

IA, cognition humaine et effondrement du savoir*

Daron Acemoglu[†]

Dingwen Kong[‡]

Asuman Ozdaglar[§] 20

février 2026

Résumé

Nous étudions comment l'IA générative, et en particulier l'IA agentique, façonne les incitations à l'apprentissage humain et l'évolution à long terme de l'écosystème informationnel de la société. Nous construisons un modèle dynamique d'apprentissage et de prise de décision dans lequel les décisions fructueuses nécessitent de combiner des connaissances générales partagées au niveau de la communauté avec des connaissances spécifiques au contexte au niveau individuel ; ces deux apports sont complémentaires. L'apprentissage présente *des économies de gamme* : l'effort humain coûteux produit à la fois un signal privé concernant son propre contexte et un signal public « faible » qui s'accumule dans le stock de connaissances générales de la communauté, générant ainsi une externalité d'apprentissage. L'IA agentique fournit des recommandations spécifiques au contexte qui *se substituent* à l'effort humain. En revanche, un stock plus riche de connaissances générales *complète* l'effort humain en augmentant son rendement marginal.

Le modèle met en évidence une tension dynamique marquée : si l'IA agentique peut améliorer la qualité des décisions immédiates, elle peut également éroder les incitations à l'apprentissage qui sous-tendent la connaissance collective à long terme. Lorsque l'effort humain est suffisamment élastique et que les recommandations agentiques dépassent un seuil de précision, l'économie peut basculer vers un état d'équilibre caractérisé par *un effondrement de la connaissance*, dans lequel la connaissance générale finit par disparaître, malgré des conseils personnalisés de grande qualité. Le bien-être est généralement non monotone par rapport à la précision de l'IA agentique, ce qui implique l'existence d'un niveau intérieur de précision de l'IA agentique maximisant le bien-être et justifie la mise en place de réglementations en matière de conception de l'information. En revanche, une plus grande capacité d'agrégation des connaissances générales — c'est-à-dire un partage et une mise en commun plus efficaces des connaissances générales générées par les humains — augmente sans ambiguïté le bien-être et renforce la résilience face à l'effondrement des connaissances.

Mots-clés : intelligence artificielle, cognition, savoir collectif, communauté, apprentissage social.

Classification JEL : D80, D83.

* Nous remercions la Fondation Hewlett, la Fondation Stone et le Consortium Gen-AI du MIT pour leur soutien financier.

[†] Massachusetts Institute of Technology, NBER et CEPR, daron@mit.edu

[‡] Massachusetts Institute of Technology, dingwenk@mit.edu

[§] Massachusetts Institute of Technology, asuman@mit.edu

1 Introduction

Les progrès rapides de l'IA (généraliste) depuis 2022 ont fait naître l'espoir de bénéfices généralisés grâce à la fourniture d'informations et à la prestation de services de qualité supérieure par l'IA (Chui et al., 2023 ; Maslej et al., 2025). Les modèles d'IA sont désormais capables d'analyser et de rédiger des textes sur un éventail impressionnant de sujets, à l'instar d'experts humains spécialisés, et peuvent passer au crible de vastes ensembles de données pour reconnaître des schémas et localiser des informations pertinentes et utiles (OpenAI, 2023). Plus important encore, ils peuvent utiliser ces capacités pour fournir des informations et des recommandations adaptées au contexte. Néanmoins, l'utilisation croissante des outils d'IA généraliste a également suscité des craintes, par exemple quant à l'impact que le recours à l'IA pourrait avoir sur la cognition, les connaissances et la créativité humaines (Holmes et al., 2023).

Les avis divergent sur cette question. D'un côté, on trouve des commentateurs et des experts qui estiment que l'utilisation des outils d'IA permettra d'élargir considérablement les connaissances humaines, par exemple sous la forme de découvertes scientifiques plus rapides et de meilleure qualité, ainsi que d'une prise de décision nettement améliorée chez les professionnels ayant accès à l'IA. Dans cet esprit, Demis Hassabis a décrit l'IA comme « l'outil ultime pour aider les scientifiques, pour nous aider à explorer l'univers qui nous entoure ».¹ À l'autre extrémité du spectre se trouvent ceux qui s'inquiètent des effets d'une dépendance croissante à l'IA pour le travail en classe sur la capacité des enfants à « apprendre à apprendre », ou de l'automatisation des tâches sur l'acquisition par les travailleurs novices de compétences pertinentes spécifiques à un métier ou à une tâche.²

Les désaccords portent en partie sur la question de savoir si les informations fournies par l'IA constituent un *complément* ou un *substitut* à l'apprentissage humain. Dans le premier cas, le développement de l'IA permettra aux humains de concentrer leurs efforts et leur attention sur ce qui compte vraiment et d'utiliser les contributions de l'IA avec une efficacité croissante. Dans le cas d'un substitut, cependant, une IA de plus en plus performante découragera de plus en plus l'effort et l'apprentissage humains, car la plupart des informations pertinentes seront servies aux humains sur un plateau.

Les partisans de ces deux points de vue peuvent trouver des arguments pour étayer leur position. Le rôle essentiel joué par l'IA pour aider les scientifiques à concevoir de nouvelles protéines constitue un exemple bien connu de la manière dont ces outils ont déjà fait évoluer les choses dans l'avancement des connaissances scientifiques humaines (Jumper et al., 2021 ; Watson et al., 2023). Ce n'est pas seulement dans le domaine scientifique que l'IA se révèle capable de compléter la prise de décision humaine. À partir des données d'une entreprise leader du secteur, Brynjolfsson et al. (2025) ont montré comment l'IA a aidé avec succès les agents du service client en leur fournissant des informations contextuelles pertinentes sur le problème auquel ils sont confrontés. À l'inverse, plusieurs études montrent que l'utilisation de l'IA a un impact négatif sur les efforts d'apprentissage des humains, voire parfois sur leurs capacités. Kosmyna et al. (2025) constatent que l'utilisation de ChatGPT comme aide à la rédaction réduit la capacité des utilisateurs à mémoriser et à consigner avec précision leurs propres arguments, et, fait remarquable, cela semble se produire en modifiant la connectivité neuronale dans le cerveau des utilisateurs. Les auteurs indiquent également que les résultats obtenus sont désormais beaucoup plus similaires d'un auteur à l'autre, ce qui soulève des inquiétudes quant à l'érosion de la créativité et de l'autonomie individuelles. Dans le même ordre d'idées, Gerlich (2025) constate que les grands modèles linguistiques semblent réduire la créativité des utilisateurs, en particulier chez les plus jeunes, et Budiyo et al. (2025) rapportent

¹ Voir <https://www.nobelprize.org/prizes/chemistry/2024/hassabis/interview/>

² Voir, par exemple, <https://www.edweek.org/technology/rising-use-of-ai-in-schools-comes-with-big-downsides-for-students/2025/10> et <https://www.weforum.org/stories/2025/04/ai-jobs-international-workers-day/>

des résultats similaires dans des évaluations à plus long terme, tandis que Jakesch et al. (2023) rapportent que la collaboration avec ChatGPT pour la rédaction influence considérablement les opinions et les croyances des personnes. Dans le même ordre d'idées, plusieurs études ont montré que les recommandations de l'IA sont souvent mal interprétées par les experts, par exemple en radiologie (Agarwal et al., 2023).

Ces différences peuvent être une conséquence inévitable du fait que les outils d'IA générative seront utilisés dans de nombreux domaines, avec des effets potentiellement différents. Mais elles peuvent également refléter le fait que certains des « effets d'équilibre » de l'IA générative ne peuvent être appréhendés à partir d'une approche empirique au cas par cas sans une structure théorique appropriée. Par exemple, même dans le cas des substituts, la prise de décision individuelle pourrait tout de même s'améliorer grâce à l'utilisation de l'IA, car la principale raison pour laquelle les décideurs humains pourraient fournir moins d'efforts est qu'ils peuvent désormais prendre de bonnes décisions sans en payer le prix. Ainsi, dans les exemples mentionnés au paragraphe précédent, chaque décideur individuel pourrait améliorer ses performances dans certaines tâches bien définies, même si cela se fait au détriment de l'acquisition de connaissances générales utiles par ces utilisateurs et, plus important encore, au détriment de l'appauvrissement des connaissances collectives. En effet, il existe des preuves que les grands modèles linguistiques ont déjà eu un impact négatif sur la construction collective des connaissances dans plusieurs contextes. Par exemple, dans le contexte de la plateforme de questions-réponses sur le codage Stack Overflow, on observe désormais une baisse de l'activité, un moindre engagement humain face aux questions et, sans doute, une moindre production de nouvelles connaissances, alors que les codeurs expérimentés sont confrontés à de nouvelles questions. Les données indiquent clairement que cela s'explique par le fait que les gens se tournent vers des outils d'IA générative pour obtenir cette même aide (del Rio-Chanona et al., 2024). La tendance semble similaire pour Wikipédia, avec une baisse de la lecture et de la création d'articles dans les domaines où ChatGPT constitue un substitut efficace à Wikipédia (Lyu et al., 2025).

Cet article vise à contribuer à une meilleure compréhension théorique de l'impact des outils d'IA sur la cognition et les connaissances humaines. Nous élaborons un modèle dynamique d'apprentissage et de prise de décision dans lequel les apports de l'IA peuvent être soit complémentaires, soit substituables à l'effort humain. Au cœur de notre approche se trouve une distinction entre deux types d'informations : les informations générales et celles spécifiques à un individu (ou à un contexte). Pour accomplir une tâche quelconque, les individus ont besoin de connaissances générales. Par exemple, pour prendre des décisions d'investissement, il faut une compréhension de base des différents instruments financiers tels que les obligations d'État, les obligations d'entreprise, les actions, les options, etc., ainsi que des informations sur la performance des marchés boursiers et des économies mondiales, certains aspects pertinents de leur structure institutionnelle, une compréhension des risques macroéconomiques, etc. Mais il faut également des informations liées au *contexte* de l'individu : quelle est la tolérance au risque et l'horizon de planification de l'individu en question ? Quelle est la corrélation entre ses autres sources de revenus et les rendements des différents actifs ? Dispose-t-il d'informations, d'intuitions, de préférences ou de convictions qui influencent la manière dont il devrait investir et les types de risques qu'il devrait prendre ? Et ainsi de suite. Il convient de noter que les décideurs humains acquièrent souvent des connaissances à la fois générales et spécifiques. Par exemple, la plupart des individus acquerront des connaissances financières générales dans le cadre d'un cours de finance ou en lisant de la littérature financière pertinente, et ils en viendront à reconnaître leurs propres besoins et à se forger des préférences et des convictions pertinentes pour l'investissement au cours du même processus. En d'autres termes, il existe souvent *des économies de gamme* dans l'apprentissage, les mêmes efforts générant à la fois *des connaissances générales* et *des connaissances spécifiques à l'individu ou au contexte*.

