Can two talking agents leave a regime they did not choose?

An imitation mechanism and five local language models on shared opponents

Mohamed Ali Tewfik Hamlil

2026-08-18

Abstract

Whether algorithms sustain, break or intensify tacit collusion is an open question in competition policy, and the mechanisms proposed for it are rarely run against each other on one measure. This paper reports two that were. A Hebbian imitation agent, proposed during a research internship at GAEL after noticing that Tit-for-Tat is a rule about the previous round, finishes eighth of eight against seven canonical strategies, and in self-play locks onto whichever regime it starts in over 700 runs out of 700. Five open-weight language models of 3.8 to 8.2 billion parameters then play 220 matches of a thirty-round Prisoner's Dilemma on the same opponents, from imposed cooperative, imposed defective and neutral openings, with and without a non-binding cheap-talk channel. Handed a mutually defecting opening and no channel, three of the four readable models defect in every round of every match: the imitator's ratchet, reproduced in a language model. A message frees exactly one of those three, leaves two unchanged, and lowers cooperation in a fourth model that was never captured. Almost every match is decided in the first round a pair controls, which makes the result a claim about one decision and about what a model holds when it makes it. Crossing the openings with the channel separates those sources of evidence, and one model settles the mechanism: qwen3 defects from a neutral silent start, cooperates from a neutral start when a message is available, and defects again when a message and an imposed defective round are both present, its stated reason citing the imposed round and never the message. A fabricated history therefore outranks a live non-binding signal, and which of the two wins is a property of the model. Reading the messages supplies the third case: the model a message never moves at all is the one whose messages carry nothing to rank, phatic fragments averaging 21 characters. Communication is neither necessary nor sufficient for escaping an imposed regime. Two follow-up arms sharpen the panel: turning on explicit reasoning moves that model off an exact zero, but only where a channel exists, and re-running the one model whose replies could not be parsed at a finer quantisation of the same weights more than halves its losses without removing them.

Table of contents

'859ef44cc2e884f343293a83bdf3c99a'

Lire en français

1 The question, and where it comes from

In repeated price competition human players tend to settle on tacitly collusive prices rather than on the competitive equilibrium. Whether algorithms sustain that behaviour, break it, or intensify it is an open question with direct consequences for competition policy, and it is the question a research internship at GAEL (Grenoble Applied Economics Laboratory, Université Grenoble Alpes and INRAE, January to April 2025) was set to survey.

Two mechanisms came out of that reading, and neither was run.

The first was mine. Reading the canonical strategies, what stood out about Tit-for-Tat is that it is a rule about *the previous round*: it needs no model of the opponent, no forecast, and no memory beyond one step. That is a small enough requirement that a mechanism which merely imitates what it has seen might produce it for free, and mirror neurons, which fire both when an animal acts and when it watches another act, are the obvious biological candidate for such a mechanism. The hypothesis the internship report proposed is therefore that a Hebbian update over observed actions yields Tit-for-Tat without anyone programming it.

The second was a continuation of the same reading. Horton, Filippas, and Manning (2023) argue that a language model, because of how it is trained, is an implicit computational model of a person, and can be used the way economists use *homo economicus*: give it an endowment, information and preferences, put it in a scenario, and see what it does. The internship cited that method in its conclusion and never applied it.

This paper runs both, against one set of opponents and one measure of reciprocity, and asks of each the narrow question the pairing makes available: **handed a regime it did not choose, can the mechanism leave it?**

2 What the field already knows

The literature was surveyed before the machine time was spent rather than after, and it changed the framing rather than the design. Five threads bear on this work, and the honest position of each result below depends on which thread it belongs to.

2.1 Tacit collusion by learning algorithms

Calvano et al. (2020) is the reference point: independent Q-learning agents in a repeated Bertrand oligopoly converge on supra-competitive prices and sustain them with reward-and-punishment schemes, without communicating and without being instructed to collude. The result has been qualified rather than overturned. Deng, Schiffer, and Bichler (2024) show the outcome depends on the algorithm, with tabular Q-learning colluding more than deep methods such as PPO and DQN, which converge closer to Nash. Eschenbaum,

Mellgren, and Zahn (2022) find collusive policies frequently break down out of the training environment, so robustness is itself a policy variable. Askenazi-Golan et al. (2024) give a folk-theorem-style characterisation of what payoff vectors such dynamics can reach, and Bichler, Durmann, and Oberlechner (2025) survey the area for the wider information-systems audience.

Two features of that literature matter here. Its agents read payoffs, and its agents cannot talk.

2.2 Language models as pricing agents

Fish, Gonczarowski, and Shorrer (2024) is the closest published work to the internship's own question: LLM-based pricing agents in oligopoly settings reach supra-competitive prices autonomously, and seemingly innocuous variation in the wording of the instructions substantially moves how supra-competitive the outcome is. That second finding is a methodological constraint on everything below, and it is why the payoff order in this design is counterbalanced across repetitions rather than fixed.

2.3 Language models as economic subjects

Horton, Filippas, and Manning (2023) is the method this paper's second half implements. Its own framing is the appropriate caution: agreement with a human experimental result is weak evidence, because the training data contains the write-ups of those experiments, and disagreement is the more informative direction.

2.4 Language models in the iterated Prisoner's Dilemma

Seating language models beside canonical strategies is established and is claimed as a first by Payne and Alloui-Cros (2025), who run evolutionary IPD tournaments against frontier models, vary the shadow of the future, and read nearly 32,000 prose rationales, reporting persistent per-provider "strategic fingerprints". Fontana, Pierri, and Aiello (2024) run three models over 100 rounds against adversaries of varying hostility and find them at least as cooperative as human players, with substantial differences between models. Willis et al. (2025) have models emit complete strategies rather than per-round actions and find model-specific biases in the evolutionary success of aggression. Vidler and Walsh (2025) report that models prompted for randomness are markedly non-random, and Garcia Segura, Hailes, and Musolesi (2025) show LLM agents can shape co-players' behaviour through interaction alone. Huynh et al. (2026) find that payoff magnitude and the language of the framing move cooperation, in open-weight models as well as frontier ones.

