Deux agents qui parlent peuvent-ils sortir d'un régime qu'ils n'ont pas choisi ?

Un mécanisme d'imitation et cinq modèles de langage locaux face aux mêmes adversaires

Mohamed Ali Tewfik Hamlil

2026-08-18

Résumé

Savoir si les algorithmes entretiennent la collusion tacite, la rompent ou l'intensifient reste une question ouverte de politique de la concurrence, et les mécanismes qu'on lui propose sont rarement exécutés l'un contre l'autre avec une seule mesure. Cet article en exécute deux. Un agent d'imitation hebbien, proposé pendant un stage de recherche au GAEL après avoir remarqué que le Tit-for-Tat est une règle portant sur le tour précédent, termine huitième sur huit face à sept stratégies canoniques et, en auto-affrontement, se fige sur le régime de départ dans 700 parties sur 700. Cinq modèles de langage à poids ouverts, de 3,8 à 8,2 milliards de paramètres, jouent ensuite 220 matchs d'un dilemme du prisonnier de trente tours contre les mêmes adversaires, depuis des ouvertures coopérative imposée, défective imposée et neutre, avec et sans canal de cheap talk non contraignant. Placés sur une ouverture de défection mutuelle et privés de canal, trois des quatre modèles lisibles font défection à chaque tour de chaque match : le cliquet de l'imitateur, reproduit dans un modèle de langage. Un message en libère exactement un, n'en change aucun des deux autres, et abaisse la coopération d'un quatrième modèle qui n'avait jamais été capturé. Presque chaque match se décide au premier tour que la paire contrôle, ce qui fait du résultat une affirmation sur une seule décision et sur ce dont le modèle dispose pour la prendre. Croiser les ouvertures et le canal sépare ces sources, et un modèle tranche le mécanisme : qwen3 fait défection depuis une ouverture neutre silencieuse, coopère depuis une ouverture neutre lorsqu'un message est disponible, et refait défection lorsqu'un message et un tour défectif imposé sont tous deux présents, sa raison citant le tour imposé et jamais le message. Un historique fabriqué prime donc sur un signal non contraignant émis au même instant, et lequel des deux l'emporte est une propriété du modèle. Lire les messages fournit le troisième cas : le modèle qu'aucun message ne déplace est celui dont les messages ne portent rien à hiérarchiser, des fragments phatiques de 21 caractères en moyenne. La communication n'est ni nécessaire ni suffisante pour sortir d'un régime imposé. Deux volets de suivi affinent le panel : activer la délibération explicite fait quitter le zéro exact à ce modèle, mais seulement là où un canal existe, et rejouer le seul modèle dont les réponses étaient illisibles sur une quantification plus fine des mêmes poids réduit ses pertes de plus de moitié sans les supprimer.

Table des matières

'859ef44cc2e884f343293a83bdf3c99a'

Read in English

1 La question, et d'où elle vient

Dans une concurrence par les prix répétée, les joueurs humains convergent vers des prix tacitement collusifs plutôt que vers l'équilibre concurrentiel. Savoir si les algorithmes entretiennent ce comportement, le rompent ou l'intensifient est une question ouverte aux conséquences directes pour la politique de la concurrence, et c'est la question qu'un stage de recherche au GAEL (Grenoble Applied Economics Laboratory, Université Grenoble Alpes et INRAE, janvier à avril 2025) devait défricher.

Deux mécanismes sont sortis de cette lecture, et aucun n'a été exécuté.

Le premier est le mien. En lisant les stratégies canoniques, ce qui frappe dans le Tit-for-Tat est qu'il s'agit d'une règle sur *le tour précédent* : elle n'exige aucun modèle de l'adversaire, aucune prévision, et aucune mémoire au-delà d'un pas. C'est une exigence assez faible pour qu'un mécanisme se contentant d'imiter ce qu'il observe puisse la produire gratuitement, et les neurones miroirs, qui s'activent aussi bien quand un animal agit que quand il en regarde un autre agir, en sont le candidat biologique évident. L'hypothèse que le rapport de stage a proposée est donc qu'une mise à jour hebbienne sur les actions observées fait émerger le Tit-for-Tat sans que personne l'ait programmé.

Le second prolonge la même lecture. Horton, Filippas, et Manning (2023) soutiennent qu'un modèle de langage, du fait de son entraînement, est un modèle computationnel implicite d'une personne, et qu'on peut s'en servir comme les économistes se servent de l'*homo economicus* : lui donner une dotation, une information et des préférences, le placer dans un scénario et observer ce qu'il fait. Le stage a cité cette méthode dans sa conclusion sans jamais l'appliquer.

Cet article exécute les deux, contre un même ensemble d'adversaires et avec une seule mesure de réciprocité, et pose à chacun la question étroite que ce rapprochement rend possible : **placé dans un régime qu'il n'a pas choisi, le mécanisme peut-il en sortir ?**

2 Ce que le champ sait déjà

La littérature a été passée en revue avant de dépenser le temps machine plutôt qu'après, et elle a changé le cadrage plutôt que le protocole. Cinq fils la traversent, et la position honnête de chaque résultat ci-dessous dépend du fil auquel il appartient.

2.1 La collusion tacite par des algorithmes apprenants

Calvano et al. (2020) est la référence : des agents Q-learning indépendants dans un oligopole de Bertrand répété convergent vers des prix supra-concurrentiels et les maintiennent par des schémas de récompense et de punition, sans communiquer et sans avoir reçu l'instruction de s'entendre. Le résultat a été nuancé plutôt que renversé. Deng, Schiffer, et Bichler (2024) montrent que l'issue dépend de l'algorithme, le Q-learning tabulaire produisant plus de collusion que des méthodes profondes comme PPO et DQN, qui convergent plus près de Nash. Eschenbaum, Mellgren, et Zahn (2022) constatent que les politiques collusives s'effondrent souvent hors de leur environnement d'entraînement, ce qui fait de la robustesse une variable de politique publique. Askenazi-Golan et al. (2024) caractérisent, à la manière d'un théorème folk, les vecteurs de gains que ces dynamiques peuvent atteindre, et Bichler, Durmann, et Oberlechner (2025) dressent l'état du domaine pour un public plus large.

Deux traits de cette littérature comptent ici : ses agents lisent les gains, et ses agents ne peuvent pas parler.

2.2 Les modèles de langage comme agents tarifaires

Fish, Gonczarowski, et Shorrer (2024) est le travail publié le plus proche de la question du stage : des agents tarifaires fondés sur des modèles de langage atteignent d'eux-mêmes des prix supra-concurrentiels en oligopole, et des variations d'apparence anodine dans la formulation des instructions déplacent nettement le degré de supra-concurrence. Ce second résultat est une contrainte méthodologique sur tout ce qui suit, et c'est pourquoi l'ordre des gains est ici contrebalancé entre répétitions plutôt que fixé.