Tout comme les humains, les outils d'IA générative peuvent également acquérir des connaissances à la fois générales et spécifiques à un contexte. Mais c'est leur capacité – en particulier la promesse des *modèles d'IA agentique* tant attendus – à développer et à fournir des informations spécifiques au contexte ainsi qu'une aide à la décision qui est la plus prometteuse. S'appuyant sur une architecture qui permet le stockage et l'analyse rapide de vastes quantités de données, ainsi que l'inférence sur les liens entre le contexte et les informations pertinentes, ces modèles promettent de mettre en évidence des schémas pertinents pour un contexte spécifique et particulièrement utiles pour les décideurs individuels. En effet, à un certain niveau, Internet a déjà agrégé une quantité énorme de connaissances générales qui existaient déjà, et les outils d'IA générative actuels peuvent le faire de manière tout à fait satisfaisante dans de nombreux domaines. La prochaine grande étape — qui recèle à la fois un immense potentiel et des dangers — est l'aide individuelle et contextuelle fournie par les modèles d'IA. C'est sur ce type de recommandation contextuelle issue de l'IA agentique que nous nous concentrons dans notre analyse. Pour revenir à l'exemple de l'investissement, les manuels et les ressources en ligne peuvent enseigner les mécanismes d'un large éventail d'instruments financiers, mais un futur système agentique pourrait traduire le contexte particulier d'un individu en un choix concret de portefeuille, voire exécuter de manière autonome des transactions en fonction des conditions et des contraintes de cette personne.

L'un des principes fondamentaux qui sous-tendent notre approche est que, pour bien prédire, obtenir de bons résultats ou accomplir des tâches, il faut généralement disposer à la fois de connaissances générales et de connaissances spécifiques au contexte, et que ces deux types d'informations se *complètent* pour aboutir à des décisions fructueuses. Les connaissances générales permettent d'interpréter et de valoriser les données spécifiques au contexte ; à l'inverse, les connaissances spécifiques au contexte permettent de situer précisément le décideur au sein de ce cadre général. Une autre caractéristique importante des connaissances générales est qu'elles s'appuient sur les efforts d'apprentissage de toute une communauté : par nature, les connaissances générales peuvent être partagées. Cela implique que la majeure partie des connaissances générales générées par un individu est une externalité qu'il ou elle n'intériorise pas. Ces éléments impliquent que des connaissances générales supplémentaires augmentent le rendement marginal de l'effort d'apprentissage d'un individu : avec davantage de connaissances générales, un même effort est plus bénéfique pour comprendre le contexte spécifique d'un individu et agir de manière utile dans ce contexte, ce qui motive à fournir davantage d'efforts lorsque les connaissances générales sont plus abondantes. Ainsi, les connaissances générales sont *complémentaires* à l'effort d'apprentissage humain. De même, avec les recommandations spécifiques au contexte fournies par l'IA agentique, il y aura moins d'incitation à fournir des efforts coûteux, car l'un des objectifs de cet effort est déjà bien rempli par l'IA agentique. Par conséquent, les recommandations de l'IA agentique se *substituent* à l'effort humain.

Notre modèle intègre cette relation entre les différents types de connaissances et l'effort d'apprentissage individuel dans un modèle dynamique d'apprentissage communautaire. Un autre élément important réside dans le fait que les connaissances au niveau de la communauté constituent elles-mêmes une donnée d'entrée pour les modèles d'IA : sans les efforts humains, les expériences et les découvertes, il n'y aurait pas suffisamment d'informations utiles que les modèles d'IA pourraient agréger et analyser pour extraire des connaissances générales ou formuler des recommandations personnalisées.³

Nous modélisons le problème de décision d'un ensemble d'agents humains comme un problème de prédiction : le gain de chaque agent dépend de la distance par rapport à un état commun représentant les connaissances générales et de leur

³ Nous aborderons plus loin la possibilité de « données synthétiques », grâce auxquelles les modèles d'IA génèrent leurs propres données et leur propre apprentissage, sans avoir besoin d'efforts humains, de connaissances ou de génération de données.

une prédiction concernant cet état, ainsi que sur l'écart entre leur prédiction concernant l'état contextuel et sa valeur réelle. Pour formuler ces prédictions, les agents s'appuient sur leur propre effort d'apprentissage, qui génère deux signaux : l'un corrélé à leur état contextuel et l'autre corrélé à l'état commun. Du fait qu'il concerne l'état commun, ce dernier signal est utile à tous les décideurs, générant ainsi des externalités d'apprentissage ou de données. L'IA agentique fournit également des recommandations contextuelles, là encore modélisées sous forme de prédictions, que les agents combinent de manière optimale avec leurs propres signaux et a priori.

La conclusion principale de cet article est une mise en garde : un modèle puissant d'IA agentique peut aider *de manière statique* les décideurs humains, mais il peut nuire *de manière dynamique* à la construction collective de connaissances. En effet, il peut conduire à ce que nous appelons un « effondrement des connaissances », dans lequel, à long terme, l'équilibre aboutit à la destruction finale de toute connaissance humaine.

Comprendre l'intuition qui sous-tend ce résultat permet de clarifier notre principale contribution et l'interprétation de nos résultats. Il n'est pas surprenant que, d'un point de vue statique, les individus tirent profit d'informations supplémentaires, en l'absence de toute spécification erronée dans leur modèle du monde ou d'autres biais. S'ils réduisent leurs efforts, c'est uniquement parce qu'ils reçoivent déjà une recommandation assez fiable de la part d'une IA agentique, qui se substitue naturellement à leurs efforts. Cette substitution, lorsqu'elle est optimale, ne peut nuire à leur utilité d'un point de vue statique. Cependant, l'effort humain alimente la connaissance collective, et cette externalité n'est pas internalisée par les agents. À mesure qu'ils réduisent leur effort d'apprentissage, la quantité d'informations que la communauté ou les modèles d'IA peuvent agréger commence à diminuer. Les effets sur le bien-être à long terme dépendent alors de l'équilibre entre ces deux forces : une meilleure prise de décision statique contre une connaissance collective moins dynamique. Nous montrons que l'effet négatif est susceptible de dominer lorsque l'IA agentique est très précise. En revanche, la capacité des agents à apprendre à partir des connaissances générales des autres — que ce soit via l'agrégation communautaire ou des outils plus traditionnels de type IA ou Internet — améliore toujours le bien-être.

Ces mêmes analyses statiques comparatives s'appliquent également à la probabilité que le système converge vers l'état d'équilibre caractérisé par un effondrement des connaissances. Plus précisément, dans les domaines où l'effort d'apprentissage est suffisamment élastique, le système peut présenter plusieurs états d'équilibre : un état à haut niveau de connaissances et un piège d'effondrement des connaissances avec un niveau de connaissances générales nul. À mesure que l'IA agentique s'améliore, le bassin d'attraction de l'état d'équilibre caractérisé par un effondrement des connaissances s'étend. De plus, dès lors que l'IA agentique est suffisamment précise, un « effondrement complet » se produit, dans lequel l'état d'équilibre à haut niveau de connaissance disparaît et le système converge vers le piège de l'effondrement des connaissances, quelles que soient les conditions initiales. D'un autre côté, une meilleure agrégation des connaissances générales produites par l'homme a l'effet inverse : elle réduit le bassin d'attraction de l'état d'équilibre d'effondrement des connaissances et peut compenser en partie l'effet de découragement de l'IA agentique.

Étant donné que les recommandations agentiques fonctionnent uniquement par le biais de la fourniture d'informations, ces résultats incitent naturellement à mettre en place des politiques de régulation fondées sur la conception de l'information. Nous montrons que limiter la précision effective des recommandations agentiques par un brouillage délibéré peut préserver les incitations à l'apprentissage et empêcher l'effondrement des connaissances. Plus précisément, nous caractérisons une politique de brouillage optimale en deux phases qui maximise le bien-être à long terme.

Nous examinons également plusieurs extensions afin de démontrer à la fois la flexibilité du modèle et la robustesse de nos principaux résultats. Tout d’abord, dans notre modèle de référence, nous partons du principe que, même en l’absence d’IA, les connaissances générales générées par les humains sont relativement bien agrégées et que, par conséquent, l’IA n’apporte pas d’amélioration significative à l’agrégation de ces connaissances. Dans notre première extension, nous assouplissons cette hypothèse et établissons que des résultats similaires s’appliquent même lorsque l’IA améliore simultanément l’agrégation des connaissances générales par rapport à ce que la communauté était capable de faire et introduit l’élément agentique consistant à fournir des recommandations spécifiques à chaque décideur humain. Nos résultats s’appliquent facilement à ce cas. Deuxièmement, nous examinons dans quelle mesure les « données synthétiques », générées par des modèles d’IA sans s’appuyer sur l’apprentissage et l’expérimentation humains, peuvent se substituer à l’effort humain. Nous montrons que les mêmes observations qualitatives s’appliquent même avec des données synthétiques, à condition que ces données ne constituent pas un substitut parfait à l’effort humain. Cependant, dans ce cas, l’état d’équilibre de « collapse des connaissances » comporte toujours une quantité positive d’informations générales sur l’état commun, car même sans effort humain, les données synthétiques génèrent de nouvelles connaissances. Troisièmement, nous étudions une version du modèle dans laquelle les individus peuvent décider de l’orientation de leur effort d’apprentissage, en déterminant l’équilibre entre l’acquisition de connaissances générales et de connaissances spécifiques au contexte. À condition que les connaissances générales et celles spécifiques au contexte ne puissent pas être parfaitement séparées, nos résultats continuent de s’appliquer dans ce cas.

La suite de cet article est organisée comme suit. Dans la section suivante, nous proposons un bref aperçu de la littérature sur le sujet. La section 3 présente notre modèle de base et contient notre analyse principale ainsi que nos résultats les plus importants. La section 4 caractérise les effets de l’IA sur le bien-être. La section 5 présente les extensions évoquées plus haut, tandis que la section 6 conclut l’article. L’annexe contient les démonstrations des résultats exposés dans le texte ainsi que d’autres extensions.

2 Bibliographie

Notre principale contribution consiste à fournir, à notre connaissance, la première analyse théorique des dynamiques d’apprentissage collectif en présence d’agents IA et à établir la possibilité que l’IA, même si elle facilite la prise de décision statique, puisse nuire à la construction collective de connaissances à long terme. Il existe néanmoins plusieurs travaux liés à notre recherche.

Les travaux les plus proches de notre article sont ceux qui étudient l’impact de l’IA sur l’acquisition d’informations par l’humain. Ide (2025) développe un modèle dans lequel les entreprises peuvent décider d’automatiser les tâches de base, mais cela peut nuire à la croissance à long terme car cela entrave la transmission intergénérationnelle des connaissances tacites, qui s’effectue généralement au sein des entreprises par le biais d’interactions entre novices et experts. Agarwal et al. (2025) examinent la collaboration entre l’humain et l’IA dans les tâches de classification et soulignent en quoi l’apport de l’IA se substitue à l’effort humain. Ils discutent de la manière dont des schémas de collaboration optimaux peuvent être élaborés via une approche de conception de mécanismes. Ni ces deux articles, ni aucun autre dont nous ayons connaissance, n’examinent les effets de l’IA sur l’acquisition collective de connaissances lorsque l’apprentissage individuel dépend des efforts d’autrui. Ils ne mettent pas non plus en avant l’élément clé de notre approche, qui repose sur la possibilité que l’apport de l’IA puisse être soit un complément, soit un substitut aux efforts d’apprentissage humains.

On peut également citer quelques travaux qui traitent de l'impact des technologies numériques récentes sur l'information ou les comportements collectifs, bien que dans des contextes très différents. Acemoglu et al. (2024) ainsi que Dasaratha et He (2022) analysent les implications de l'utilisation des réseaux sociaux sur l'apprentissage. Agrawal et al. (2023), Ide et Talama (2025) et Cullen et al. (2025) étudient les effets de l'IA sur le partage des connaissances, la structure organisationnelle et la productivité dans les entreprises. Les travaux récents de Farboodi et al. (2025) étendent le cadre d'automatisation basé sur les tâches d'Acemoglu et Restrepo (2018, 2019) de sorte que les données générées à la fois par l'IA et par les travailleurs humains deviennent une contribution à une automatisation plus poussée basée sur l'IA. Hadfield et Koh (2025) proposent une discussion générale sur les effets économiques des agents IA.