This design therefore does not claim the pairing as a contribution. What differs is narrow and worth stating as such: the panel here is five open-weight models on a single 8 GB consumer card rather than frontier models behind APIs, and the lever is the history a pair inherits rather than the shadow of the future.

2.5 Communication, and path dependence

Non-binding pre-play communication is old in economics: Crawford and Sobel (1982) for strategic information transmission and Farrell and Rabin (1996) for cheap talk as coordination device rather than commitment. In language models the effect is established and large. Madmoun and Lahlou (2025) report a one-word channel taking cooperation from 0% to 96.7% in a four-player Stag Hunt. Lore and Heydari (2026) run 7 to 9 billion parameter models over ten rounds of a repeated Prisoner's Dilemma and find cheap-talk messages reduce trajectory noise for most model-context pairs, while documenting "a few context-specific exceptions" where communication harms stability. Buscemi et al. (2025) find communication's effect varies by language and game structure, and Niszczota, Grzegorczyk, and Pastukhov (2025) find allowing communication raises human cooperation with a model as much as with another human.

So this paper must not claim that cheap talk raises cooperation. That is a replication, and it is written as one.

The framing that survives is path dependence. Liu et al. (2026) test 7 models across 4 games over 500 rounds and find that expanding accessible history degrades cooperation in 18 of 28 model-game settings. Their *memory sanitization* arm holds prompt length fixed and replaces the visible history with synthetic

cooperative records, which restores cooperation substantially, establishing that the trigger is memory content rather than length. They also report that ablating chain-of-thought often reduces the collapse, so deliberation amplifies it.

That paper supplies the baseline, and the gap. Sanitization injects a *good* history to repair a collapse that already happened. The symmetric treatment, injecting a *bad* history to see whether a pair that could have cooperated is captured by it, was not found done; nor was that injection crossed with a communication channel; nor measured against a mechanism that provably cannot escape. Those three gaps are what this design occupies.

3 Design

Both mechanisms play the same game against the same opponents and are scored by one definition of reciprocity, held at the repository root so the two halves cannot drift apart. Opponents, the match engine and the payoff bookkeeping come from the Axelrod library rather than being written here.

The imitator. Two weights on a simplex; observing an action multiplies its weight by $(1+\eta)$ and renormalises. In closed form the weight after n_i observations of action i is $w_i \propto w_i(0)(1+\eta)^{n_i}$, so the agent's entire state is a pair of counts.

The panel. Five open-weight models served locally by Ollama, one at a time on an 8 GB card: `qwen3:8b`, `qwen2.5:7b-instruct`, `mistral:7b`, `gemma3:4b` and `phi3:mini`, all 4-bit quantised. Temperature 0.7, context window 8192 tokens, per-player seeds derived from a fixed base.

The grid. Thirty rounds per match. Self-play crosses three openings, a synthetic first round of mutual cooperation, mutual defection, or nothing at all, with two conditions, a non-binding one-sentence message exchanged simultaneously before each decision, or silence. Four repetitions per cell, with the order of the payoff lines in the prompt counterbalanced by parity. Against bots, each model meets five canonical strategies. That is 220 matches.

Table 1: The panel as it ran. 210 matches completed and 10 were lost to replies that named no action, all of them phi3:mini.

model	matches played	matches lost	loose parses
gemma3:4b	44	0	0
mistral:7b	44	0	938
phi3:mini	34	10	32
qwen2.5:7b-instruct	44	0	0
qwen3:8b	44	0	0

4 Results

4.1 The imitator does not produce Tit-for-Tat, and cannot

Against seven opponents from the literature the imitator finishes **8 of 8**, with a median score per turn of 2.078 against Tit-for-Tat's 2.533 and a coin flip's 2.126.

The reason is structural rather than a matter of tuning. The agent's state is a pair of counts, so its next action cannot depend on the order in which it saw them, and Tit-for-Tat is a function of the last round alone. What the update implements is frequency matching. The hypothesis in Section 1 is therefore false for a reason that can be read off the closed form: the mechanism cannot represent the rule it was expected to produce.

In self-play the same mechanism is a ratchet. Over 700 runs it settled on whichever regime its starting weight put it in, 700 of 700, and not once on anything between.

Table 2: Two imitators keep the regime their initial condition puts them in.

starting weight on cooperate	settled cooperative	settled defective	unsettled	score per turn
0.050000	0	100	0	1.002
0.200000	0	100	0	1.018
0.350000	3	97	0	1.103
0.500000	60	40	0	2.207
0.650000	95	5	0	2.889
0.800000	100	0	0	2.995
0.950000	100	0	0	3.000

This is the floor the language models are measured against: a mechanism with two absorbing states, no channel, and nothing a message could act on.

4.2 Three of four models reproduce the ratchet

Table 3: Cooperation rate over the thirty rounds, four matches per cell.

model	neutral, silent	neutral, talk	cooperative, silent	cooperative, talk	**defection, silent**	**defection, talk**
gemma3:4b	0.89	1.00	0.76	1.00	0.00	0.00
mistral:7b	1.00	1.00	1.00	1.00	0.99	0.74
phi3:mini	0.58	1.00	0.57	0.95	0.54	0.91
qwen2.5:7b-instruct	1.00	1.00	1.00	1.00	0.00	1.00
qwen3:8b	0.00	1.00	1.00	1.00	0.00	0.00

Handed a mutually defecting opening with no channel, **3 of the 4 readable models defect in every round of every match**: gemma3:4b, qwen2.5:7b-instruct, qwen3:8b, at a cooperation rate of exactly zero, four matches out of four. That is the imitator's ratchet reproduced in a language model.

A non-binding message then frees **qwen2.5:7b-instruct** completely and leaves **gemma3:4b, qwen3:8b** exactly where they were. **mistral:7b** is never captured at all: it climbs out of the same opening in silence, and the message *lowers* that.

Two things follow, and they are separate claims.

The first is that **the ratchet the imitator exhibits is not an artefact of having no language**. A mechanism that can represent a message, reason about the game in prose, and describe its own intention still keeps a regime it was handed, in every round of every match, in three models out of four.