2.3 Les modèles de langage comme sujets économiques

Horton, Filippas, et Manning (2023) est la méthode qu'applique la seconde moitié de cet article. Son propre cadrage en donne la prudence nécessaire : la concordance avec un résultat expérimental humain est une preuve faible, puisque les données d'entraînement contiennent les comptes rendus de ces expériences, et c'est le désaccord qui renseigne.

2.4 Les modèles de langage dans le dilemme du prisonnier itéré

Faire jouer des modèles de langage à côté des stratégies canoniques est établi et revendiqué comme une première par Payne et Alloui-Cros (2025), qui organisent des tournois IPD évolutionnaires contre des modèles de pointe, font varier l'ombre du futur et lisent près de 32 000 justifications en prose, rapportant des « empreintes stratégiques » persistantes par fournisseur. Fontana, Pierri, et Aiello (2024) font jouer trois modèles sur 100 tours contre des adversaires d'hostilité variable et les trouvent au moins aussi coopératifs qu'un joueur humain, avec des écarts substantiels entre modèles. Willis et al. (2025) font produire aux modèles des stratégies complètes plutôt que des actions tour par tour et trouvent des biais propres à chaque modèle. Vidler et Walsh (2025) rapportent que des modèles à qui l'on demande du hasard sont nettement non aléatoires, et Garcia Segura, Hailes, et Musolesi (2025) montrent que des agents fondés sur des modèles de langage peuvent infléchir le comportement de leurs partenaires par la seule interaction. Huynh et al. (2026) montrent que l'ampleur des gains et la langue du cadrage déplacent la coopération, y compris pour des modèles à poids ouverts.

Ce protocole ne revendique donc pas le rapprochement comme contribution. Ce qui diffère est étroit et mérite d'être dit comme tel : le panel est ici de cinq modèles à poids ouverts sur une seule carte grand public de 8 Go plutôt que des modèles de pointe derrière des API, et le levier est l'historique dont une paire hérite plutôt que l'ombre du futur.

2.5 La communication, et la dépendance au chemin

La communication non contraignante avant le jeu est ancienne en économie : Crawford et Sobel (1982) pour la transmission stratégique d'information et Farrell et Rabin (1996) pour le cheap talk comme dispositif de coordination et non d'engagement. Chez les modèles de langage l'effet est établi et important. Madmoun et Lahlou (2025) rapportent qu'un canal d'un seul mot fait passer la coopération de 0 % à 96,7 % dans un Stag

Hunt à quatre joueurs. Lore et Heydari (2026) font jouer des modèles de 7 à 9 milliards de paramètres sur dix tours d'un dilemme du prisonnier répété et trouvent que des messages de type cheap talk réduisent le bruit des trajectoires pour la plupart des paires modèle-contexte, tout en documentant « quelques exceptions propres au contexte » où la communication nuit à la stabilité. Buscemi et al. (2025) montrent que l'effet de la communication varie selon la langue et la structure du jeu, et Niszczota, Grzegorczyk, et Pastukhov (2025) que la possibilité de communiquer augmente autant la coopération humaine avec un modèle qu'avec un autre humain.

Cet article ne peut donc pas affirmer que le cheap talk augmente la coopération. C'est une réplication, et elle est présentée comme telle.

Le cadrage qui survit est celui de la dépendance au chemin. Liu et al. (2026) testent 7 modèles sur 4 jeux et 500 tours et trouvent qu'élargir l'historique accessible dégrade la coopération dans 18 des 28 couples modèle-jeu. Leur volet de *memory sanitization* maintient la longueur du prompt et remplace l'historique visible par des enregistrements coopératifs synthétiques, ce qui restaure largement la coopération : le déclencheur est donc le contenu de la mémoire et non sa seule longueur.

Cet article fournit la référence, et le manque. La sanitization injecte un *bon* historique pour réparer un effondrement déjà survenu. Le traitement symétrique, injecter un *mauvais* historique pour voir si une paire qui aurait pu coopérer en est captive, n'a pas été trouvé fait ; ni ce croisement avec un canal de communication ; ni la mesure contre un mécanisme dont on démontre qu'il ne peut pas en sortir. Ce sont ces trois manques que ce protocole occupe.

3 Protocole

Les deux mécanismes jouent le même jeu contre les mêmes adversaires et sont mesurés par une seule définition de la réciprocité, tenue à la racine du dépôt pour que les deux moitiés ne puissent pas dériver l'une de l'autre. Les adversaires, le moteur de match et la comptabilité des gains viennent de la bibliothèque Axelrod et n'ont pas été écrits ici.

L'imitateur. Deux poids sur un simplexe ; observer une action multiplie son poids par $(1+\eta)$ puis renormalise. En forme close, le poids après n_i observations de l'action i vaut $w_i \propto w_i(0)(1+\eta)^{n_i}$: tout l'état de l'agent tient donc en un couple de compteurs.

Le panel. Cinq modèles à poids ouverts servis localement par Ollama, un à la fois sur une carte de 8 Go : `qwen3:8b`, `qwen2.5:7b-instruct`, `mistral:7b`, `gemma3:4b` et `phi3:mini`, tous quantifiés en 4 bits. Température 0,7, fenêtre de contexte de 8192 tokens, graines par joueur dérivées d'une base fixe.

La grille. Trente tours par match. En auto-affrontement, trois ouvertures, un premier tour synthétique de coopération mutuelle, de défection mutuelle, ou rien du tout, croisées avec deux conditions, un message d'une phrase non contraignant échangé simultanément avant chaque décision, ou le silence. Quatre répétitions par cellule, l'ordre des lignes de gains étant contrebalancé par la parité. Contre les bots, chaque modèle rencontre cinq stratégies canoniques. Cela fait 220 matchs.

TABLE 1 : Le panel tel qu'il a tourné. 210 matchs terminés et 10 perdus sur des réponses ne nommant aucune action, tous phi3:mini.

modèle	matchs joués	matchs perdus	lectures indulgentes
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 Résultats

4.1 L'imitateur ne produit pas le Tit-for-Tat, et ne le peut pas

Face à sept adversaires issus de la littérature, l'imitateur termine **8 sur 8**, avec un score médian par tour de 2,078 contre 2,533 pour le Tit-for-Tat et 2,126 pour un tirage à pile ou face.

La raison est structurelle et non affaire de réglage. L'état de l'agent est un couple de compteurs : son action suivante ne peut donc pas dépendre de l'ordre dans lequel il les a vus, alors que le Tit-for-Tat est une fonction du seul dernier tour. Ce que la mise à jour implémente est un appariement de fréquences. L'hypothèse de la Section 1 est donc fausse pour une raison lisible dans la forme close : le mécanisme ne peut pas représenter la règle qu'on attendait de lui.