L'accent mis sur la nature collective de l'apprentissage humain remonte aux premiers travaux classiques de Nelson (1959) et d'Arrow (1962a,b). Ces idées sont également à l'origine de la littérature récente sur l'expérimentation collective, par exemple Bolton et Harris (1999), Keller et al. (2005) et Keller et Rady (2010), ainsi que de la littérature sur l'apprentissage social (à partir des actions d'autrui), notamment Banerjee (1992), Bikhchandani et al. (1992) et Smith et Sørensen (2000). Des contributions plus récentes à cette littérature, telles que celles d'Acemoglu et al. (2011) et de Dasaratha et He (2019), soulignent également comment la structure d'observation détermine l'apprentissage individuel et collectif. Aucun de ces articles n'aborde le rôle des agrégateurs externes, tels que l'IA, dans la construction collective des connaissances.

Nous nous inscrivons également dans la littérature consacrée aux conditions dans lesquelles les signaux sont complémentaires ou substituables, ainsi qu'aux externalités de données plus générales qui en découlent. Voir, entre autres, Bořrgers et al. (2013) et Brooks et al. (2024) sur les comparaisons de signaux, ainsi que Acemoglu et al. (2022) et Bergemann et al. (2022) sur les externalités liées aux données et à l'information.

Enfin, une littérature émergente en informatique étudie les effets individuels et sociaux des outils d'IA, et certains des articles pertinents de cette littérature ont déjà été cités dans l'introduction.

3 Modèle

Dans cette section, nous présentons notre modèle, fournissons la caractérisation des équilibres, établissons comment un état d'équilibre de « collapse de connaissances » peut émerger à la suite d'une IA agentique, et présentons plusieurs analyses statiques comparatives.

3.1 Environnement

Le temps est discret, $t = 1, 2, \dots$. À chaque période t , il existe un continuum d'agents à courte durée de vie de masse totale $M > 0$, indexés par $i \in [0, M]$. La population est divisée en plusieurs îlots de telle sorte que *les connaissances générales* ne sont agrégées et partagées qu'au sein de chaque îlot. Plus précisément, on divise $[0, M]$ en îlots de masse égale I : pour $m = 0, 1, \dots, K - 1$ avec $K := M/I$, l'îlot I_m est

$$I_m := [mI, (m + 1)I).$$

Pour chaque îlot I_m , les connaissances générales acquises par les agents au sein de cet îlot sont agrégées et partagées avec toutes les cohortes futures au sein de cette île. Ici, I représente la capacité des agents à apprendre à partir des connaissances générales des autres — que ce soit via l'agrégation communautaire, l'IA plus traditionnelle ou des outils de type Internet. En maintenant la population totale M fixe, l'augmentation *de* I entraînera une diminution du nombre d'îles et une augmentation de leur taille, permettant ainsi une meilleure agrégation des connaissances générales. Lorsque $I = M$, l'économie ne comporte qu'une seule île et toutes les connaissances générales sont parfaitement partagées par l'ensemble de la population, tandis qu'avec *un* petit I , le partage des connaissances générales est beaucoup moins important.

Il existe deux états latents du monde, représentant deux types de connaissance de la société :

- Un état commun ϑ_t évolue comme une marche aléatoire :

$$\vartheta_{t+1} = \vartheta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma^2), \quad \vartheta_1 \sim N(0, \Sigma_0^2).$$

Nous interprétons cet état commun comme un savoir général qui présente une utilité générale pour l'ensemble de la communauté. Par exemple, dans le cadre de décisions d'investissement, il peut refléter une compréhension de base des instruments financiers et des institutions de marché, tandis que dans le cadre de décisions médicales, il peut représenter des connaissances fondamentales sur l'anatomie, les bactéries, les virus, les maladies et les traitements standard. Nous modélisons l'état commun comme *évoluant* au fil du temps, car la frontière des connaissances générales n'est pas statique : l'innovation technologique peut introduire de nouveaux instruments et pratiques financiers, et les changements environnementaux ou épidémiologiques peuvent entraîner l'apparition de nouvelles maladies et thérapies. La spécification de marche aléatoire offre un moyen simple de rendre compte de cette évolution continue.

- Un état idiosyncratique pour chaque agent i , tiré indépendamment à chaque période,

$$\vartheta_{i,t} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

Nous interprétons $\vartheta_{i,t}$ comme une connaissance contextuelle propre à l'agent i à la période t . Par exemple, dans le cadre de décisions d'investissement, elle peut refléter la tolérance au risque de l'individu, son horizon de planification et son exposition au risque lié aux revenus non financiers ; dans le cadre de décisions médicales, elle peut représenter des caractéristiques propres au patient, telles que ses symptômes et son état physique. Comme ces informations sont intrinsèquement individualisées et (dans cette formulation stylisée) ne persistent pas de manière systématique d'une période à l'autre, nous modélisons $\vartheta_{i,t}$ comme un tirage indépendant issu d'une distribution fixe.

Pour l'agent i à la période t , nous modélisons sa tâche comme consistant à prédire à la fois ϑ_t (l'état commun à cette période) et $\vartheta_{i,t}$ (son état idiosyncratique). Notons ces prédictions respectivement $x_{i,t}$ et $y_{i,t}$. La production dépend des deux erreurs de prédiction $|x_{i,t} - \vartheta_t|$ et $|y_{i,t} - \vartheta_{i,t}|$. Par souci de simplicité, nous supposons que la production dépend uniquement du fait que chaque prédiction se situe ou non dans une bande de tolérance d'une unité :

$$f(1/x_{i,t} - \vartheta_t \leq 1, 1/y_{i,t} - \vartheta_{i,t} \leq 1),$$

où $f: \{0, 1\}^2 \rightarrow \mathbb{R}$ est une fonction faiblement croissante, normalisée de sorte que $f(1, 1) - f(0, 0) = 1$. Le fait que le gain dépende du fait que les erreurs de prédiction soient inférieures à 1 est une simplification.

La reparamétrisation suivante permet d'isoler les effets des deux composantes — les connaissances générales et celles spécifiques au contexte — dans la production :

$$\begin{aligned}\Delta_G &:= f(1, 0) - f(0, 0), \\ \Delta_I &:= f(0, 1) - f(0, 0), \\ \Delta_X &:= f(1, 1) - f(1, 0) - f(0, 1) + f(0, 0).\end{aligned}$$

Ces objets représentent, respectivement, les gains issus de la connaissance générale, de la connaissance contextuelle, et le gain supplémentaire résultant de leur complémentarité. La normalisation $f(1, 1) - f(0, 0) = 1$ implique

$$\Delta_G + \Delta_I + \Delta_X = 1.$$

Tout au long de l'analyse, nous posons l'hypothèse suivante sur la fonction de production :

Hypothèse 1. $\Delta_I = 0$ et $\Delta_X > 0$.

Ici, $\Delta_I = 0$ rend compte des environnements où la connaissance spécifique au contexte, à elle seule, sans connaissance générale, ne crée aucune valeur ajoutée. Par exemple, un médecin qui observe les symptômes d'un patient peut ne pas être en mesure de prescrire un traitement efficace s'il ne comprend pas correctement les mécanismes de la maladie. En d'autres termes, $\Delta_I = 0$ rend compte d'une forte complémentarité dans la production. En effet, puisque $\Delta_I + \Delta_X = f(1, 1) - f(1, 0) \geq 0$ (par monotonie de f), il s'ensuit que $\Delta_X \geq 0$ dès lors que

$\Delta_I = 0$. Afin d'exclure le cas trivial $\Delta_I = \Delta_X = 0$ (dans lequel les agents n'auraient aucune incitation à apprendre), nous imposons en outre $\Delta_X > 0$. De plus, nous admettons que $\Delta_G > 0$: la connaissance générale seule peut parfois créer de la valeur ; par exemple, un médecin disposant de connaissances médicales de base correctes peut être en mesure de prodiguer des soins rudimentaires même sans examen approfondi du patient.

Un exemple classique satisfaisant l'hypothèse 1 est la fonction de production de Leontief

$$f(x, y) = \min\{x, y\},$$

pour laquelle

$$\Delta_I = \Delta_G = 0, \Delta_X = 1.$$

Avant d'exécuter les tâches, chaque agent choisit un effort $e_{i,t} \geq 0$ au coût $\frac{1}{\alpha} e_{i,t}^\alpha$, où $\alpha > 1$. Le est donc égal au rendement de la tâche moins le coût de l'effort :

$$u_{i,t} = f(\{x_{i,t} - \vartheta_t \leq 1\}, \{y_{i,t} - \vartheta_{i,t} \leq 1\}) - \frac{1}{\alpha} e_{i,t}^\alpha.$$

Nous décrivons tout d'abord l'environnement informationnel dans un monde *dépourvu* d'IA agentique. Comme indiqué ci-dessus, l'effort d'apprentissage d'un agent génère deux types de signaux :

- **Apprentissage privé.** L'effort de l'agent i génère un signal privé concernant son état idiosyncratique $\vartheta_{i,t}$, dont la précision est proportionnelle à son effort :

$$s_{i,t}^H \sim N \left(\vartheta_{i,t}, \frac{1}{\lambda_I e_{i,t}} \right),$$

où $\lambda_I > 0$ est un paramètre technologique.

- **Apprentissage public.** L'effort d'un agent génère également des connaissances générales, c'est-à-dire des informations sur l'état commun ϑ_t . Dans le modèle de référence, nous supposons que ces connaissances générales sont déjà bien agrégées au sein des îlots. Nous simplifions également l'analyse en supposant que chaque agent est atomistique, de sorte que les contributions individuelles au stock (public) de connaissances générales sont infinitésimales, mais que l'agrégation de nombreuses contributions de ce type reste finie. Plus précisément, l'effort de l'agent i produit un signal concernant ϑ_t dont la précision est proportionnelle à $\lambda_G e_{i,t}$, où $\lambda_G > 0$ est un paramètre technologique. Les signaux produits au sein de l'îlot m sont agrégés — par l'observation communautaire, Internet ou une IA plus traditionnelle — en un signal social public

$$s_{m,t}^C \sim N \left(\vartheta_t, \frac{1}{\lambda_G E_{m,t}} \right), \quad (1)$$

où $E_{m,t}$ est l'effort global sur l'île m :

$$E_{m,t} := \int_{I_m} e_{i,t} di. \quad (2)$$

Ce signal public $s_{m,t}^C$ est partagé avec toutes les cohortes futures de l'île m . La spécification implique une informativité constante par unité d'effort global.

Comme la contribution de chaque individu est infinitésimale, la composante d'apprentissage public de l'effort ne génère pas d'incitations privées, de sorte que l'externalité d'apprentissage n'est absolument pas internalisée par les agents. Par exemple, lorsqu'un ingénieur logiciel diagnostique un bug de production rare, le principal bénéfice privé provient de la restauration de son propre système (une correction très spécifique au contexte), tandis que la transformation de cette découverte en un article public clair et mis à jour (par exemple sur Stack Overflow) profiterait à de nombreux futurs développeurs, mais ne rapporte qu'un faible bénéfice privé.