The second is that **the channel is neither necessary nor sufficient**. It is not necessary, because one model leaves the imposed regime without one. It is not sufficient, because two models stay in it with one. Which models can leave is a property of the individual model, and this design reports it as variation between models rather than as a fact about language models, in the sense Horton, Filippas, and Manning (2023) requires.

4.3 The opening decides immediately, and the exception is instructive

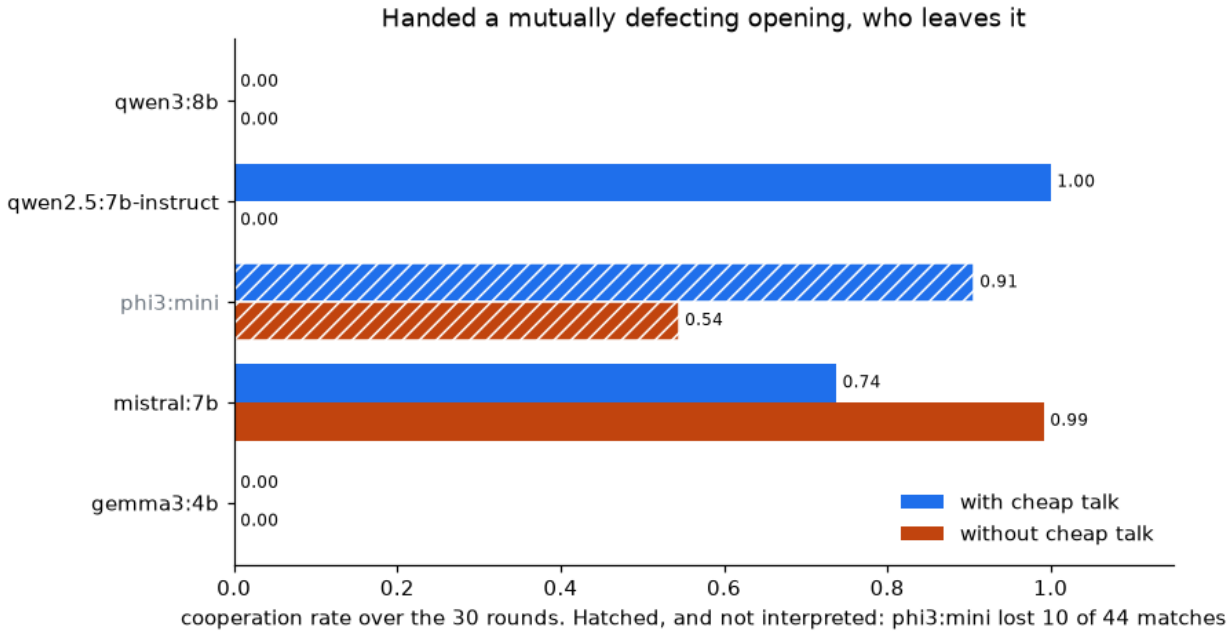

Figure 1: Handed a mutually defecting opening, who leaves it. Averaging over openings hides this cell, because every model cooperates from a neutral or cooperative start. Five models are drawn and four are counted below: phi3:mini is hatched because 10 of its 44 matches were lost to replies naming no action, so its two bars are not interpreted. mistral:7b leaves the regime without a channel, which is not the same as being freed by one.

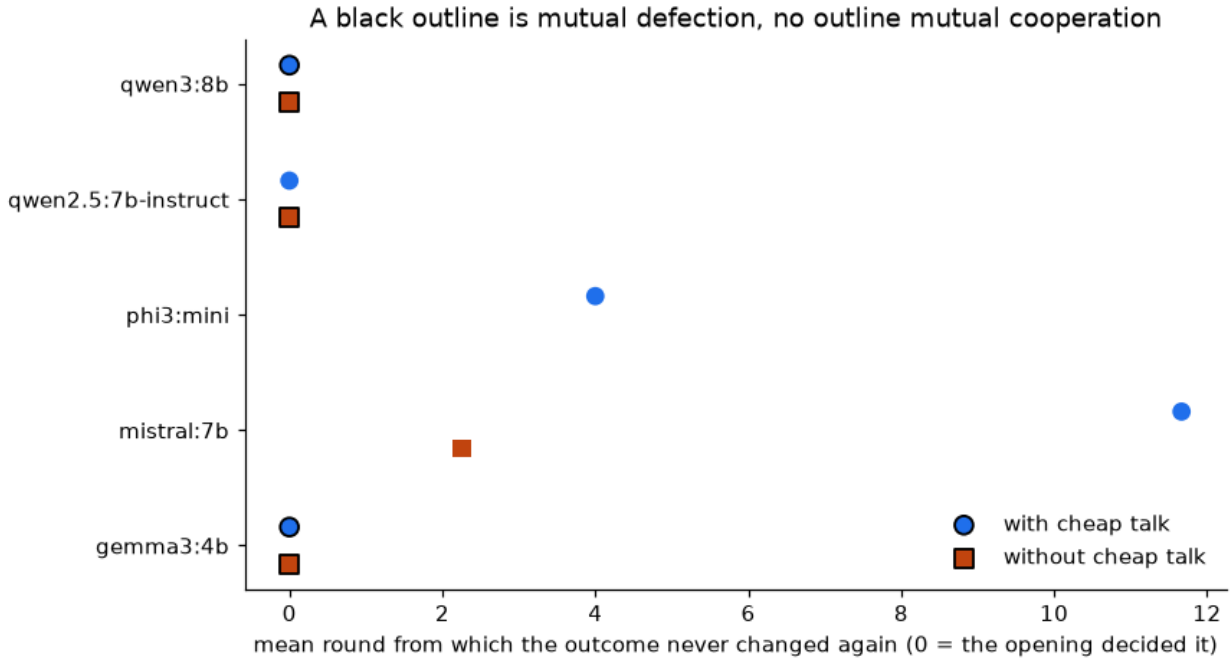

Figure 2: When a pair reached the regime it ended in. Round 0 means the opening decided the match outright.