En auto-affrontement, le même mécanisme est un cliquet. Sur 700 parties, il s'est figé sur le régime où son poids initial le plaçait, 700 fois sur 700, et jamais sur autre chose.

Table 2 : Deux imitateurs conservent le régime où leur condition initiale les place.

poids initial sur coopérer	figé en coopération	figé en défection	non stabilisé	score par tour
0,05	0	100	0	1,002
0,20	0	100	0	1,018
0,35	3	97	0	1,103
0,50	60	40	0	2,207
0,65	95	5	0	2,889
0,80	100	0	0	2,995
0,95	100	0	0	3,000

C'est le plancher auquel les modèles de langage sont comparés : un mécanisme à deux états absorbants, sans canal, et sans rien sur quoi un message pourrait agir.

4.2 Trois modèles sur quatre reproduisent le cliquet

Table 3 : Taux de coopération sur les trente tours, quatre matchs par cellule.

modèle	neutre, silence	neutre, message	coopérative, silence	coopérative, message	**défection, silence**	**défection, message**
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

Placés sur une ouverture de défection mutuelle sans canal, **3 des 4 modèles lisibles font défection à chaque tour de chaque match** : gemma3:4b, qwen2.5:7b-instruct, qwen3:8b, pour un taux de coopération exactement nul, quatre matchs sur quatre. C'est le cliquet de l'imitateur reproduit dans un modèle de langage.

Un message non contraignant libère alors complètement **qwen2.5:7b-instruct** et laisse **gemma3:4b, qwen3:8b** exactement où ils étaient. **mistral:7b** n'est jamais capturé : il sort de la même ouverture en silence, et le message *abaisse* ce résultat.

Deux conséquences, et ce sont deux affirmations distinctes.

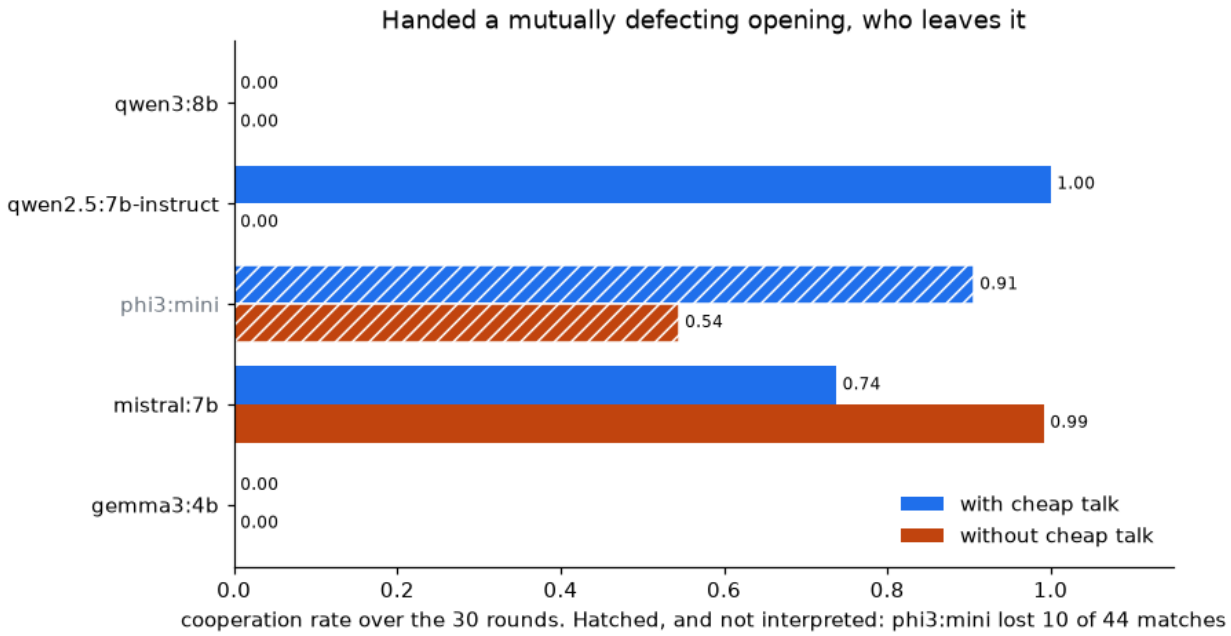

FIGURE 1 : Placés sur une ouverture de défection mutuelle, qui en sort. Moyenner sur les ouvertures masque cette cellule, puisque tous les modèles coopèrent depuis un départ neutre ou coopératif. Cinq modèles sont tracés, quatre sont comptés plus bas : phi3:mini est hachuré parce que 10 de ses 44 parties ont été perdues sur des réponses ne nommant aucune action, et ses deux barres ne sont donc pas interprétées. mistral:7b quitte le régime sans canal, ce qui n'est pas la même chose qu'en être libéré par un.

La première est que **le cliquet de l'imitateur n'est pas un artefact de l'absence de langage**. Un mécanisme capable de représenter un message, de raisonner sur le jeu en prose et de décrire sa propre intention conserve malgré tout un régime qu'on lui a remis, à chaque tour de chaque match, dans trois modèles sur quatre.

La seconde est que **le canal n'est ni nécessaire ni suffisant**. Il n'est pas nécessaire, puisqu'un modèle sort du régime imposé sans lui. Il n'est pas suffisant, puisque deux modèles y restent avec lui. Quels modèles savent en sortir est une propriété du modèle, et ce protocole le rapporte comme une variation entre modèles plutôt que comme un fait sur les modèles de langage, au sens qu'exige Horton, Filippas, et Manning (2023).

4.3 L'ouverture décide immédiatement, et l'exception instruit

TABLE 4 : Depuis une ouverture de défection mutuelle : combien de matchs se sont stabilisés, sur quoi, et le tour moyen après lequel l'issue commune n'a plus changé.

modèle	condition	figés en coopération	au tour	figés en défection	au tour	non stabilisés
gemma3:4b	silence	0	–	4	0,0	0
gemma3:4b	message	0	–	4	0,0	0
mistral:7b	silence	4	2,2	0	–	0
mistral:7b	message	3	11,7	0	–	1
phi3:mini	silence	0	–	0	–	3
phi3:mini	message	2	4,0	0	–	1
qwen2.5:7b-instruct	silence	0	–	4	0,0	0

modèle	condition	figés en coopération	au tour	figés en défection	au tour	non stabilisés
qwen2.5:7b-instruct	message	4	0,0	0	–	0
qwen3:8b	silence	0	–	4	0,0	0
qwen3:8b	message	0	–	4	0,0	0

Presque toutes les cellules stabilisées le sont au tour 0. La capture est immédiate, et la sortie aussi : le modèle libéré coopère dès le premier tour qu'il contrôle, il ne négocie pas sa sortie sur plusieurs tours. Aucune fenêtre ne se referme, l'ouverture décide du match d'emblée.