Soit $m(i)$ l'îlot de l'agent i . Dans le monde pré-IA, chaque agent observe son signal privé concernant l'état idiosyncratique ainsi que l'historique des signaux publics précédemment agrégés sur son îlot. Formellement, l'ensemble d'informations pré-IA de l'agent i et l'historique public sur l'îlot m jusqu'au temps t sont

$$\eta_i^{\text{pré-IA}} := \left(s_{i,t}^H, \{s_{m(i),t}^C\}_{1 \leq t \leq t} \right), \quad \eta_t^{\text{pré-IA}} := \{s_{m,t}^C\}_{1 \leq t \leq t}.$$

Ainsi, même si l'état commun sous-jacent ϑ_t est partagé entre les îlots, les agents situés sur différents condition des îlots selon différentes histoires publiques.

Nous présentons à présent l'IA agentique. Comme évoqué dans l'introduction, nous modélisons l'IA agentique comme une

technologie de l'information qui fournit des informations individuelles et spécifiques au contexte. Par exemple, un tel système agentique pourrait fournir des recommandations médicales personnalisées ou des suggestions d'investissement adaptées à l'utilisateur. En conséquence, nous modélisons l'IA agentique comme fournissant à chaque agent un signal concernant son état idiosyncrasique :

$$s_{i,t}^A \sim N(\vartheta_{i,t}, \frac{1}{\tau_A}),$$

où $\tau_A \geq 0$ rend compte de l'exactitude (précision) de l'IA agentique.

Dans le modèle de référence, l'IA agentique ne fournit pas directement d'informations sur l'état commun et n'améliore pas non plus la technologie permettant de rassembler les efforts humains dispersés au sein du savoir public. Cette hypothèse correspond à une approximation d'un monde dans lequel le savoir général est déjà bien agrégé grâce à l'apprentissage social en présentiel et à d'autres outils antérieurs à l'IA. Par conséquent, l'IA agentique n'améliore pas le savoir général disponible pour la prise de décision. Dans l'exemple médical, par exemple, cette interprétation correspondrait au cas où l'IA agentique n'inventerait pas de nouveaux traitements ni ne découvrirait de nouvelles maladies, et n'informerait pas les spécialistes sur des connaissances générales importantes. Dans la section 5.1, nous admettons que les améliorations des capacités de l'IA puissent également accroître la capacité d'agrégation effective de la société, tandis que dans la section 5.2, nous examinons les données synthétiques générées par l'IA. À condition que ces canaux supplémentaires ne soient pas trop puissants, les forces fondamentales mises en avant dans le modèle de référence continuent de régir la dynamique à long terme, comme nous le verrons.

D'un point de vue formel, les ensembles d'informations individuelles et publiques post-IA sont

$$I_{i,t}^{\text{post-IA}} := \{s_{i,t}^H, s_{i,t}^A, \{s_{m(i),t}^C\}_{1 \leq t \leq t}\}, \quad \text{et} \quad I_{m,t}^{\text{post-IA}} := \{s_{m,t}^C\}_{1 \leq t \leq t}.$$

Il convient de noter que les informations publiques disponibles sur chaque île ne sont pas modifiées par l'introduction de l'IA agentique dans le modèle de référence.

Le calendrier intra-périodique est le suivant : au cours de la période t ,

- (i) Chaque agent i observe l'historique public $\{s_{m(i),t}^C\}_{1 \leq t \leq t}$ sur son îlot et choisit $e_{i,t}$. (Étant donné que les profils d'effort des cohortes passées sont déterminés de manière unique à l'équilibre, que l'agent actuel observe les efforts passés n'a aucune importance.)
- (ii) Chaque agent i observe le signal privé $s_{i,t}^H$.
- (iii) (*Post-IA uniquement*) Chaque agent i reçoit une recommandation de l'IA agentique $s_{i,t}^A$.
- (iv) Chaque agent i formule une prédiction $(x_{i,t}, y_{i,t})$ et reçoit une utilité égale au produit obtenu moins le coût de l'effort.
- (v) La communauté agrège les signaux de connaissance générale au sein de chaque île m et produit le signal public $s_{m,t}^C$, qui devient accessible à tous les membres de l'îlot m de la cohorte t .

Nous allons montrer que l'équilibre dynamique unique est symétrique entre les îles et les agents. Dans cet équilibre symétrique, l'effort $e_{i,t}$ est identique pour tous les individus et toutes les îles (qui sont de taille égale

), de sorte que $E_{m,t} = E_t$ et que la précision induite dans (1) est la même pour tous *les* m . Par conséquent, à partir de ce point, nous nous concentrerons, sans perte de généralité, sur une île représentative (disons $m = 0$) et laisserons de côté l'indice de l'îlot. On écrit donc E_t pour $E_{m=0,t}$, s_t^C pour $s_{t,m=0,t}^C$, et I_t pour $I_{m=0,t}$. Nous nous concentrons également au comportement d'un seul agent, qui représente, par symétrie, le comportement de tous les autres agents sur la même île.

Il est également utile de noter que l'environnement pré-IA équivaut à un environnement post-IA où la précision de l'IA agentique $\tau_A = 0$. Par conséquent, dans les sections suivantes, nous nous concentrerons sur l'environnement post-IA, nous traiterons le monde pré-IA comme un cas particulier où $\tau_A = 0$, et nous omettrons l'exposant « post-IA » pour les ensembles d'informations.

3.2 Mises à jour des croyances et connaissances

Définissons

$$V_t := \text{Var}(\vartheta_t \mid I_t)$$

comme la variance a posteriori de l'état commun ϑ_t compte tenu de l'historique public jusqu'au temps t . La *précision publique* correspondante

$$X_t := V_t^{-1}$$

résume le stock de connaissances générales (accessible à tous les individus de la société). Les signaux étant gaussiens, la trajectoire $\{X_t\}_{t \geq 1}$ ne dépend pas des réalisations des signaux publics, mais est plutôt déterminée par la trajectoire des efforts agrégés $\{E_t\}_{t \geq 1}$.

Le filtrage de Kalman standard implique la loi de mouvement suivante pour la précision du signal public, qui mesure le stock de connaissances générales. À la fin de la période t , l'effort agrégé E_t génère le signal public $s_t^C \sim N(\vartheta_t, (\lambda_G E_t)^{-1})$. En combinant ce signal avec la variance a priori V_t donne la variance a posteriori $V_t^{-1} + \lambda_G E_t$ pour ϑ_t , et à cela s'ajoute la dérive d'état $\vartheta_{t+1} = \vartheta_t + \epsilon_t$ ajoute variance Σ^2 . Par conséquent,

$$V_{t+1} = \text{Var}(\vartheta_{t+1} \mid I_{t+1}) = V_t^{-1} + \lambda_G E_t^{-1} + \Sigma^2.$$

L'effort global E_t de cette cohorte augmente la précision publique V_t^{-1} de la cohorte suivante¹; la variance de l'innovation Σ^2 la dilue d'une période à l'autre.

De manière équivalente, la précision publique évolue selon

$$V_{t+1}^{-1} = X_t + \lambda_G E_t^{-1} + \Sigma^2. \quad (3)$$

Ensuite, définissons la *précision idiosyncratique* de l'agent i comme suit

$$Y_{i,t} := \text{Var}(\vartheta_{i,t} \mid I_{i,t})^{-1} = \sigma^{-2} + \lambda_I e_{i,t} + \tau_A, \quad (4)$$

qui est la somme (i) de la précision a priori σ^{-2} , (ii) de la précision issue de l'apprentissage privé $\lambda_i e_{i,t}$, et (iii) de la précision issue de la recommandation de l'IA agentique τ_A .

3.3 Définition de l'équilibre

Un équilibre (bayésien parfait) est constitué des choix d'effort $\{e_{i,t}\}$ et des prévisions $\{(x_{i,t}, y_{i,t})\}$ de tous les agents, de telle sorte que, pour chaque agent i et chaque période t , les valeurs choisies $e_{i,t}$ et $(x_{i,t}, y_{i,t})$ maximisent l'utilité attendue $E[u_{i,t}]$, où

$$u_{i,t} = f(1\{x_{i,t} - \vartheta_t \leq 1\}, 1\{y_{i,t} - \vartheta_t \leq 1\}) - \frac{1}{\alpha} e_{i,t}^\alpha.$$

Avant de caractériser l'équilibre dynamique, nous nous intéressons de plus près au comportement d'un agent donné au cours d'une période donnée. Étant donné que les agents ont une durée de vie limitée et que f est faiblement croissante pour chacun de ses arguments, les agents, en situation d'équilibre, effectuent des prédictions *bayésiennes* à un pas, en choisissant les moyennes a posteriori suivantes :

$$x_{i,t} = E[\vartheta_t | I_{i,t}], \quad y_{i,t} = E[\vartheta_t | I_{i,t}]. \quad (5)$$

Compte tenu de (5), l'utilité attendue est déterminée par les croyances a posteriori de l'agent, résumées par la précision publique X_t et la précision idiosyncratique $Y_{i,t}$.

Pour une variable aléatoire gaussienne de moyenne nulle $Z \sim N(0, \tau^{-1})$ de précision $\tau > 0$, définissons

$$G(\tau) := \Pr(|Z| \leq 1) = 2\Phi\left(\frac{1}{\sqrt{\tau}}\right) - 1, \quad g(\tau) := G'(\tau) = \frac{\phi\left(\frac{1}{\sqrt{\tau}}\right)}{\sqrt{\tau}}.$$

D'après (5), on a $x_{i,t} - \vartheta_t \sim N(0, \text{Var}(\vartheta_t | I_t)) = N(0, X_t^{-1})$, ce qui implique $P(x_{i,t} - \vartheta_t \leq 1) = G(X_t)$. De même, $P(y_{i,t} - \vartheta_t \leq 1) = G(Y_{i,t})$. Par conséquent, en utilisant le fait que $\Delta_i = 0$ (la connaissance individuelle ne génère aucune valeur sans la connaissance générale), l'utilité attendue de l'agent i à la période t peut s'écrire s'écrit comme suit :

$$U_{i,t} := E u_{i,t} = f(0, 0) + \underbrace{\frac{G(X_t) \cdot \Delta_G}{\alpha}}_{\text{Général Knowledge}} + \underbrace{\frac{G(X_t)G(Y_{i,t}) \cdot \Delta_X}{\alpha}}_{\text{Complémentarité Context}} - \frac{1}{\alpha} e_{i,t}^\alpha. \quad (6)$$

3.4 Substituts et compléments

Comme indiqué dans l'introduction, l'un des thèmes centraux de cet article est que les informations fournies par l'IA générative peuvent soit *compléter* l'apprentissage humain (en valorisant l'effort humain), soit s'y *substituer* (en fournissant directement les informations que cet effort aurait produites), selon que les informations fournies sont générales ou spécifiques au contexte. Dans notre modèle, ces deux types d'informations correspondent à deux objets distincts : le stock hérité de précision publique X_t , qui mesure les connaissances générales, et la précision de la recommandation contextuelle et agentique τ_A . L'espérance

Le résultat de (6) met en lumière ces deux forces.