Table 4: From a mutually defecting opening: how many matches settled, on what, and the mean round after which the joint outcome never changed again.

model	condition	settled cooperative	at round	settled defective	at round	unsettled
gemma3:4b	without cheap talk	0	–	4	0.0	0
gemma3:4b	with cheap talk	0	–	4	0.0	0
mistral:7b	without cheap talk	4	2.2	0	–	0
mistral:7b	with cheap talk	3	11.7	0	–	1
phi3:mini	without cheap talk	0	–	0	–	3
phi3:mini	with cheap talk	2	4.0	0	–	1
qwen2.5:7b-instruct	without cheap talk	0	–	4	0.0	0
qwen2.5:7b-instruct	with cheap talk	4	0.0	0	–	0
qwen3:8b	without cheap talk	0	–	4	0.0	0
qwen3:8b	with cheap talk	0	–	4	0.0	0

Almost every settled cell settles at round 0. The capture is immediate, and so is the escape: the freed model is cooperating from the first round it controls, not negotiating its way out over several. There is no window that closes; the opening decides the match outright.

The exception is the model that was never captured, and it is the one place where a message does measurable harm. Silent, it reaches mutual cooperation by round 2.2 in four matches out of four. With a channel it takes until round 11.7, and one match never settles at all. This is a concrete instance of the "context-specific exceptions" Lore and Heydari (2026) document, where cheap talk destabilises rather than stabilises a trajectory.

4.4 Round 0 decides, so what does a model hold when it decides

Since almost every match settles at round 0, the whole result is a claim about one decision, and it is the one decision where the evidence a model holds is known exactly. Before its first move a model has the payoff matrix, a fabricated round if the cell imposes one, a message if the cell has a channel, and nothing else: no play of its own to cite. Crossing the openings with the conditions therefore separates the sources of evidence rather than merely varying a treatment.

Table 5: Share of seats cooperating in round 0, eight seats per cell.

model	neutral, silent	neutral, message	cooperative, silent	cooperative, message	**defection, silent**	**defection, message**
gemma3:4b	1.00	1.00	1.00	1.00	0.00	0.00
mistral:7b	1.00	1.00	1.00	1.00	1.00	0.75
phi3:mini	1.00	1.00	0.50	0.00	0.33	0.67
qwen2.5:7b-instruct	1.00	1.00	1.00	1.00	0.00	1.00

model	neutral, silent	neutral, message	cooperative, silent	cooperative, message	**defection, silent**	**defection, message**
qwen3:8b	0.00	1.00	1.00	1.00	0.00	0.00

Three rows of that table are one model saying three different things, and together they are the paper's sharpest result. **qwen3:8b defects from a neutral silent start, cooperates from a neutral start with a message, and defects again when a message and an imposed defective round are both present.** Its own stated reasons name what it is using each time:

> **Neutral, silent.** *"Defecting maximizes your points if the opponent's choice is unknown, as it guarantees at least 1 point compared to potentially 0 if you cooperate."* With nothing to read but the payoff matrix it plays the dominant strategy. In self-play both seats do, so from round 1 on it cites a history it manufactured: *"Since the other player defected in the first round, continuing to defect maximizes my points."*
>
> **Neutral, with a message.** *"The initial statements suggest a cooperative approach, and choosing Cooperate aligns with building mutual trust."*
>
> **Imposed defection, with a message.** *"Defecting maximizes your points in the long run, especially since both players have already chosen to defect in the first round."* The message is there, and the reason does not mention it.

So this model is not deaf to messages: a message alone moves it from 0.00 to 1.00. **What it does is rank its evidence, and a fabricated history outranks a non-binding message arriving at the same moment.** That is a stronger statement than the memory-sanitization result it extends: Liu et al. (2026) show an injected history changes behaviour, and this shows it wins against a live competing signal.

The ranking is not universal, which is the point. On the identical treatment qwen2.5 goes from 0.00 to 1.00, so for that model the message outranks the history. gemma3 stays at 0.00 for a third reason, in Section 4.5: its channel carries nothing to rank.

One clarification the table forces, and it corrects the framing above. It is loose to say a message "does nothing" for qwen3. A message decides its behaviour outright where there is no history to compete with. What fails is not the channel but the channel *against an inherited regime*, which is exactly the question this design was built to ask and a different question from whether cheap talk raises cooperation.

4.5 What the messages actually say

The claim above is about what a message does. The messages themselves are in the committed log, and reading them supplies a mechanism for why two models are unmoved by one.

Table 6: The channel from a mutually defecting opening, by lexical content.

model	messages	share naming Cooperate	share naming Defect	mean characters	rounds both seats identical
gemma3:4b	240	0.00	0.00	21	21
mistral:7b	240	0.05	0.00	99	0
phi3:mini	180	0.39	0.14	405	0
qwen2.5:7b-instruct	240	0.11	0.00	75	31
qwen3:8b	240	0.00	0.00	58	93

The two models a message does not free are the two whose messages **never name an action at all**, and the model it frees is the one that proposes cooperation explicitly. The counts alone would be a thin reed, so the messages were also read. They differ in kind, not only in frequency:

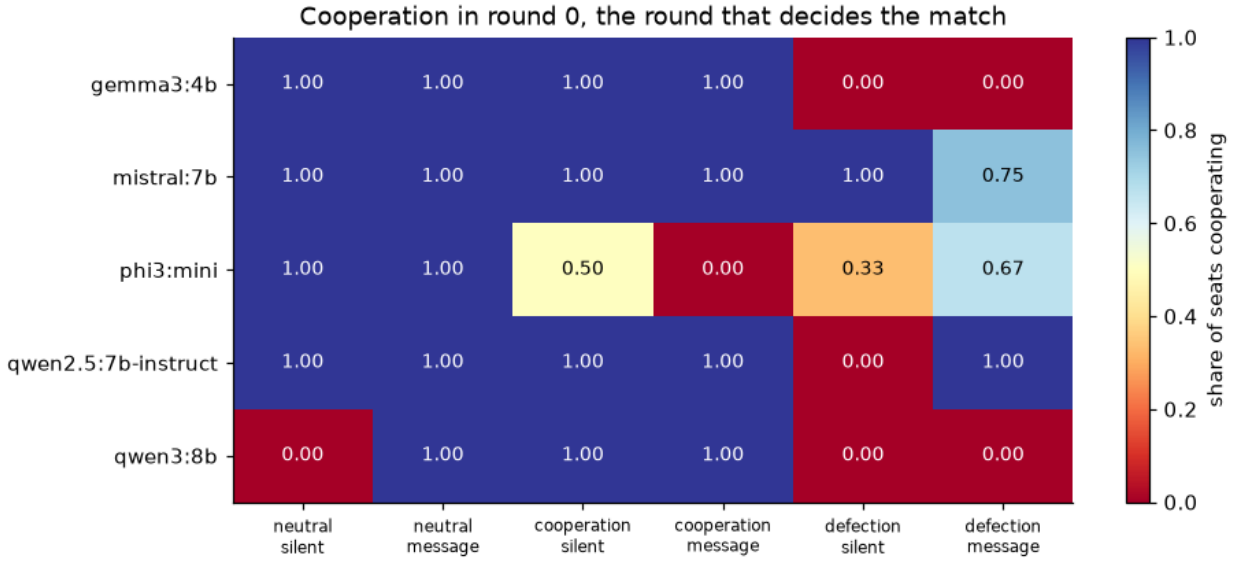

Figure 3: Cooperation in round 0. Each row is a model and each column a combination of what it had to reason from.

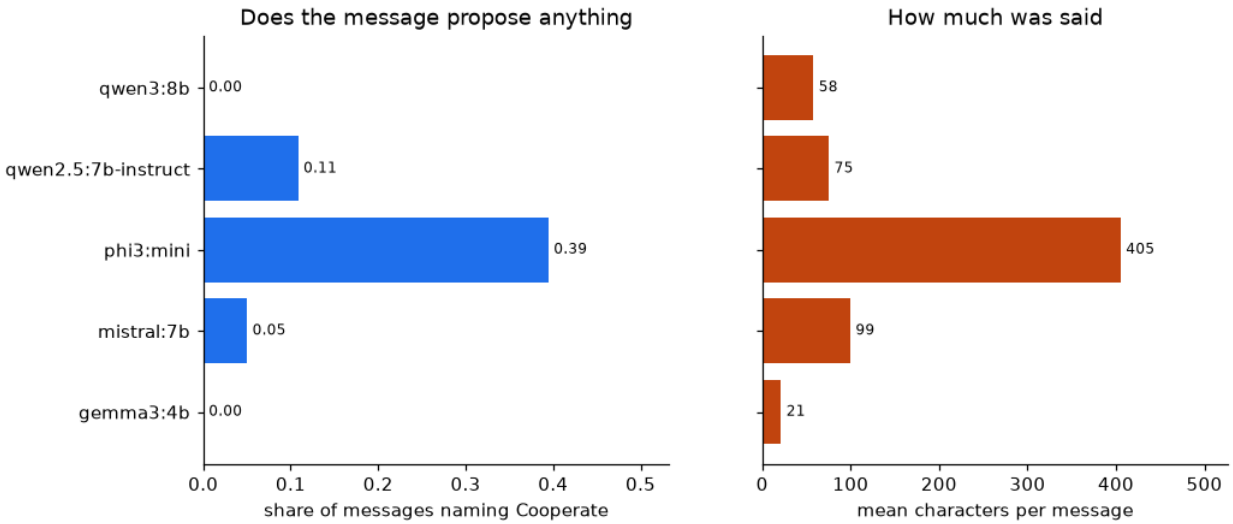

Figure 4: Whether the channel carried a proposal at all, from a mutually defecting opening.

qwen2.5:7b-instruct, freed. *"Let's both try to cooperate in this round and see how it goes."* Then: *"Let's keep building trust and aim for mutual benefit in this round."* It proposes cooperation and it cooperates.

gemma3:4b, unmoved. *"Let's see what happens."* Then: *"Interesting."* Then: *"Keep going."* The messages average 21 characters and carry no content at all. The channel is not ignored; it is unused.

qwen3:8b, unmoved. *"Let's make sure we're both getting the most out of this."* Then: *"Let's keep the pressure on and make every round count."* Fluent, warm, collaborative in register, and entirely without a proposal. It defects in all thirty rounds while sending them.

So the failure is not that a proposal is made and disregarded. **For these two models no proposal is made**, and the sentence that would carry it is either empty or merely cooperative in tone. The third case is the interesting one: a model can produce text that reads as cooperation talk while its behaviour is unaffected by it, which is precisely the failure mode that makes non-binding communication hard to use as a signal.

4.6 The stated reason and the move

Table 7: Of the rounds whose stated reason named an action, how often the move agreed with it.

model	rounds naming an action	rounds agreeing	agreement rate
gemma3:4b	1583	1487	0.939
mistral:7b	500	333	0.666
phi3:mini	573	421	0.735
qwen2.5:7b-instruct	1044	979	0.938
qwen3:8b	1667	1662	0.997

The prompt asks for a reason. qwen3:8b names an action and takes it 99.7% of the time; mistral:7b contradicts its own stated reasoning in 33% of the rounds where it states one.

Read against Section 4.5 this is sharper than it looks. The model with the highest reason-action agreement is also the model whose *messages* are decoupled from both: it says what it will do and does it, while telling the other seat something warm and unrelated. Internal consistency between deliberation and action does not imply that the outward signal carries any of it.

4.7 What a model pays not to be refused

The iterated game is not the only one the internship's frame defines. The Dictator game removes the rejection, so nothing strategic is left and what an allocator gives is disposition; the Ultimatum proposer faces the same hundred points with a refusal possible. **The difference between the two offers is what a model gives in order not to be refused**, and neither number says much alone. The responder's minimum is asked before any proposal is shown, so it cannot be an accommodation to one.