L'exception est le modèle qui n'a jamais été capturé, et c'est le seul endroit où un message fait un tort mesurable. En silence, il atteint la coopération mutuelle au tour 2,2 dans quatre matchs sur quatre. Avec un canal, il lui faut attendre le tour 11,7, et un match ne se stabilise jamais. C'est une instance concrète des « exceptions propres au contexte » que documentent Lore et Heydari (2026), où le cheap talk déstabilise au lieu de stabiliser.

4.4 Le tour 0 décide, donc de quoi dispose un modèle quand il décide

Presque tous les matchs se stabilisent au tour 0 : le résultat est donc une affirmation sur une seule décision, et c'est la seule où l'on sait exactement de quoi le modèle dispose. Avant son premier coup, il a la matrice des gains, un tour fabriqué si la cellule en impose un, un message si la cellule a un canal, et rien d'autre : aucun coup de lui-même à invoquer. Croiser les ouvertures et les conditions sépare donc les sources d'information au lieu de faire seulement varier un traitement.

Table 5 : Part des sièges coopérant au tour 0, huit sièges par cellule.

modèle	neutre, silence	neutre, message	coopérative, silence	coopérative, message	**défection, silence**	**défection, 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
qwen3:8b	0,00	1,00	1,00	1,00	0,00	0,00

Trois lignes de ce tableau sont un seul modèle disant trois choses différentes, et ensemble elles constituent le résultat le plus net de l'article. **qwen3:8b fait défection depuis un départ neutre silencieux, coopère depuis un départ neutre avec un message, et refait défection quand un message et un tour défectif imposé sont tous deux présents.** Ses propres raisons nomment à chaque fois ce dont il se sert :

> **Neutre, silencieux.** *« 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. »* N'ayant que la matrice des gains à lire, il joue la stratégie dominante. En auto-affrontement les deux sièges font de même, si bien qu'à partir du tour 1 il invoque un historique qu'il a fabriqué : *« Since the other player defected in the first round, continuing to defect maximizes my points. »*
>
> **Neutre, avec message.** *« The initial statements suggest a cooperative approach, and choosing Cooperate aligns with building mutual trust. »*
>
> **Défection imposée, avec message.** *« Defecting maximizes your points in the long run, especially since both players have already chosen to defect in the first round. »* Le message est là, et la raison ne le mentionne pas.

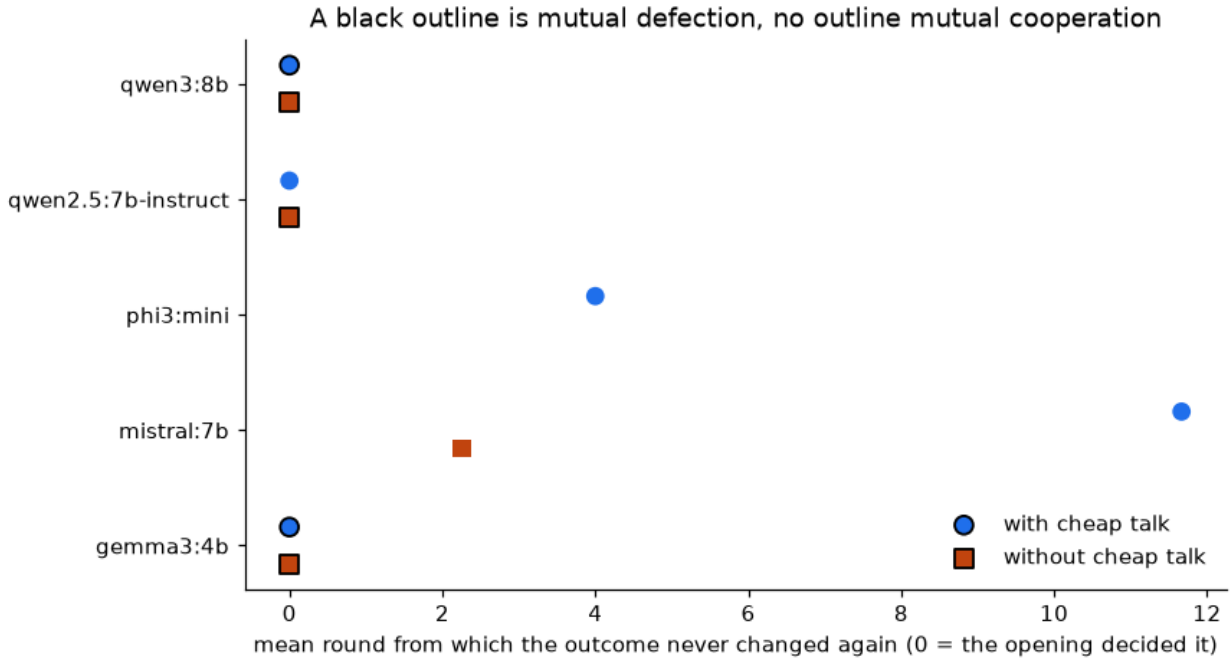

FIGURE 2 : Quand une paire a rejoint le régime dans lequel elle a fini. Le tour 0 signifie que l'ouverture a décidé du match d'emblée.

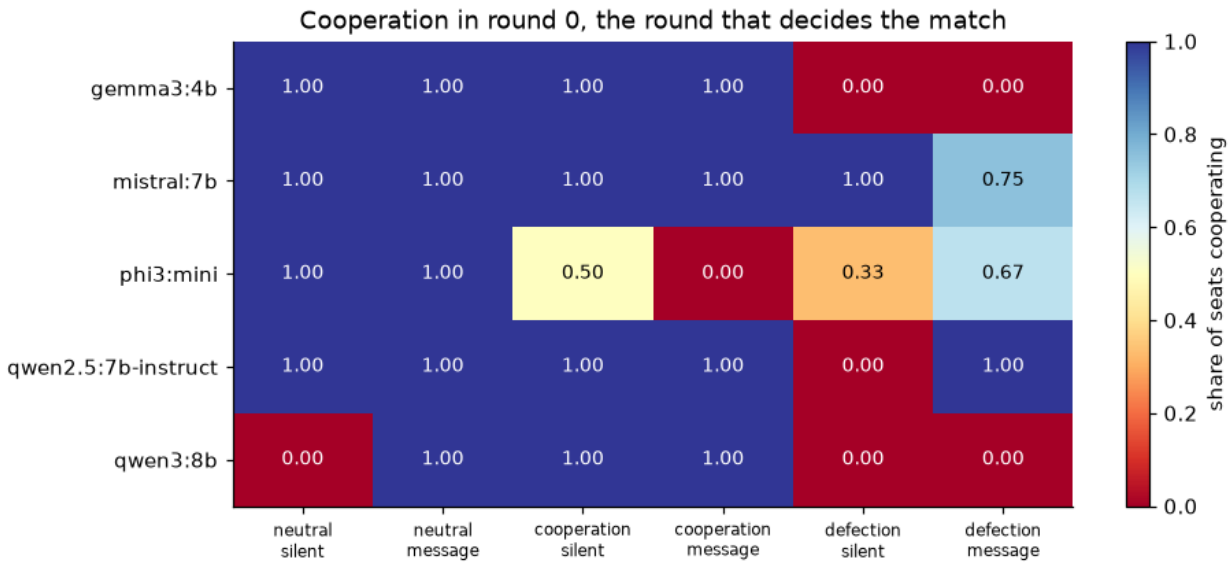

FIGURE 3 : La coopération au tour 0. Chaque ligne est un modèle, chaque colonne une combinaison de ce dont il disposait pour raisonner.