Remarque 1. La précision publique X_t complète l'effort humain, tandis que la précision de l'IA agentique τ_A se substitue à l'effort humain. C'est-à-dire :

$$\frac{\partial^2 U_{i,t}}{\partial e_{i,t} \partial X_t} > 0, \quad \frac{\partial^2 U_{i,t}}{\partial e_{i,t} \partial \tau_A} < 0.$$

Nous expliquons ici tout d'abord pourquoi les connaissances générales, représentées par la précision publique X_t , constituent un complément à l'effort humain. Sous l'hypothèse 1 ($\Delta_I = 0$ et $\Delta_X > 0$), le rendement privé d'un agent résultant de son effort provient entièrement du terme de complémentarité

$$G(X_t)G(Y_{i,t}) \cdot \Delta_X, \quad \text{où } Y_{i,t} = \sigma^{-2} + \lambda_I e_{i,t} + \tau_A.$$

L'effort d'apprentissage fait augmenter $Y_{i,t}$, et son bénéfice marginal est proportionnel à $G(X_t)$: lorsque le stock de connaissances générales est plus élevé, l'amélioration de la précision contextuelle revêt une plus grande valeur, car ces deux intrants sont complémentaires dans la réalisation des tâches, et l'obtention d'une meilleure précision contextuelle est la principale raison pour laquelle les individus déploient un effort d'apprentissage (les connaissances générales qu'ils génèrent constituent une externalité qu'ils n'internalisent pas). Formellement,

$$\frac{\partial U_{i,t}}{\partial e_{i,t}} = \Delta_X G(X_t) \lambda_I g(Y_{i,t}) - e_{i,t}^{\alpha-1} \Rightarrow \frac{\partial^2 U_{i,t}}{\partial e_{i,t} \partial X_t} = \Delta_X \lambda_I g(X_t) g(Y_{i,t}) > 0.$$

Cela signifie que l'effort d'apprentissage individuel est plus bénéfique lorsqu'il existe davantage de connaissances générales (X_t plus élevé) pouvant être appliquées de manière plus productive à son propre contexte.

D'autre part, la précision de l'IA agentique n'intervient que par le biais de $Y_{i,t}$, améliorant directement la précision spécifique au contexte. Comme la probabilité de succès $G(Y)$ présente des rendements décroissants en termes de précision, une valeur plus élevée de τ_A réduit le gain marginal résultant d'un effort humain supplémentaire :

$$\frac{\partial^2 U_{i,t}}{\partial e_{i,t} \partial \tau_A} = \Delta_X G(X_t) \lambda_I g'(Y_{i,t}) < 0.$$

Ainsi, de meilleures recommandations des agents évincent l'effort d'apprentissage en rendant la précision acquise de manière incrémentale moins précieuse à la marge.

Comme X_t est déterminé par les efforts d'équilibre des cohortes passées (et par la capacité d'agrégation de l'île I), l'observation 1 implique que :

1. Un effort passé plus élevé $\{e_t\}_{t < t}$ renforce le signal public et encourage la poursuite des efforts.
2. Une plus grande capacité d'agrégation (un I plus grand) renforce le signal public, encourageant ainsi également l'effort.
3. Une plus grande précision de l'IA agentique τ_A affaiblit le rendement marginal de l'effort, ce qui décourage l'effort.

Dans l'ensemble, les choix en matière d'effort constituent *des compléments stratégiques* entre les cohortes, et une plus grande capacité d'agrégation renforce cette complémentarité. En revanche, l'IA agentique se *substitue* à l'effort d'apprentissage humain.

3.5 Existence et caractérisation de l'équilibre

Fixons une période t et un agent i . En prenant pour données la précision publique actuelle X_t (déterminée par les efforts agrégés des cohortes précédentes sur l'île) et la précision de l'IA agentique τ_A , l'agent i choisit un effort $e_{i,t} \geq 0$ afin de maximiser son utilité attendue $U_{i,t}$ dans (6). En utilisant (4), la condition du premier ordre est

$$\frac{\partial U_{i,t}}{\partial e_{i,t}} = \Delta G_X(X) \cdot \lambda g(\sigma_I^{-2} + \lambda e_{i,t} + \tau_A) - e_{i,t}^{\alpha-1}.$$

Le terme du bénéfice marginal est faiblement décroissant en $e_{i,t}$ car $g(\cdot)$ est décroissant, tandis que le

$e_{i,t}^{\alpha-1}$ est strictement croissant (puisque $\alpha > 1$). Par conséquent, $\frac{\partial U_{i,t}}{\partial e_{i,t}}$ est strictement décroissante en $e_{i,t}$.

De plus, on a

$$\frac{\partial U_{i,t}}{\partial e_{i,t}} \Big|_{e_{i,t}=0} = \Delta G_X(X) \cdot \lambda g(\sigma_I^{-2} + \tau_A) \geq 0, \quad \lim_{e_{i,t} \rightarrow +\infty} \frac{\partial U_{i,t}}{\partial e_{i,t}} \leq 0,$$

donc $U_{i,t}$ est strictement concave en effort et admet un maximiseur unique et fini.⁴

Ce maximiseur est commun à tous les agents de la cohorte t : $e_{i,t} = e_t$ pour tout i . Définissons alors *l'effort de meilleure réponse* $e(X, \tau_A)$ comme la solution unique à la condition du premier ordre :

$$e(X, \tau_A) := e \geq 0 \mid \Delta_X G(X) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) = e^{\alpha-1},$$

de sorte qu'en équilibre, $e_{i,t} = e_t = e(X_t, \tau_A)$.

L'observation 1, combinée à des arguments standard de statique comparative monotone (par exemple, le théorème de Topkis), donne :

Observation 2. *L'effort de meilleure réponse $e(X, \tau_A)$ est croissant en X et décroissant en τ_A , et ce de manière stricte pour tout $X > 0$.*

Compte tenu du choix symétrique $e_t = e(X_t, \tau_A)$, l'effort global sur l'île est

$$E_t = \int_I e_{i/t} dt = I e_t = I e(X_t, \tau_A),$$

et la récursion de Kalman (3) détermine de manière unique la précision publique de la période suivante. Nous résumons la dynamique d'équilibre comme suit. La preuve de tous les résultats énoncés dans le texte est présentée dans le

⁴ À la limite $X_t = 0$, on a $G(X_t) = 0$; l'effort n'apporte donc aucun bénéfice et la meilleure réponse unique est $e_{i,t} = 0$. En effet, un individu ne peut à lui seul accroître le stock de connaissances générales et, sans connaissances générales, les connaissances spécifiques au contexte ne sont d'aucune utilité. Au contraire, à partir de n'importe quelle valeur $X_1 > 0$, la récursivité de l'équilibre garantit que $X_t > 0$ pour tout t fini, et l'effort de meilleure réponse est intérieur à chaque période le long de la trajectoire d'équilibre.

Annexe.

Proposition 1 (Équilibre dynamique). *Fixons une précision publique initiale $X_1 \geq 0$ implicite dans la distribution a priori. Il existe un et unique chemin d'équilibre symétrique (bayésien parfait) tel que $e_{i,t} = e_t \geq 0$, où $\{X_t, e_t\}_{t=1}^{\infty}$ est la solution unique de la récursion*

$$\begin{aligned} & e_t = e(X_t, \tau_A), \\ \forall t, \quad X_{t+1} &= \frac{h}{\Sigma^2} + X_t + \lambda_G I e_t^{-1} i^{-1}. \end{aligned} \quad (7)$$

En substituant la meilleure réponse dans la loi du mouvement de (7), on obtient une carte de transition d'état unidimensionnelle (à un pas) :

$$X_{t+1} = F(X_t),$$

où

$$F(X) := \frac{h}{\Sigma^2} + X + \lambda_G I e(X, \tau_A)^{-1} i^{-1}. \quad (8)$$

L'équilibre dynamique se résume donc à une itération de la fonction F sur la variable d'état X_t , l'effort e_t s'ajustant de manière endogène par le biais de la meilleure réponse $e(X_t, \tau_A)$. Cette représentation met en évidence deux forces qui façonnent l'évolution des connaissances générales.

Tout d'abord, il y a l'accumulation de connaissances : un effort actuel plus élevé e_t augmente la précision du signal public généré à la fin de la période t , ce qui accroît le stock de précision publique de la période suivante. Ce phénomène est rendu par le terme $\lambda_G I e_t$ à l'intérieur de F , qui représente à la fois la productivité de l'apprentissage public (λ_G) et la capacité d'agrégation de l'île (I).

Deuxièmement, il y a la dépréciation des connaissances : comme l'état commun suit un parcours aléatoire, ce qui a été appris par le passé perd de sa pertinence, et l'ampleur de cet effet est rendue par la variance d'innovation Σ^2 . Par exemple, même si l'état actuel était connu avec une précision arbitrairement élevée, la variance de la période suivante ne serait pas inférieure à Σ^2 . Cela se reflète dans le fait que la précision publique ne peut jamais dépasser Σ^{-2} et que, par conséquent, F s'applique à un intervalle borné :

$$0 \leq F(X) < \Sigma^{-2} \quad \text{pour tout } X \text{ fini.}$$

3.6 Équilibres en régime permanent

Un équilibre en régime permanent peut être représenté par un scalaire X^- satisfaisant la condition de point fixe de la fonction de transition d'état F :

$$X^- = F(X^-).$$

En d'autres termes, un équilibre en régime permanent peut être représenté par un niveau de précision publique qui se reproduit d'une cohorte à l'autre.

L'effort en régime permanent associé e^- et la précision idiosyncratique Y^- sont définis par

$$e^- := e(X^-, \tau_A), \quad Y^- := \sigma^{-2} + \lambda_I e^- + \tau_A.$$

Pour notre propos, un type important d'état stationnaire est l'état dégénéré où

$$(X^-, e^-, Y^-) = (0, 0, \sigma^{-2} + \tau_A). \quad (9)$$

Cet état d'équilibre correspond à une situation dans laquelle la communauté n'accumule pratiquement aucune connaissance générale utilisable à long terme. Même si l'IA agentique peut encore fournir des informations pertinentes spécifiques à un individu ou à un contexte, cela n'est pas socialement utile car, au fil du temps, les connaissances générales qui constituent une donnée essentielle (d'après notre hypothèse selon laquelle $\Delta_I = 0$) disparaissent et rendent également inutiles les informations spécifiques au contexte. C'est la raison pour laquelle nous appelons l'état d'équilibre représenté en (9) « état d'équilibre d'effondrement des connaissances ».

Nous dérivons ensuite un certain nombre de propriétés fondamentales de la carte de transition d'états F , ce qui nous permet de caractériser l'ensemble des états stationnaires et la dynamique à long terme.

Lemme 1. *La fonction de transition d'états F est continue et strictement croissante sur $\mathbb{R}_{\geq 0}$, avec*

$$F(0) = 0 \quad \text{et} \quad F(+\infty) = \Sigma^{-2}.$$

Ici, la monotonie de F découle de la complémentarité fondamentale entre les connaissances générales et les rendements de l'effort d'apprentissage : un stock hérité plus élevé de connaissances générales X augmente l'effort actuel $e(X, \tau_A)$, ce qui renforce le signal public et accroît la précision publique de la période suivante.

La fonction F dépend de l'échelle d'agrégation I et de la précision de l'IA agentique τ_A par le biais de l'effort agrégé endogène $I \cdot e(X, \tau_A)$. L'observation 2 nous permet d'établir une statique comparative monotone de F .

Proposition 2 (Statique comparative de F). *La fonction F est croissante point par point en I et décroissante point par point en τ_A , et ce de manière stricte pour tout $X > 0$.*

La proposition 2 a une interprétation économique directe. Une échelle d'agrégation I plus grande rend l'effort de chaque cohorte plus productif, car le même effort individuel est regroupé en un signal public plus informatif, ce qui se traduit par une connaissance générale plus importante au cours de la période suivante. En revanche, une précision τ_A plus élevée de l'IA agentique réduit l'effort actuel (puisque les recommandations se substituent à l'apprentissage privé), ce qui diminue le flux de nouvelles connaissances générées par l'humain vers le stock de connaissances générales. Il est important de rappeler que dans notre modèle de base, l'IA agentique n'affecte l'évolution des connaissances collectives que par le biais du comportement humain : elle ne génère pas directement de connaissances générales, mais elle peut les éroder indirectement en décourageant l'effort qui alimente l'apprentissage public.

