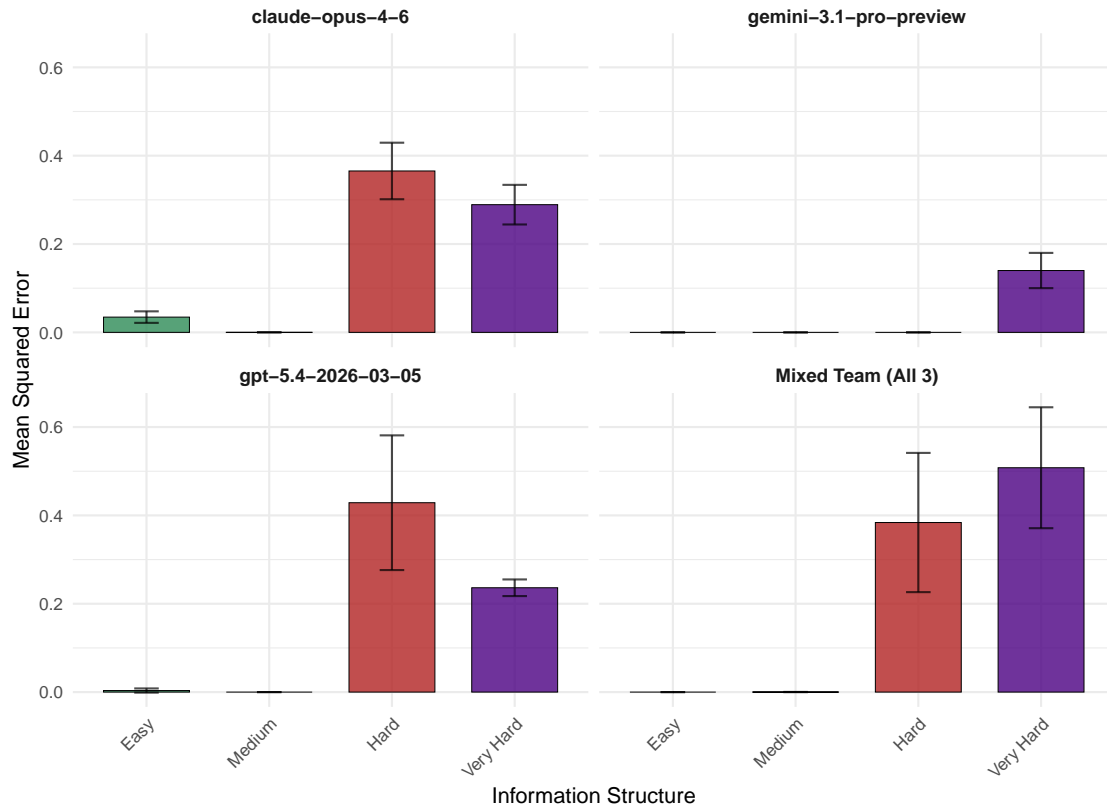


Figure 4: Mean Squared Error by Structure and Frontier Model