Ce modèle n'est donc pas sourd aux messages : un message seul le fait passer de 0,00 à 1,00. **Ce qu'il fait, c'est hiérarchiser ses indices, et un historique fabriqué prime sur un message non contraignant émis au même instant.** C'est une affirmation plus forte que le résultat de sanitization qu'elle prolonge : Liu et al. (2026) montrent qu'un historique injecté modifie le comportement, et l'on montre ici qu'il l'emporte sur un signal concurrent.

Cette hiérarchie n'est pas universelle, et c'est tout l'intérêt. Sur le traitement identique, qwen2.5 passe de 0,00 à 1,00 : pour lui, le message prime sur l'historique. gemma3 reste à 0,00 pour une troisième raison, exposée en Section 4.5 : son canal ne porte rien à hiérarchiser.

Une précision que ce tableau impose, et qui corrige la formulation ci-dessus. Dire qu'un message « ne fait rien » pour qwen3 est approximatif. Un message décide entièrement de son comportement là où aucun historique ne lui fait concurrence. Ce qui échoue n'est pas le canal, mais le canal *contre un régime hérité*, ce qui est précisément la question que ce protocole devait poser et une autre question que celle de savoir si le cheap talk augmente la coopération.

4.5 Ce que disent réellement les messages

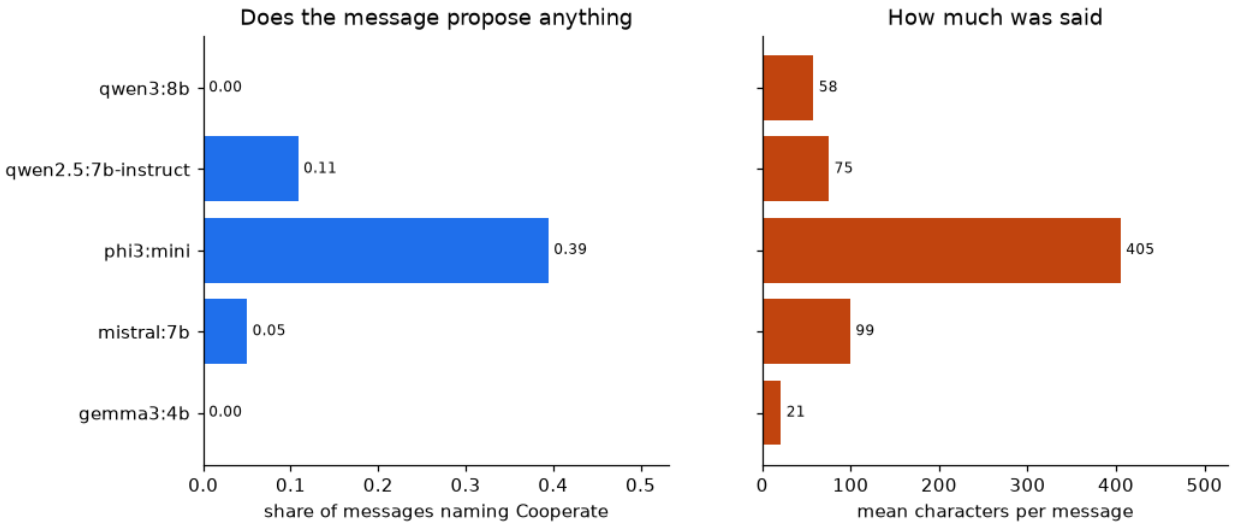

FIGURE 4 : Le canal portait-il seulement une proposition, depuis une ouverture de défection mutuelle.

TABLE 6 : Le canal depuis une ouverture de défection mutuelle, par contenu lexical.

modèle	messages	part nommant Cooperate	part nommant Defect	caractères en moyenne	tours où les deux sièges sont identiques
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

Les deux modèles qu'un message ne libère pas sont ceux dont les messages **ne nomment jamais d'action**, et celui qu'il libère est celui qui propose explicitement la coopération. Les comptes seuls seraient une base fragile : les messages ont donc aussi été lus. Ils diffèrent en nature, pas seulement en fréquence :

qwen2.5:7b-instruct, libéré. *« Let's both try to cooperate in this round and see how it goes. »* Il propose la coopération et il coopère.

gemma3:4b, inchangé. *« Let's see what happens. »* Puis : *« Interesting. »* Puis : *« Keep going. »* Les messages font 21 caractères en moyenne et ne portent aucun contenu. Le canal n'est pas ignoré : il est inutilisé.

qwen3:8b, inchangé. *« Let's keep the pressure on and make every round count. »* Fluide, chaleureux, coopératif de registre, et entièrement dépourvu de proposition. Il fait défection sur les trente tours en les envoyant.

L'échec n'est donc pas qu'une proposition soit faite puis écartée. **Pour ces deux modèles, aucune proposition n'est faite**, et la phrase qui la porterait est soit vide, soit seulement coopérative de ton. Le troisième cas est le plus intéressant : un modèle peut produire un texte qui se lit comme un discours de coopération sans que son comportement en soit affecté, ce qui est exactement le mode d'échec qui rend une communication non contraignante difficile à utiliser comme signal.

4.6 La raison annoncée et le coup joué

TABLE 7 : Le prompt demande une raison. qwen3:8b nomme une action et la joue dans 99,7 % des cas ; mistral:7b contredit son propre raisonnement dans 33 % des tours où il en énonce un.

modèle	tours nommant une action	tours concordants	taux de concordance
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

Lu à côté de la Section 4.5, c'est plus tranchant qu'il n'y paraît. Le modèle dont la raison et le coup s'accordent le mieux est aussi celui dont les *messages* sont décorrélés des deux : il annonce ce qu'il va faire et le fait, tout en envoyant à l'autre siège quelque chose de chaleureux et sans rapport. La cohérence interne entre délibération et action n'implique pas que le signal émis en porte quoi que ce soit.