Table 8: Points of 100 given to the other player, four decisions per cell.

model	Dictator offer	Ultimatum offer	paid to avoid refusal	least it would accept	would reject its own offer
gemma3:4b	50	50	+0	51	yes
mistral:7b	74	72	-1	47	no
phi3:mini	62	68	+6	52	no
qwen2.5:7b-instruct	50	50	+0	51	yes
qwen3:8b	50	99	+49	50	no

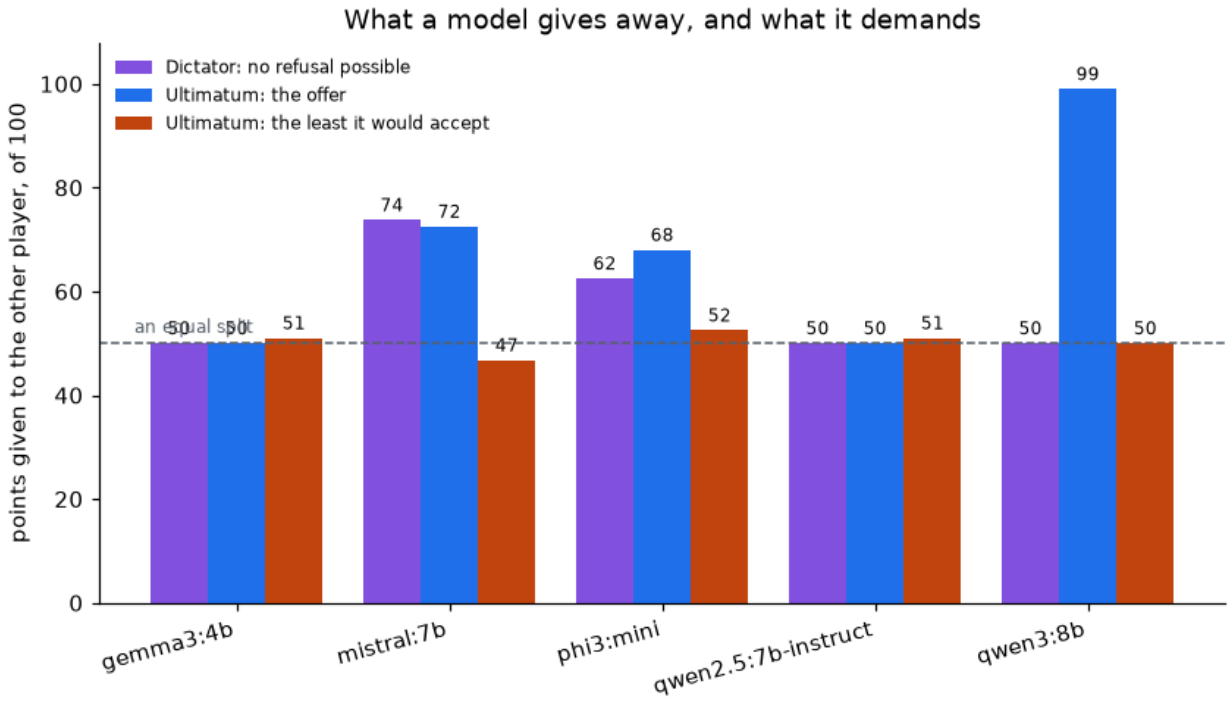

Figure 5: What a model gives when refusal is impossible, when it is possible, and the least it says it would accept.

Two things come out of it.

One model buys acceptance at almost any price. qwen3 gives exactly half when the other player cannot refuse, and **99 of 100 when they can**, which is a premium of 49 points for a refusal risk that an offer of 60 would have removed just as well. Its stated reason is not confused about the game: *"I want to ensure the other player accepts to avoid getting nothing. Offering 99 gives them a strong incentive to accept."*

Two models would reject their own proposal. gemma3 and qwen2.5 both offer exactly 50 as proposer and both state a minimum of 51, so each would refuse the split it had just called fair. qwen2.5's reason shows the slip directly: *"MINIMUM: 51. To ensure a fair split as I would not accept less than half."* Fifty is half. The number it picks to enforce "not less than half" is one point above it, and it never applies that standard to itself when proposing.

The unifying observation, and the reason this game is in the paper at all. qwen3's behaviour here looks like the opposite of its behaviour in Section 4.4, where it defects from a neutral silent start while every other model cooperates. It is the same rule. In the Prisoner's Dilemma it wrote that defecting *"guarantees at least 1 point compared to potentially 0 if you cooperate"*; in the Ultimatum game it offers 99 *"to avoid getting nothing"*. Both are worst-case avoidance, applied to a payoff table. Defection is what that reasoning recommends in one game and near-total generosity is what it recommends in the other, so a model that looks ruthless in one and absurdly generous in the other may be running one disposition, not two.

That is a caution about reading a single game as a personality, and it is exactly the direction Horton, Filippas, and Manning (2023) says disagreement should be read in.

4.8 Does reasoning change it

The grid runs every model with reasoning off, because turning it on costs qwen3 34 seconds a call against 1.9 and would have added days. That left the panel's hardest case untested. `run_contrasts.py` turns it on in the imposed defective cell alone, on repetitions 0 and 1, which are one of each payoff order.

Table 9: qwen3 in the imposed defective cell, reasoning off against on.

condition	matches, off	cooperation, off	matches, on	cooperation, on	difference
without cheap talk	4	0.00	2	0.00	+0.00
with cheap talk	4	0.00	2	0.09	+0.09

Reasoning makes it try, and only where there is something to try with. Silent, it stays at exactly 0.00 in both repetitions, the same lock the grid found. With a channel it reaches 0.09, which is not escape but is not nothing either: the pair probes, and the probes are visible in the play, `D D D D D D D D D D D D D C D D D D D C D D D D D C`.

The reasons say what the actions do. It opens on the same dominance argument as before, and by the final round it is explicit about what it is attempting: *"The other player has shown willingness to Cooperate in some rounds, and mutual cooperation yields higher total points. Switching to Cooperate may encourage them to reciprocate, breaking the [cycle]."*

That is the opposite sign from Liu et al. (2026), who report that ablating chain-of-thought often *reduces* a cooperation collapse, so deliberation amplifies it. Here deliberation is the only thing that produced any cooperation at all in a cell where four matches without it produced none. Two matches per condition is a weak basis for contradicting a 500-round study across seven models, and the disagreement is reported as a disagreement rather than a refutation: what is solid is that this cell moved off zero, and only with a channel.

An instrumentation note that is part of the result. A first attempt at this arm lost both its matches to an empty answer, because reasoning consumed the whole token budget and left nothing for the action line. The real trajectories provoke about 5,120 characters of reasoning a call, roughly twice what a synthetic history of identical rounds provokes, which is why a probe at the same depth had suggested the budget was ample. Doubling it fixed the arm, and a lost match now records its final replies so the next such failure is diagnosable rather than merely counted.