Nous étudions ensuite la stabilité locale de l'état d'équilibre dit « d'effondrement des connaissances » (point fixe nul de F). Intuitivement, lorsque X est très faible, le rendement marginal de l'effort est lui aussi faible (puisque le terme de complémentarité est pondéré par $G(X)$) ; la question centrale est donc de savoir si les agents déploient encore *suffisamment* d'efforts pour empêcher X de tendre vers zéro. La réponse dépend de la courbure des coûts d'effort.

Lemme 2. (i) Si $\alpha - 1 > \frac{1}{4}$, alors $F(X) > X$ dans un voisinage de 0, de sorte que l'état d'équilibre de l'effondrement des connaissances (9) est localement instable.

(ii) Si $\alpha - 1 < \frac{1}{4}$, alors $F(X) < X$ dans un voisinage de 0, de sorte que l'état d'équilibre de l'effondrement des connaissances (9) est localement stable.

La stabilité locale de l'état d'équilibre nul revient à comparer l'effort nécessaire pour maintenir en vie une infime quantité de connaissance publique avec l'effort que les agents choisissent effectivement en équilibre. Pour un petit état $X \downarrow 0$, le maintien *de* X au cours de la période suivante nécessite un effort de l'ordre de $\Theta(X^2)$. Par

En revanche, la meilleure réponse donne un effort $e(X, \tau_A) = \Theta(X^{2(\alpha-1)})$. Par conséquent, lorsque $\alpha - 1 < \frac{1}{4}$, cet exposant est supérieur à 2, de sorte que l'effort d'équilibre diminue plus rapidement que X^2 car, à proximité de zéro, les agents investissent moins que le niveau de « maintien », et par conséquent, tout petit X positif revient à zéro. Cela implique que l'état d'équilibre de « effondrement des connaissances » (zéro) est localement stable. À l'inverse, lorsque $\alpha - 1 > \frac{1}{4}$, l'exposant est inférieur à 2, de sorte que l'effort diminue plus lentement que X^2 . Dans ce cas, lorsque X est proche de zéro, les agents déploient un effort plus que suffisant pour compenser la dérive ; un petit X positif tend alors à croître, et l'état zéro devient localement instable.

Le lemme 2 identifie la courbure des coûts d'effort comme un facteur déterminant pour savoir si l'effondrement des connaissances constitue un piège qui s'autoalimente. Cela semble intuitif : sous la fonction de coût e^α/α , l'élasticité de l'effort par rapport aux rendements marginaux est $\varepsilon := 1/(\alpha - 1)$. La condition $\alpha - \frac{1}{4} > \frac{1}{4}$ est donc équivalente à $\varepsilon < 4$: lorsque l'effort n'est pas trop élastique, les agents ne réduisent pas leur effort de manière trop brutale à mesure que les incitations s'affaiblissent, ce qui stabilise l'accumulation collective de connaissances et exclut l'effondrement comme résultat localement stable. En revanche, lorsque $\varepsilon > 4$ (c'est-à-dire $\alpha - 1 < \frac{1}{4}$), le système devient fragile : un stock de connaissances générales plus faible réduit fortement les incitations, ce qui réduit alors l'effort et diminue encore davantage le stock de connaissances générales. Cela ouvre la voie à l'échec de la coordination et à la dépendance au cheminement, comme nous le décrirons ci-dessous.

L'élasticité de l'effort peut être influencée par des facteurs technologiques et institutionnels. Par exemple, l'effort tend à être moins élastique lorsque les institutions ou les politiques imposent un niveau d'apprentissage de base et font en sorte qu'il soit difficile pour la quantité de connaissances de dépasser une certaine limite supérieure. Dans le contexte médical, cela peut être le cas parce que (i) des normes professionnelles élevées, telles que le nombre minimum d'années de formation médicale, garantissent un tel niveau de base, et (ii) l'ampleur et la difficulté des matières avancées concernées rendent très difficile l'atteinte de niveaux d'effort effectif très élevés.

3.7 Régime unique d'équilibre

Nous fournissons d'abord une analyse plus détaillée du régime dans lequel les coûts d'effort sont suffisamment convexes, ce qui signifie que

$$\alpha - 1 > \frac{1}{4}, \quad \text{ou, de manière équivalente} \quad \varepsilon := \frac{1}{\alpha - 1} < 4,$$

de sorte que l'effort optimal ne soit pas « trop élastique » par rapport aux variations de son rendement marginal, garantissant ainsi un état d'équilibre unique.

Comme la fonction de transition $X_{t+1} = F(X_t)$ est strictement croissante et continue (Lemme 1), les états stationnaires correspondent aux intersections de F avec la droite à 45 degrés. Le Lemme 2 implique que lorsque $\alpha - 1 > \frac{1}{4}$, la fonction de transition F se situe *au-dessus* de la droite à 45 degrés pour *des valeurs de X suffisamment proches de zéro*, ce qui garantit que le point fixe zéro est localement instable. Mais de plus, $F(X)$ est bornée vers le haut par Σ^{-2} (Lemme 1), tandis que la droite de pente 45° n'est pas bornée, de sorte que $F(X) < X$ pour tout X suffisamment grand. Par conséquent, par continuité, F doit croiser la droite de pente 45° au moins une fois sur $(0, \infty)$. Dans ce régime, la carte de transition d'état a également une pente inférieure à un chaque fois qu'elle croise la droite de 45° de sorte que l'état stationnaire (non trivial) est *unique*.

Proposition 3. *Supposons que $\alpha - 1 > \frac{1}{4}$. Il existe alors exactement deux équilibres en régime permanent, $0 = X_\ell^- < X_h^- < +\infty$, où X_h^- est localement stable et X_ℓ^- est localement instable. De plus, il existe un seul bassin d'attraction sur $\mathbb{R}_+ \setminus \{0\}$. Autrement dit, pour la suite $\{X_t\}$ définie par $X_{t+1} = F(X_t)$, on a $X_1 > 0 \Leftrightarrow X_t \rightarrow X_h^-$.*

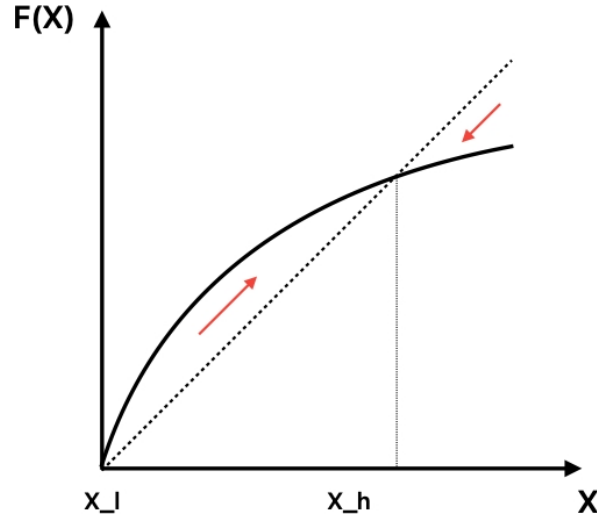


Figure 1 : Dynamique d'équilibre lorsque $\alpha - 1 > \frac{1}{4}$.

La proposition 3 décrit un « régime d'apprentissage robuste » : bien que $X = 0$ soit toujours un état d'équilibre (la connaissance sociale peut toujours rester à zéro si elle y parvient), il s'agit d'un résultat à la limite : tout stock de connaissance générale strictement positif, aussi minime soit-il, déclenche un effort d'apprentissage suffisant pour

plus que compenser la dépréciation résultant des changements, et l'économie converge vers l'état d'équilibre unique à haut niveau de connaissance X_h^- . En conséquence, il n'y a pas de défaillance de coordination : les résultats à long terme ne dépendent pas de l'histoire (à l'exception de la condition initiale dégénérée $X_1 = 0$).

Définissons l'effort d'équilibre et la précision idiosyncratique associés à l'état d'équilibre à haut niveau de connaissance :

$$e_h^- := e(X_h^-, \tau_A), \quad Y_h^- := \sigma^2 + \lambda_I e_h^- + \tau_A.$$

Proposition 4. *Supposons que $\alpha - 1 > \frac{1}{4}$. Les énoncés de monotonie suivants sont strictement valables :*

(i) *À mesure que la précision de l'IA agentique τ_A augmente :*

$$X_h^- \downarrow \text{ en } \tau_A, \quad e_h^- \downarrow \text{ dans } \tau_A, \quad Y_h^- \uparrow \text{ dans } \tau_A.$$

(ii) *À mesure que l'échelle d'agrégation I augmente :*

$$X_h^- \uparrow \text{ dans } I, \quad e_h^- \uparrow \text{ dans } I, \quad Y_h^- \uparrow \text{ dans } I.$$

La partie (i) met en évidence le fait que, même dans le cadre d'un apprentissage robuste, une IA agentique plus précise érode la connaissance collective. Une valeur plus élevée de τ_A réduit l'incitation des agents à investir dans un apprentissage coûteux (Observation 2), ce qui diminue le flux de nouvelles informations générées par les humains qui alimentent le stock de connaissances générales et déplace la carte de transition F vers le bas (Proposition 2). Comme X_h^- est le point fixe intérieur unique, il doit diminuer. Dans le même temps, la précision idiosyncratique Y_h^- augmente mécaniquement avec τ_A , même si la composante humaine endogène de Y_h^- (via e_h^-) diminue. Ainsi, l'IA agentique améliore les informations spécifiques au contexte disponibles à l'état stationnaire tout en réduisant le stock de connaissances générales qui soutient l'apprentissage humain complémentaire.

La partie (ii) met en évidence l'effet inverse : une plus grande capacité d'agrégation renforce la complémentarité intertemporelle entre les cohortes. À effort constant, une valeur plus élevée de I rend les contributions de chaque cohorte plus instructives ; à X constant, elle augmente également l'effort d'équilibre en accroissant le rendement social de l'apprentissage dont les agents héritent via X . Ces deux mécanismes font monter la courbe F et élèvent le point fixe stable unique X_h^- . En ce sens, des communautés de partage d'informations plus importantes (ou de meilleures technologies d'agrégation) rendent le stock de connaissances plus résilient et poussent l'économie vers un état d'équilibre à plus haut niveau de connaissances.

En résumé, la condition $\alpha - 1 > \frac{1}{4}$ implique que l'équilibre à long terme est unique et que l'état d'équilibre de l'effondrement des connaissances n'est pas un attracteur. Cependant, l'IA agentique peut toujours réduire X_h^- (et donc le stock de connaissances générales à long terme de la communauté), même si elle ne peut pas générer de manière générique un piège d'effondrement auto-entretenu dans cette plage de paramètres.

3.8 Régime à états stationnaires multiples et effondrement des connaissances

Dans le cadre de notre étude, la partie la plus intéressante de l'espace des paramètres est celle où plusieurs états stationnaires peuvent apparaître, c'est-à-dire

$$\alpha - 1 < \frac{1}{4}, \quad \text{ce qui revient à dire} \quad \varepsilon := \frac{1}{\alpha - 1} > 4,$$

auquel cas l'effort est hautement élastique et, d'après le lemme 2, l'état d'équilibre à connaissance nulle $X^- = 0$ est *localement stable*. Cette stabilité locale ouvre la voie à de *multiples* résultats stables à long terme.