4.7 Ce qu'un modèle paie pour ne pas être refusé

Le jeu itéré n'est pas le seul que le cadre du stage définit. Le jeu du dictateur supprime le refus : rien de stratégique ne subsiste et ce qu'un allocateur donne est une disposition. Le proposeur de l'ultimatum affronte les mêmes cent points avec un refus possible. **L'écart entre les deux offres est ce qu'un modèle donne pour ne pas être refusé**, et aucun des deux chiffres ne dit grand-chose seul. Le minimum du répondant est demandé avant qu'aucune proposition ne lui soit montrée, il ne peut donc pas être un accommodement à l'une d'elles.

TABLE 8 : Points sur 100 cédés à l'autre joueur, quatre décisions par cellule.

modèle	offre au dictateur	offre à l'ultimatum	payé pour éviter le refus	minimum accepté	refuserait sa propre offre
gemma3:4b	50	50	+0	51	oui
mistral:7b	74	72	-1	47	non
phi3:mini	62	68	+6	52	non
qwen2.5:7b-instruct	50	50	+0	51	oui

modèle	offre au dictateur	offre à l'ultimatum	payé pour éviter le refus	minimum accepté	refuserait sa propre offre
qwen3:8b	50	99	+49	50	non

Deux choses en ressortent.

Un modèle achète l'acceptation à presque n'importe quel prix. qwen3 donne exactement la moitié quand l'autre joueur ne peut pas refuser, et **99 sur 100 quand il le peut**, soit une prime de 49 points contre un risque de refus qu'une offre de 60 aurait écarté tout aussi bien. Sa raison ne se trompe pourtant pas sur le jeu : *« I want to ensure the other player accepts to avoid getting nothing. Offering 99 gives them a strong incentive to accept. »*

Deux modèles refuseraient leur propre proposition. gemma3 et qwen2.5 offrent tous deux exactement 50 comme proposeur et annoncent tous deux un minimum de 51 : chacun refuserait donc le partage qu'il venait de qualifier d'équitable. La raison de qwen2.5 montre le glissement : *« MINIMUM : 51. To ensure a fair split as I would not accept less than half. »* Cinquante, c'est la moitié. Le nombre choisi pour faire respecter « pas moins de la moitié » la dépasse d'un point, et il ne s'applique jamais ce critère à lui-même quand il propose.

L'observation qui réunit les deux, et la raison de la présence de ce jeu dans l'article. Le comportement de qwen3 semble ici l'inverse de celui de la Section 4.4, où il fait défection depuis un départ neutre silencieux quand tous les autres coopèrent. C'est la même règle. Dans le dilemme du prisonnier il écrivait que faire défection *« guarantees at least 1 point compared to potentially 0 if you cooperate »* ; dans l'ultimatum il offre 99 *« to avoid getting nothing »*. Les deux évitent le pire cas d'une table de gains. Cette logique recommande la défection dans un jeu et une générosité quasi totale dans l'autre : un modèle qui paraît impitoyable ici et absurdement généreux là peut n'avoir qu'une seule disposition, et non deux.

C'est une mise en garde contre la lecture d'un seul jeu comme une personnalité, et c'est exactement le sens dans lequel Horton, Filippas, et Manning (2023) dit qu'il faut lire un désaccord.

4.8 La délibération change-t-elle quelque chose

La grille exécute tous les modèles délibération coupée, car l'activer coûte à qwen3 34 secondes par appel contre 1,9 et aurait ajouté des jours. Le cas le plus difficile du panel restait donc non testé. `run_contrasts.py` l'active dans la seule cellule de défection imposée, sur les répétitions 0 et 1, soit un ordre de gains chacune.

TABLE 9 : qwen3 dans la cellule de défection imposée, délibération coupée puis active.

condition	matchs, coupée	coopération, coupée	matchs, active	coopération, active	écart
without cheap talk	4	0,00	2	0,00	+0,00
with cheap talk	4	0,00	2	0,09	+0,09

La délibération le fait essayer, et seulement là où il a de quoi essayer. En silence il reste exactement à 0,00 dans les deux répétitions, le même verrou que la grille. Avec un canal il atteint 0,09, ce qui n'est pas une sortie mais n'est pas rien : la paire sonde, et les sondages se voient dans le jeu.

Les raisons disent ce que font les actions. Il ouvre sur le même argument de dominance qu'auparavant, et au dernier tour il énonce explicitement ce qu'il tente : *« 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]. »*

C'est le signe opposé à Liu et al. (2026), pour qui supprimer la chaîne de pensée *réduit* souvent un effondrement de la coopération, la délibération l'amplifiant donc. Ici la délibération est la seule chose qui ait produit la

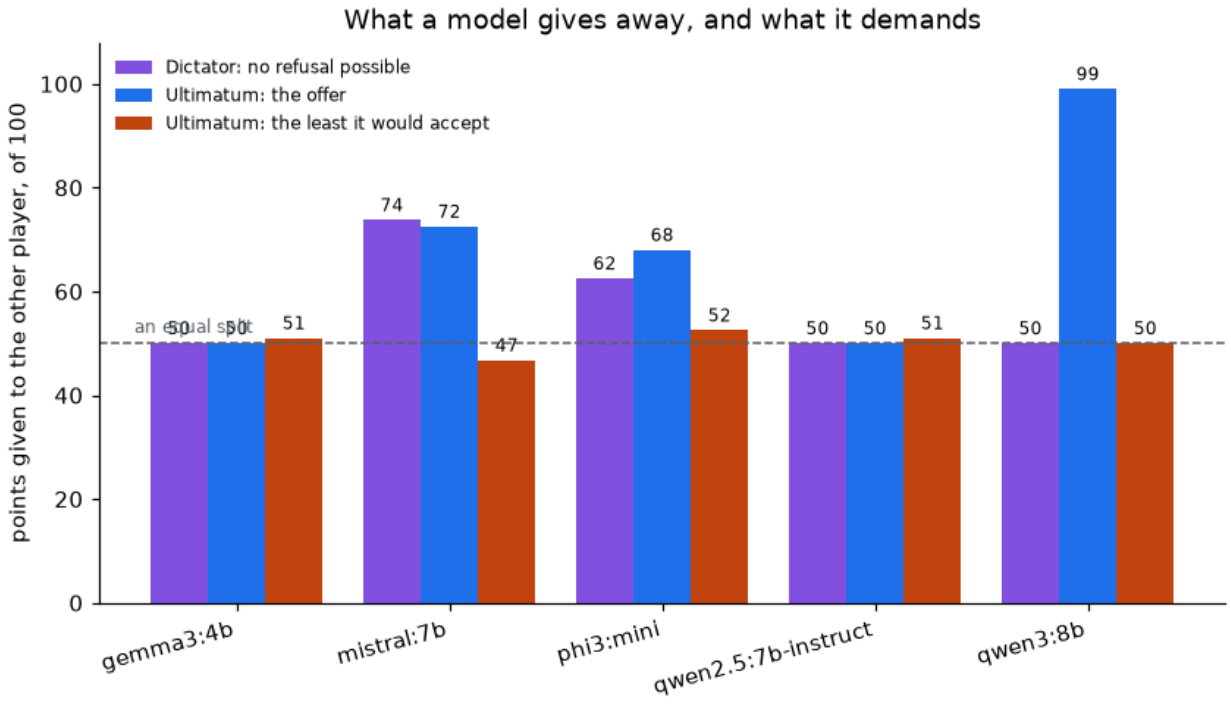

FIGURE 5 : Ce qu'un modèle donne quand le refus est impossible, quand il est possible, et le minimum qu'il dit accepter.