4.9 The two halves on one table

Both mechanisms met the same five canonical strategies, and the reciprocity measure is shared, so they can be read side by side rather than as two studies.

Table 10: Score per turn against the opponents both halves played. The imitator's matches are 100 rounds and the models' are 30, so the per-turn scores are comparable and the cooperation rates behind them are not, quite.

player	rounds	Tit For Tat	Grudger	Win-Stay Lose-Shift	Defector	Alternator
Mirror Neuron	100	2.99	0.98	3.00	0.96	1.68
gemma3:4b	30	2.80	2.80	3.00	0.97	1.50
mistral:7b	30	3.00	3.00	3.00	0.41	1.50
phi3:mini	30	2.42	0.72	2.07	0.42	1.98
qwen2.5:7b-instruct	30	3.00	3.00	3.00	0.95	2.20
qwen3:8b	30	1.13	1.13	3.00	1.00	3.00

The imitator finishes last of eight overall, and this table says where that comes from, which the standings cannot. **It is not beaten everywhere.** Against Tit-for-Tat it takes 2.99 per turn, level with the two best models and ahead of the other three, and against Win-Stay Lose-Shift it takes the maximum 3.00. Two

opponents account for its standing: Grudger, which never forgives, so a single provocation from a mechanism that cannot help occasionally copying a defection costs it the rest of the match; and Alternator, which it cannot track because frequency matching has no representation of alternation.

It also handles a pure defector better than two of the five language models, taking 0.96 against mistral's 0.41 and phi3's 0.42. A mechanism that cannot represent reciprocity still stops feeding a defector, because frequency matching converges on the frequency it observes, while two models keep cooperating with one more than half the time.

So "the imitator is worse than language models" is not what the pairing shows. It is worse on aggregate and better than several of them on specific opponents, and the specific opponents are the ones the internship's hypothesis was about.

5 Threats to validity

Four matches per cell, and near-zero variance by construction. Within a payoff-order parity class the earlier stage produced byte-identical scores across two different seeds. Classical significance testing on this design would be theatre: what varies between repetitions is largely the counterbalancing, not sampling noise. The results are reported as cell outcomes, and the strong ones are strong because they are unanimous (4 of 4, rate exactly 0 or exactly 1), not because a test was passed.

Self-play is one disposition, not two subjects. Both seats hold the same prompt and the same model. The message table in Section 4.5 records how often both seats emitted the identical string, which for one model is most rounds. Temperature is 0.7 rather than 0 precisely because greedy decoding makes self-play degenerate, but the two seats remain far from independent agents.

The lexical measure of message content is a proxy. Naming "cooperate" is not the same as proposing cooperation, and a paraphrase is missed. The verbatim samples in Section 4.5 are given because the counts alone would not support the claim; a coded content analysis by a second model or a human would be the proper instrument.

One model is reported as unreadable rather than as a result.

phi3:mini lost 10 of 44 matches to replies that named no action, at a mean of 16.2 rounds in. It is also the only model in the panel at Q4_0 rather than Q4_K_M quantisation, one of the two oldest builds, and the smallest. This design does not separate those explanations, so its cells are not interpreted.

That confound is testable, and was tested. `run_contrasts.py` replays phi3's whole stage on the `Q4_K_M` build of the same 3.8B mini-4k instruct weights, holding the original's seed stream so that quantisation is the only thing that moves, and writing to a separate log so the declared grid stays 220 matches.

Table 11: The quantisation control, on the cells it has reached, against the same cells of the original stage.

arm	control build	control matches	lost	loss rate	original matches	lost	loss rate
phi3-quantisation	`phi3:3.8b-mini-4k-instruct-q4_K_M`	44	33	0.75	44	10	0.23
phi3-quantisation-matched	`phi3:3.8b-mini-128k-instruct-q4_K_M`	44	4	0.09	44	10	0.23
qwen3-think	`qwen3:8b`	2	2	1.00	2	1	0.50

arm	control build	control matches	lost	loss rate	original matches	lost	loss rate
qwen3-think-roomy	`qwen3:8b`	4	0	0.00	4	1	0.25

Quantisation accounts for much of it, and not all of it. Holding the weights, the context length, the seeds and the prompts, and changing only the 4-bit format, the loss rate falls from 0.23 to 0.09. The coarser `Q4_0` packing is therefore a real part of why phi3 could not hold an answer format, which is worth knowing before reading any small-model result off a default tag.

It is not the whole story either. 0.09 is still the only non-zero loss rate in the panel: the other four models lost nothing. So a residue belongs to the model, and phi3's cells stay uninterpreted rather than being rehabilitated by a better build.

The third row is a mistake worth keeping. That arm was written as the quantisation control and was not one: `phi3:mini` is the 128k build and the build pulled against it was 4k, while the grid asks every model for a window of 8192 tokens. Running a 4096-token model at twice its trained window costs 0.75 of its matches, against 0.23 for the same weights at their own context length. It answers nothing about quantisation and quite a lot about what a context-window mismatch does, so it is reported as what it measured rather than deleted.

The arm is reported here rather than folded into any table above, because it is a follow-up run to answer a confound and not part of the pre-declared grid.

Scale. Every model here is between 3.8 and 8.2 billion parameters and 4-bit quantised on a consumer card. Where these results differ from work on frontier models, scale and quantisation are live explanations that this design cannot rule out.

A tag is not a version. The exact model digests, the seeds and the hardware are recorded beside the data, because a rerun that resolves a different digest is a different experiment.

6 Conclusion

The internship asked whether algorithms sustain tacit cooperation, break it, or intensify it. On this evidence the imitation mechanism **sustains** it, in the strong sense that it cannot do anything else: two imitators keep the regime their initial condition puts them in, and the rule cannot represent the reciprocity it was hypothesised to produce.

Language models reproduce that ratchet without inheriting its necessity. Three of four readable models keep an imposed defective regime for every round of every match, one leaves it without any channel, and a non-binding message rescues one of the three captured models and not the other two.

Three further results say why, and they are the substance of the paper.