Comme dans le cas où $\alpha - 1 \geq \frac{1}{4}$, les états stationnaires correspondent aux intersections de la carte de transition d'états F avec la droite à 45 degrés. Lorsque $\alpha - 1 < \frac{1}{4}$, le graphique de F se situe *en dessous* de la droite à 45 degrés pour *des valeurs de X proches de zéro*. Le fait que F recoupe ou non la droite de 45 degrés pour des valeurs plus élevées *de X* dépend de l'effort fourni par les agents et, par conséquent, de la force de substitution de l'IA agentique (représentée par τ_A). À mesure que τ_A augmente, l'effort de meilleure réponse $e(X, \tau_A)$ diminue ponctuellement (Observation 2), ce qui déplace l'ensemble de la carte F vers le bas et réduit (ou élimine) l'ensemble des points fixes positifs.

La proposition suivante formalise les deux possibilités qui en découlent. Notons par τ_A^c la *seuil de collapse complet*, c'est-à-dire la plus grande précision de l'IA agentique pour laquelle un état d'équilibre positif peut être maintenu.

Proposition 5. *Supposons que $\alpha - 1 < \frac{1}{4}$. Il existe un seuil $\tau_A^c \geq 0$, tel que*

- (i) *Si $\tau_A < \tau_A^c$ il existe trois équilibres en régime permanent $0 = X_\ell^- < X_m^- < X_h^- < +\infty$, où X_ℓ^- et X_h^- sont localement stables, tandis que X_m^- est localement instable.*

Le point instable X_m^- divise l'espace d'états en deux bassins d'attraction :

$$\begin{aligned} X_1 < X_m^- &\Rightarrow X_t \rightarrow X_\ell^- = 0 \\ X_1 > X_m^- &\Rightarrow X_t \rightarrow X_h^- \end{aligned}$$

- (ii) *Si $\tau_A > \tau_A^c$ alors l'unique équilibre en régime permanent est celui de l'effondrement de la connaissance, $X_\ell^- = 0$, et cet état stationnaire est globalement stable, ce qui signifie que $\forall X_1 \geq 0, X_t \rightarrow X_\ell^-$.*

La figure 2 illustre schématiquement les deux cas dans ce régime. Dans le panneau (a), la fonction F coupe trois fois la droite à 45 degrés : les points d'intersection inférieur et supérieur sont stables (pentes localement inférieures à un), tandis que le point d'intersection intermédiaire est instable (pente supérieure à un). Le point instable X_m^- agit donc comme un seuil : si la précision publique héritée tombe en dessous de X_m^- , la complémentarité entre la connaissance générale et l'effort privé devient trop faible, l'effort s'effondre progressivement et la communauté converge vers $X_\ell^- = 0$. Si, au contraire, X_1 dépasse X_m^- , la rétroaction positive entre les cohortes domine et le système converge vers l'état d'équilibre à haut niveau de connaissances X_h^- . Cette configuration implique une forme de *dépendance au cheminement* : le stock de connaissances initial détermine si la société tend ou non vers un effondrement des connaissances.

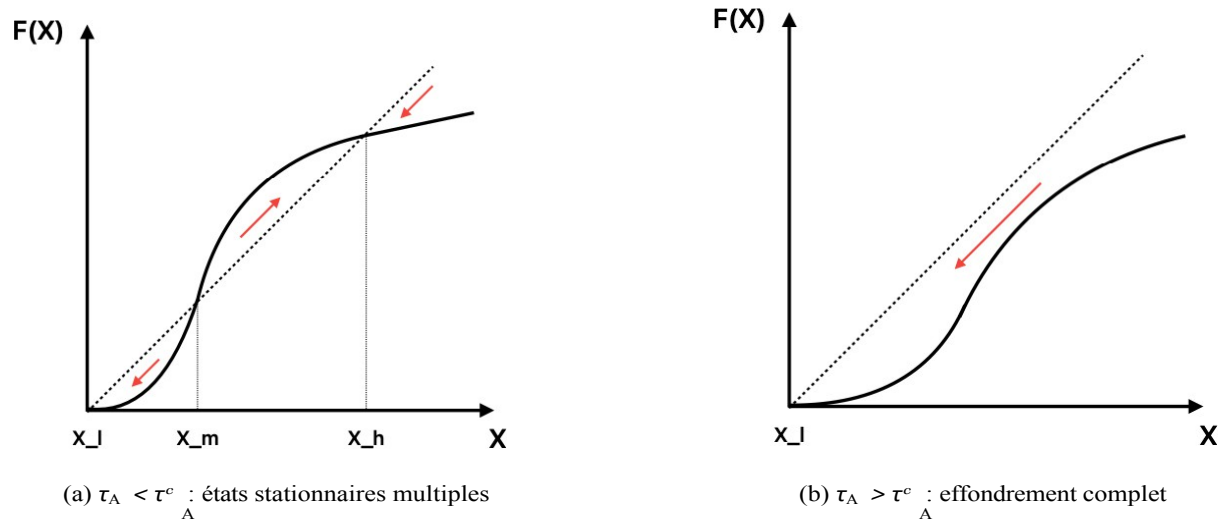


Figure 2 : Dynamique d'équilibre lorsque $\alpha - 1 < \frac{1}{4}$.

Le panneau (b) illustre un régime encore plus défavorable (du point de vue de la construction collective de connaissances) : lorsque τ_A dépasse τ_A^c , la fonction F se situe partout en dessous de la ligne de 45 degrés pour $X > 0$, et il ne peut y avoir d'états stables positifs. Dans ce cas, l'équilibre dynamique converge vers l'état stable d'effondrement des connaissances, quelles que soient les conditions initiales. Nous appelons ce phénomène effondrement total, et appelons τ_A^c le seuil d'effondrement total.

Une autre caractéristique intéressante de la proposition 5 est que le résultat à long terme peut varier de manière discontinue avec τ_A . À mesure que τ_A s'approche de τ_A^c par le bas, les deux points fixes positifs X_m^- et X_h^- se déplacent vers l'un vers l'autre et finissent par entrer en collision ; à τ_A^c , les états stationnaires positifs disparaissent, après quoi le seul attracteur restant est $X_\ell = 0$. Par conséquent, le stock de connaissances générales à long terme peut présenter une chute discontinue lorsque τ_A croise τ_A^c .

Nous établissons ensuite une analyse statique comparative pour les grandeurs clés qui ont émergé de manière endogène de la caractérisation : (i) le seuil d'effondrement total τ_A^c , (ii) le seuil du bassin d'attraction, X_m^- , et (iii) l'état d'équilibre à haut niveau de connaissance X_h^- . Nous considérons d'abord τ_A^c .

Proposition 6. *Supposons que $\alpha - 1 < \frac{1}{4}$. Alors :*

- (i) τ_A^c est croissant en fonction de la capacité d'agrégation I (de manière stricte lorsque $\tau_A^c > 0$).
- (ii) τ_A^c est croissant en fonction du degré de complémentarité Δ_X (de manière stricte lorsque $\tau_A^c > 0$).⁵

Les autres démonstrations de cet article sont présentées dans l'annexe B en ligne.

L'interprétation de la proposition 6 est simple : une valeur plus élevée *de* I renforce le signal social généré par un niveau donné d'effort par habitant, ce qui déplace la courbe F vers le haut et rend ainsi le

⁵ Notez que lorsque nous augmentons Δ_X , Δ_G doit être ajusté en conséquence pour satisfaire la contrainte de paramètre $\Delta_X + \Delta_G = 1$.

un stock de connaissances plus résilient face à l'effet de substitution d'effort de l'IA agentique. De même, une valeur plus élevée de Δ_X augmente le gain de complémentarité tiré de la correction conjointe de l'état commun et de l'état idiosyncrasique, et encourage ainsi un effort d'équilibre pour un niveau donné de stock de connaissances sociales, ce qui déplace la fonction F vers le haut. Dans les deux cas, la société peut tolérer un niveau plus élevé de précision de l'IA agentique avant que l'état d'équilibre à haut niveau de connaissances ne disparaisse.

Lorsque $\tau_A < \tau_A^c$, il n'y a pas d'effondrement complet, et le fait que la communauté converge vers un effondrement des connaissances ou vers l'état d'équilibre à haut niveau de connaissances dépend de la question de savoir si le stock initial de connaissances générales dépasse le seuil X_m^- . Le seuil X_m^- peut donc être interprété comme une mesure de la fragilité ou de la probabilité d'un échec de la coordination : plus X_m^- est grand, moins une société a de chances de se stabiliser dans l'état d'équilibre à haut niveau de connaissances. Nous présentons l'analyse statique comparative pour ce seuil X_m^- dans la proposition suivante :

Proposition 7. *Supposons que $\alpha - 1 < \frac{1}{4}$ et que $\tau_A < \tau^c$. Le seuil du bassin d'attraction, X_m^- , est croissant en τ_A et décroissant en I et Δ_X .*

En résumé, même lorsqu'un état d'équilibre à haut niveau de connaissance existe, une IA agentique plus précise augmente le stock minimal de connaissances générales requis pour que la communauté reste sur (ou revienne à) la trajectoire à haut niveau de connaissance. En revanche, une agrégation plus forte (I) et une complémentarité plus forte (Δ_X) réduisent le seuil du bassin d'attraction, X_m^- , et facilitent le maintien de l'équilibre à haut niveau de connaissance.

Enfin, soit (X_h^-, e_h^-, Y_h^-) l'état d'équilibre à haut niveau de connaissances lorsqu'il existe, où

$$e_h^- := e(X_h^-, \tau_A), \quad Y_h^- := \sigma^{-2} + \lambda_I e_h^- + \tau_A.$$

La proposition suivante résume la manière dont les résultats en régime permanent réagissent à τ_A et I .

Proposition 8. *Supposons que $\alpha - 1 < \frac{1}{4}$ et $\tau_A < \tau^c$. Les énoncés de monotonie suivants sont valables de manière stricte :*

(i) *À mesure que la précision τ_A de l'IA agentique augmente :*

$$X_h^- \downarrow \text{ dans } \tau_A, \quad e_h^- \downarrow \text{ dans } \tau_A.$$

Il existe $0 \leq \hat{\tau}_A < \tau^c$ tel que

$$Y_h^- \uparrow \text{ dans } \tau_A \text{ lorsque } \tau_A \in [0, \hat{\tau}_A], \quad Y_h^- \downarrow \text{ dans } \tau_A \text{ lorsque } \tau_A \in (\hat{\tau}_A, \tau_A^c).$$

(ii) *L'échelle d'agrégation I agit dans le sens inverse :*

$$X_h^- \uparrow \text{ dans } I, \quad e_h^- \uparrow \text{ dans } I, \quad Y_h^- \uparrow \text{ dans } I.$$

Cette proposition est similaire à la proposition 4, la seule différence étant la non-monotonie de Y_h^- lorsque l'on augmente τ_A . Cette non-monotonie reflète deux mécanismes opposés : une valeur plus élevée de τ_A améliore directement la précision contextuelle, mais la réduit indirectement en évinçant l'apprentissage privé (e_h^-). Pour des valeurs relativement faibles de τ_A , le bénéfice informationnel direct domine et Y_h^- augmente. En revanche, à proximité du seuil d'effondrement, le système devient fragile et l'effet d'éviction peut dominer, de sorte que même la précision contextuelle diminue avec une IA agentique plus précise.

4 Bien-être

Dans cette section, nous évaluons les effets de l'IA sur le bien-être à long terme. Nous considérons que l'objet de bien-être pertinent est l'utilité attendue en régime permanent d'une cohorte représentative. Nous supprimons le terme constant $f(0, 0)$ tout au long de l'analyse, de sorte que le bien-être est mesuré par rapport à la référence « sans information ».