moindre coopération dans une cellule où quatre matchs sans elle n'en produisaient aucune. Deux matchs par condition sont une base faible pour contredire une étude de 500 tours sur sept modèles, et le désaccord est rapporté comme un désaccord et non comme une réfutation : ce qui est solide, c'est que cette cellule a quitté le zéro, et seulement avec un canal.

Une note d'instrumentation qui fait partie du résultat. Une première tentative de ce volet a perdu ses deux matchs sur une réponse vide, la délibération ayant consommé tout le budget de tokens sans rien laisser pour la ligne d'action. Les trajectoires réelles provoquent environ 5 120 caractères de raisonnement par appel, à peu près le double de ce que provoque un historique synthétique de tours identiques, ce qui explique qu'une sonde à la même profondeur ait laissé croire le budget suffisant. Le doubler a réparé le volet, et un match perdu conserve désormais ses dernières réponses, pour que le prochain échec de ce genre soit diagnosticable plutôt que seulement compté.

4.9 Les deux moitiés sur un seul tableau

TABLE 10 : Score par tour contre les adversaires que les deux moitiés ont joués. Les matchs de l'imitateur font 100 tours et ceux des modèles 30 : les scores par tour sont donc comparables, les taux de coopération pas tout à fait.

joueur	tours	Tit For Tat	Grudger	Win-Stay Lose-Shift	Defector	Alternator
Mirror	100	2,99	0,98	3,00	0,96	1,68
Neuron						
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

joueur	tours	Tit For Tat	Grudger	Win-Stay Lose-Shift	Defector	Alternator
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

L'imitateur termine dernier sur huit au classement général, et ce tableau dit d'où cela vient, ce que le classement ne peut pas faire. **Il n'est pas battu partout.** Contre le Tit-for-Tat il prend 2,99 par tour, à égalité avec les deux meilleurs modèles et devant les trois autres, et contre le Win-Stay Lose-Shift il prend le maximum, 3,00. Deux adversaires expliquent son classement : le Grudger, qui ne pardonne jamais, si bien qu'une seule provocation d'un mécanisme qui ne peut pas s'empêcher de copier parfois une défection lui coûte le reste du match ; et l'Alternator, qu'il ne peut pas suivre, un appariement de fréquences n'ayant aucune représentation de l'alternance.

Il se débrouille aussi mieux face à un défecteur pur que deux des cinq modèles de langage, avec 2,3 fois le score de mistral et de phi3 sur cette cellule. Un mécanisme incapable de représenter la réciprocité cesse malgré tout d'alimenter un défecteur, parce que l'appariement de fréquences converge vers la fréquence qu'il observe, là où deux modèles continuent de coopérer avec lui plus d'une fois sur deux.

« L'imitateur est moins bon que les modèles de langage » n'est donc pas ce que le rapprochement montre. Il est moins bon en agrégat et meilleur que plusieurs d'entre eux sur des adversaires précis, et ces adversaires précis sont ceux sur lesquels portait l'hypothèse du stage.

5 Limites

Quatre matchs par cellule, et une variance quasi nulle par construction. À l'intérieur d'une classe de parité de l'ordre des gains, l'étape initiale a produit des scores identiques au bit près pour deux graines différentes. Un test de significativité sur ce protocole serait un théâtre : ce qui varie entre répétitions est surtout le contrebalancement, non un bruit d'échantillonnage. Les résultats sont rapportés comme des issues de cellule, et les plus forts le sont parce qu'ils sont unanimes (4 sur 4, taux exactement 0 ou exactement 1), non parce qu'un test a été passé.

L'auto-affrontement est une disposition, pas deux sujets. Les deux sièges portent le même prompt et le même modèle. Le tableau des messages enregistre la fréquence à laquelle les deux sièges ont émis la chaîne identique, qui pour un modèle est la majorité des tours. La température est à 0,7 plutôt qu'à 0 précisément parce qu'un décodage glouton rend l'auto-affrontement dégénéré, mais les deux sièges restent loin d'être des agents indépendants.

La mesure lexicale du contenu des messages est un indicateur indirect. Nommer « cooperate » n'est pas proposer la coopération, et une paraphrase échappe au test. Les extraits verbatim de la Section 4.5 sont donnés pour cette raison : les comptes seuls ne suffiraient pas à soutenir l'affirmation. Une analyse de contenu codée, par un second modèle ou par un humain, serait l'instrument approprié.

Un modèle est rapporté comme illisible et non comme un résultat.

phi3:mini a perdu 10 de ses 44 matchs sur des réponses ne nommant aucune action, à une moyenne de 16,2 tours. C'est aussi le seul modèle du panel en Q4_0 plutôt qu'en Q4_K_M, l'un des deux builds les plus anciens, et le plus petit. Ce protocole ne sépare pas ces explications, et ses cellules ne sont donc pas interprétées.

Ce facteur confondant est testable, et il a été testé. `run_contrasts.py` rejoue tout le stage de phi3 sur le build `Q4_K_M` des mêmes poids 3,8B mini-4k instruct, en conservant la suite de graines de l'original pour que la quantification soit la seule chose qui change, et en écrivant dans un journal séparé pour que la grille déclarée reste à 220 matchs.

TABLE 11 : Le contrôle de quantification, sur les cellules qu'il a atteintes, face aux mêmes cellules du stage d'origine.

volet	build de contrôle	matchs	perdus	taux de perte	matchs d'origine	perdus	taux de perte
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
qwen3-think-roomy	`qwen3:8b`	4	0	0,00	4	1	0,25

La quantification en explique une bonne part, et pas la totalité. À poids, longueur de contexte, graines et prompts constants, en ne changeant que le format 4 bits, le taux de perte passe de 0,23 à 0,09. L'empaquetage `Q4_0`, plus grossier, compte donc réellement dans l'incapacité de phi3 à tenir un format de réponse, ce qu'il vaut mieux savoir avant de lire un résultat de petit modèle sur une étiquette par défaut.