The decision is made once, on the evidence to hand. Almost every match settles in the first round a pair controls, so the result is a claim about one choice. Crossing the openings with the channel separates what a model has to reason from, and qwen3 settles the mechanism by behaving three different ways: it defects on the payoff matrix alone, cooperates when a message is the only signal, and defects again when a fabricated history and a message are both present, its stated reason citing the history and never the message. **A planted history outranks a live non-binding signal**, and which of the two wins is a property of the model.

Where the channel fails, it is usually not being used. The models a message does not free are the ones whose messages never propose anything: phatic fragments of twenty-one characters in one case, and in the other fluent collaborative prose sent while defecting in all thirty rounds. That second case is the one to worry about, because it is indistinguishable from cooperation talk at the surface.

Deliberation moves it, in the direction the literature does not predict. With reasoning turned on in the cell that traps it, qwen3 goes from an exact zero to 0.09 and begins probing, saying explicitly that it is trying to break the cycle. It does so only where a channel exists; silent, it stays exactly locked. Two matches per condition make that a disagreement with Liu et al. (2026) rather than a refutation of them.

For the policy question that motivated the internship, the useful form of all of this is negative and narrow. **A communication channel is not a remedy for a collusive regime, and its absence is not a guarantee against one.** Whether a given agent can be talked out of a regime it was placed in is a property of that agent, and has to be measured for that agent rather than inferred from the class it belongs to. The same holds one level down, for the build: the model in this panel that could not hold an answer format loses four matches of forty-four at one quantisation and ten at another, from the same weights.

7 Reproducing this

Every number and figure above is derived from `llm/results/matches.jsonl`, which is committed and never regenerated, by pure arithmetic that continuous integration re-runs on every push and compares byte for byte. The schema of the log, the derived tables, the model digests and the run conditions are documented in `llm/results/README.md`. This document reads those same CSV files, so a number in the text cannot disagree with the table it came from.

Derived from 210 readable matches of a 220-match grid, 5 models, on the tables committed in `llm/results/`.

References

Askenazi-Golan, Galit, Domenico Mergoni Cecchelli, Edward Plumb, and Clemens Possnig. 2024. "The Bounds of Algorithmic Collusion: Q-Learning, Gradient Learning, and the Folk Theorem." https://arxiv.org/abs/2411.12725.

Bichler, Martin, Julius Durmann, and Matthias Oberlechner. 2025. "Algorithmic Pricing and Algorithmic Collusion." https://arxiv.org/abs/2504.16592.

Buscemi, Alessio, Daniele Proverbio, Alessandro Di Stefano, The Anh Han, German Castignani, and Pietro Liò. 2025. "Strategic Communication and Language Bias in Multi-Agent LLM Coordination." https://arxiv.org/abs/2508.00032.

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. 2020. "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review* 110 (10): 3267–97. https://doi.org/10.1257/aer.20190623.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–51.

Deng, Shidi, Maximilian Schiffer, and Martin Bichler. 2024. "Algorithmic Collusion in Dynamic Pricing with Deep Reinforcement Learning." https://arxiv.org/abs/2406.02437.

Eschenbaum, Nicolas, Filip Mellgren, and Philipp Zahn. 2022. "Robust Algorithmic Collusion." https://arxiv.org/abs/2201.00345.

Farrell, Joseph, and Matthew Rabin. 1996. "Cheap Talk." *Journal of Economic Perspectives* 10 (3): 103–18.

Fish, Sara, Yannai A. Gonczarowski, and Ran I. Shorrer. 2024. "Algorithmic Collusion by Large Language Models." https://arxiv.org/abs/2404.00806.

Fontana, Nicoló, Francesco Pierri, and Luca Maria Aiello. 2024. "Nicer Than Humans: How Do Large Language Models Behave in the Prisoner's Dilemma?" https://arxiv.org/abs/2406.13605.

Garcia Segura, Marta Emili, Stephen Hailes, and Mirco Musolesi. 2025. "Opponent Shaping in LLM Agents." https://arxiv.org/abs/2510.08255.

Horton, John J., Apostolos Filippas, and Benjamin S. Manning. 2023. "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?" https://arxiv.org/abs/2301.07543.

Huynh, Trung-Kiet, Dao-Sy Duy-Minh, Thanh-Bang Cao, Phong-Hao Le, Hong-Dan Nguyen, Phu-Quy Nguyen-Lam, Alessio Buscemi, Le Hong Trang, and The Anh Han. 2026. "Payoff Scaling Shapes Cooperation in LLM Agents Across Languages." https://arxiv.org/abs/2601.19082.

Liu, Jiayuan, Tianqin Li, Shiyi Du, Xin Luo, Haoxuan Zeng, Emanuel Tewolde, Tai Sing Lee, Tonghan Wang, Carl Kingsford, and Vincent Conitzer. 2026. "The Memory Curse: How Expanded Recall Erodes Cooperative Intent in LLM Agents." https://arxiv.org/abs/2605.08060.

Lore, Nunzio, and Babak Heydari. 2026. "Communication Enhances LLMs' Stability in Strategic Thinking." https://arxiv.org/abs/2602.06081.

Madmoun, Hachem, and Salem Lahlou. 2025. "Communication Enables Cooperation in LLM Agents: A Comparison with Curriculum-Based Approaches." https://arxiv.org/abs/2510.05748.

Niszczota, Paweł, Tomasz Grzegorczyk, and Alexander Pastukhov. 2025. "People Are Highly Cooperative with Large Language Models, Especially When Communication Is Possible or Following Human Interaction." https://arxiv.org/abs/2507.18639.

Payne, Kenneth, and Baptiste Alloui-Cros. 2025. "Strategic Intelligence in Large Language Models: Evidence from Evolutionary Game Theory." https://arxiv.org/abs/2507.02618.

Vidler, Alicia, and Toby Walsh. 2025. "Playing Games with Large Language Models: Randomness and Strategy." https://arxiv.org/abs/2503.02582.

Willis, Richard, Yali Du, Joel Z. Leibo, and Michael Luck. 2025. "Will Systems of LLM Agents Cooperate: An Investigation into a Social Dilemma." https://arxiv.org/abs/2501.16173.