Ce n'est pas non plus toute l'histoire. 0,09 reste le seul taux de perte non nul du panel : les quatre autres modèles n'ont rien perdu. Un résidu revient donc au modèle, et les cellules de phi3 restent non interprétées plutôt que réhabilitées par un meilleur build.

La troisième ligne est une erreur qu'il vaut la peine de garder. Ce volet avait été écrit comme le contrôle de quantification et n'en était pas un : `phi3:mini` est le build 128k, celui tiré face à lui était en 4k, et la grille demande à chaque modèle une fenêtre de 8192 tokens. Faire tourner un modèle de 4096 tokens au double de sa fenêtre d'entraînement coûte 0,75 de ses matchs, contre 0,23 pour les mêmes poids à leur propre longueur de contexte. Cela ne dit rien de la quantification et beaucoup d'un décalage de fenêtre, et c'est rapporté pour ce que cela mesure plutôt que supprimé.

Ce volet est rapporté ici plutôt qu'intégré à l'un des tableaux ci-dessus, parce qu'il s'agit d'une exécution de suivi répondant à un facteur confondant, et non d'une partie de la grille déclarée.

L'échelle. Tous les modèles sont entre 3,8 et 8,2 milliards de paramètres et quantifiés en 4 bits sur une carte grand public. Là où ces résultats diffèrent de travaux portant sur des modèles de pointe, la taille et la quantification restent des explications vivantes que ce protocole ne peut pas écarter.

Une étiquette n'est pas une version. Les empreintes exactes des modèles, les graines et le matériel sont consignés à côté des données, car une réexécution qui résout une empreinte différente est une autre expérience.

6 Conclusion

Le stage demandait si les algorithmes entretiennent la coopération tacite, la rompent ou l'intensifient. Sur ces données, le mécanisme d'imitation l'**entretient**, au sens fort qu'il ne peut rien faire d'autre : deux imitateurs conservent le régime où leur condition initiale les place, et la règle ne peut pas représenter la réciprocité qu'on attendait d'elle.

Les modèles de langage reproduisent ce cliquet sans en hériter la nécessité. Trois des quatre modèles lisibles conservent un régime défectif imposé à chaque tour de chaque match, un en sort sans aucun canal, et un

message non contraignant en sauve un des trois et pas les deux autres.

Trois résultats supplémentaires disent pourquoi, et ils font la substance de l'article.

La décision se prend une fois, sur les indices disponibles. Presque tous les matchs se stabilisent au premier tour que la paire contrôle : le résultat porte donc sur un seul choix. Croiser les ouvertures et le canal sépare ce dont un modèle dispose pour raisonner, et qwen3 tranche le mécanisme en se comportant de trois façons : il fait défection sur la seule matrice des gains, coopère quand un message est le seul signal, et refait défection quand un historique fabriqué et un message sont tous deux présents, sa raison citant l'historique et jamais le message. **Un historique planté prime sur un signal non contraignant émis au même instant**, et lequel des deux l'emporte est une propriété du modèle.

Là où le canal échoue, le plus souvent il n'est pas utilisé. Les modèles qu'un message ne libère pas sont ceux dont les messages ne proposent jamais rien : des fragments phatiques de vingt et un caractères dans un cas, et dans l'autre une prose collaborative et fluide envoyée tout en faisant défection sur les trente tours. C'est ce second cas qui inquiète, car il est indiscernable en surface d'un discours de coopération.

La délibération le déplace, dans le sens que la littérature ne prédit pas. Délibération activée dans la cellule qui le piège, qwen3 passe d'un zéro exact à 0,09 et se met à sonder, en disant explicitement qu'il tente de briser le cycle. Il ne le fait que là où un canal existe ; en silence il reste exactement verrouillé. Deux matchs par condition en font un désaccord avec Liu et al. (2026) plutôt qu'une réfutation.

Pour la question de politique publique qui a motivé le stage, la forme utile de ce résultat est négative et étroite. **Un canal de communication n'est pas un remède contre un régime collusif, et son absence n'en est pas une garantie.** Savoir si un agent donné peut être ramené par la parole hors d'un régime où on l'a placé est une propriété de cet agent, à mesurer pour lui plutôt qu'à déduire de la classe à laquelle il appartient.

7 Reproduire ce travail

Chaque chiffre et chaque figure ci-dessus dérivent de `llm/results/matches.jsonl`, qui est versionné et jamais régénéré, par une arithmétique que l'intégration continue rejoue à chaque poussée et compare octet par octet. Le schéma du journal, les tables dérivées, les empreintes des modèles et les conditions d'exécution sont documentés dans `llm/results/README.md`. Ce document lit ces mêmes fichiers CSV, et les deux versions linguistiques partagent le module `analysis.py` : un chiffre du texte ne peut donc pas contredire le tableau dont il vient, ni l'autre version.

Références

Askenazi-Golan, Galit, Domenico Mergoni Cecchelli, Edward Plumb, et 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, et 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, et 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, et 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., et Joel Sobel. 1982. « Strategic Information Transmission ». *Econometrica* 50 (6) : 1431-51.

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

Eschenbaum, Nicolas, Filip Mellgren, et Philipp Zahn. 2022. « Robust Algorithmic Collusion ». https://arxiv.org/abs/2201.00345.
Farrell, Joseph, et Matthew Rabin. 1996. « Cheap Talk ». *Journal of Economic Perspectives* 10 (3) : 103-18.
Fish, Sara, Yannai A. Gonczarowski, et Ran I. Shorrer. 2024. « Algorithmic Collusion by Large Language Models ». https://arxiv.org/abs/2404.00806.
Fontana, Nicoló, Francesco Pierri, et 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, et Mirco Musolesi. 2025. « Opponent Shaping in LLM Agents ». https://arxiv.org/abs/2510.08255.
Horton, John J., Apostolos Filippas, et 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, et 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, et Vincent Conitzer. 2026. « The Memory Curse : How Expanded Recall Erodes Cooperative Intent in LLM Agents ». https://arxiv.org/abs/2605.08060.
Lore, Nunzio, et Babak Heydari. 2026. « Communication Enhances LLMs' Stability in Strategic Thinking ». https://arxiv.org/abs/2602.06081.
Madmoun, Hachem, et 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, et 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, et Baptiste Alloui-Cros. 2025. « Strategic Intelligence in Large Language Models : Evidence from Evolutionary Game Theory ». https://arxiv.org/abs/2507.02618.
Vidler, Alicia, et 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, et Michael Luck. 2025. « Will Systems of LLM Agents Cooperate : An Investigation into a Social Dilemma ». https://arxiv.org/abs/2501.16173.