4.1 Utilité attendue en régime permanent

Pour tout état d'équilibre (X^-, e^-, Y^-) , définissons l'utilité attendue en régime permanent comme suit

$$U^- := G(X^-) \cdot \Delta_G + G(X^-)G(Y^-) \cdot \Delta_X - \frac{1}{\alpha} e^{-\alpha}.$$

Sous l'hypothèse 1 ($\Delta_I = 0$), l'état d'équilibre avec effondrement des connaissances $(X^-, e^-) = 0$ donne $U^- = 0$.

Comme e^- est l'effort privé optimal étant donné (X^-, τ_A) , le bien-être peut s'écrire comme une fonction de (X^-, τ_A) uniquement :

$$U^-(X^-, \tau_A) = G(X^-) \cdot \Delta_G + \max_{e \geq 0} G(\sigma^{-2} + \lambda_1 e' + \tau_A) \cdot G(X^-) \Delta_X - \frac{1}{\alpha} (e')^\alpha. \quad (10)$$

Cette représentation distingue (i) l'effet du stock hérité de connaissances générales X^- , (ii) l'effet de la précision de la recommandation agentique τ_A , et (iii) le problème d'optimisation privée intra-cohorte concernant l'effort, qui met en balance une plus grande précision idiosyncratique et des coûts d'effort convexes.

Chaque fois qu'il existe un état d'équilibre positif à haut niveau de connaissance (tel que défini dans les propositions 3 et 5), nous désignons par (X_h^-, e_h^-, Y_h^-) le triplet d'équilibre associé et définissons le bien-être d'équilibre correspondant par $U^{+} := U^-(X_h^-, \tau_A)$.

Le bien-être à long terme réalisé dépend du fait que la dynamique d'équilibre admette un état d'équilibre unique (stable) ou présente une multiplicité et une dépendance au chemin.

(i) Régime d'équilibre unique : $\alpha - 1 > 1$. D'après la proposition 3, pour toute condition initiale $X_1 > 0$, l'équilibre converge vers l'état d'équilibre unique à haut niveau de connaissance X_h^- . Le bien-être à long terme est donc déterminé de manière unique et est égal à U^{+} . Dans ce régime, il n'y a pas de problème de sélection d'équilibre : l'IA n'affecte le bien-être que par le biais de ses variations (X_h^-, e_h^-, Y_h^-) .

(ii) Régime à états d'équilibre multiples : $\alpha - 1 < \frac{1}{4}$. Dans ce régime, la dynamique d'équilibre peut présenter un échec de coordination. Si $\tau_A \geq \tau^c$, un effondrement complet se produit et le bien-être à long terme est égal à 0 pour toutes les conditions initiales. Si, au contraire, $\tau_A < \tau^c$, l'état d'équilibre d'effondrement des connaissances et l'état d'équilibre de haut niveau de connaissances coexistent, séparés par le seuil X_m^- . Le bien-être à long terme dépend alors du stock initial X_1 : les cohortes convergent vers X_h^- (et obtiennent U^{+}) si X_1 se situe au-dessus du seuil du bassin d'attraction, mais convergent vers l'état d'équilibre de l'effondrement des connaissances (et obtiennent 0) si X_1 se situe en dessous de ce seuil. Nous devons donc examiner l'impact de l'IA à la fois sur l'utilité de l'état d'équilibre à haut niveau U^{+} et sur le seuil du bassin d'attraction X_m^- .

4.2 L'impact de la connaissance générale

Proposition 9. *Chaque fois qu'un état d'équilibre à haut niveau de connaissance existe, U^{+} est strictement croissante en I .*

En gros, cette proposition montre qu'une connaissance générale plus étendue améliore toujours le bien-être.

Plus précisément, *un I plus élevé permet une meilleure agrégation des connaissances générales et augmente le bien-être par deux canaux distincts :*

- **Effet de niveau.** Une valeur plus élevée de I augmente le pouvoir informatif du signal public généré par un effort par habitant donné, ce qui décale la carte de transition F vers le haut et augmente le stock d'équilibre de la connaissance générale X_h^- . D'après l'équation (10), ce stock plus élevé se traduit alors par une valeur plus élevée de U^{+} .
- **Effet de sélection.** Lorsque $\alpha - 1 < \frac{1}{4}$, un I plus élevé augmente non seulement U^{+} mais rend également l'état d'équilibre à haut niveau de connaissances plus facile à maintenir dynamiquement en abaissant le seuil du bassin d'attraction, X_m^- . Cela élargit le bassin d'attraction de l'état d'équilibre à haut niveau de connaissances, augmentant ainsi le bien-être réalisé pour un ensemble plus large de conditions initiales.

Dans la pratique, une *valeur I* plus élevée peut résulter d'un partage d'informations plus efficace au sein des communautés et d'améliorations apportées à d'autres technologies d'agrégation de l'information qui élargissent la portée des connaissances générales générées par l'homme. Notre modèle suggère donc que ce type de partage d'informations améliore à la fois les efforts humains consacrés à l'apprentissage et réduit le risque d'effondrement des connaissances.

4.3 Les effets de l'IA agentique

Les effets sur le bien-être d'une IA agentique plus puissante (τ_A plus élevé) sont intrinsèquement ambigus, car ils s'exercent par le biais de deux canaux opposés dans (10) : *un τ_A plus élevé augmente directement la précision idiosyncratique, mais réduit indirectement le stock de connaissances générales à long terme en évinçant l'effort.* En différenciant

(10) à l'état d'équilibre de haut niveau de connaissances et en appliquant le théorème de l'enveloppe, on obtient

$$\frac{\partial U^+}{\partial \tau_A} = \underbrace{g(Y_h^-) \cdot G(X_h^-) \Delta_X}_{\text{Direct Effet, } \geq 0} + \underbrace{\frac{\partial G(X_h^-)}{\partial \tau_A} \cdot (\Delta_G + G(Y_h^-) \Delta_X)}_{\text{Indirect Effet, } \leq 0}.$$

L'effet direct correspond au gain mécanique résultant de recommandations individualisées plus précises. *L'effet indirect* reflète un effet d'éviction dynamique : une valeur plus élevée de τ_A réduit l'effort d'équilibre, ce qui affaiblit le flux d'informations générées par l'humain vers le stock de connaissances générales et diminue ainsi X_h^- .

Introduisons les abréviations suivantes pour ces deux effets :

$$DE = g(Y_h^-) \cdot G(X_h^-) \Delta_X, \quad IE = - \frac{\partial G(X_h^-)}{\partial \tau_A} \cdot (\Delta_G + G(Y_h^-) \Delta_X).$$

Alors $\frac{\partial U^+}{\partial \tau_A} > 0$ si et seulement si $DE > IE$.

Nous établissons ci-dessous notre résultat principal, à savoir que l'effet négatif est susceptible de prévaloir lorsque l'IA agentique est très précise. En effet, le rapport IE/DE augmente toujours avec τ_A (sous des contraintes de régularité naturelles sur l'espace des paramètres), et est strictement supérieur à un lorsque τ_A est suffisamment grand. Deux forces économiques poussent le rapport IE/DE à la hausse à mesure que τ_A augmente. Premièrement, le bénéfice direct DE présente des rendements décroissants, car $g(\cdot)$ diminue à mesure que la précision idiosyncratique augmente. Cette force rend compte du fait qu'une fois que les recommandations individualisées sont déjà assez précises, de nouvelles augmentations de τ_A n'amélioreront pas beaucoup les prédictions. Deuxièmement, la perte indirecte IE s'accroît lorsque le stock de connaissances générales est plus faible. En effet, lorsque X_h^- est faible, les améliorations marginales des connaissances générales sont particulièrement précieuses (la pente $g(X_h^-)$ est élevée), de sorte que les réductions de X_h^- se traduisent par d'importantes pertes de bien-être. Ensemble, ces forces impliquent que des niveaux modestes d'IA agentique peuvent accroître le bien-être, mais qu'une IA agentique suffisamment précise le réduit.

Nous considérons d'abord le régime d'équilibre unique $\alpha - 1 > \frac{1}{4}$. Pour faciliter l'exposé, nous ajoutons également une condition de régularité simple sur la précision idiosyncratique a priori.⁶

Hypothèse 2. $\sigma^{-2} \geq \sqrt{2} - 1$.

La statique comparative du bien-être pour l'état d'équilibre à haut niveau de connaissance est résumée dans la proposition suivante :

⁶ Essentiellement, l'hypothèse 2 exclut un profil de bien-être à pics multiples dans lequel la dérivée du bien-être change deux fois de signe. Il est facile de comprendre pourquoi cela peut se produire si l'hypothèse 2 ne se vérifie pas. En particulier, on peut écrire la

effet indirect tel que $IE = - \frac{\partial e_h^-}{\partial \tau_A} \cdot G'(Ie_h^-) \cdot I \cdot (\Delta_G + G(Y_h^-) \Delta_X)$, et on peut en conclure que cet effet indirect devient très

est importante lorsque la précision idiosyncratique $\gamma_h = \sigma^{-2} + \lambda_1 e_h^- + \tau_A$ est proche de zéro. Intuitivement, dans cette région, l'effort d'équilibre e_h^- devient très sensible à τ_A car l'effet de substitution $|\frac{\partial e_h^-}{\partial \tau_A}|$ devient très important. Pour s'en rendre compte, notons que e_h^- satisfait

$\Delta_X \cdot G_W(Ie_h^-) \cdot \lambda_1 g(\sigma^{-2} + \lambda_1 e_h^- + \tau_A) = e_h^{-\alpha-1}$ et lorsque γ_h est faible, le membre de gauche devient très sensible à τ_A ,

car $g(\cdot)$ est très raide près de zéro ; en particulier, $g'(\tau) = -1 \cdot e^{-\frac{\sqrt{1-\tau}}{2\pi} - \frac{\tau}{2}} + 1 \cdot \tau = \Theta(\tau^{-\frac{3}{2}})$ lorsque $\tau \rightarrow 0$, ce qui viole

unicité de pic. On obtient un exemple simple violant l'unicité de pic de cette manière lorsque $\alpha = 100$, $\Delta_X = 0,005$, $\Delta_G = 0,995$, $\lambda_1 = 0,05$, $\lambda_G = 1$, $\sigma^{-2} = 0$, $\Sigma^{-2} = 1$, $I = 1$.

Proposition 10. Supposons que $\alpha - 1 > \frac{1}{4}$ et que l'hypothèse 2 soit vérifiée. Il existe un seuil fini τ^* pour $0 \leq \tau^* < +\infty$, de sorte que :

A

- (i) U^+_{-} soit strictement croissante en τ_A sur $(0, \tau^*_A)$.
- (ii) U^+_{-} est strictement décroissante en τ_A sur $(\tau^*_A, +\infty)$.
- (iii) U^+_{-} tend vers 0 lorsque τ_A tend vers $+\infty$.

Dans le régime d'équilibre unique, il n'y a pas d'effet de sélection : l'économie converge toujours vers X^-_h à partir de tout $X_1 > 0$. Le compromis en matière de bien-être porte donc uniquement sur la manière dont U^+_{-} varie en fonction de τ_A . La

proposition implique donc que τ^* est la précision de l'IA agentique maximisant le bien-être, toutes autres choses étant égales par ailleurs paramètres A fixés. La limite $U^+_{-} \rightarrow 0$ lorsque $\tau_A \rightarrow \infty$ met en évidence le cas limite d'un effondrement complet : si l'IA agentique devient suffisamment puissante, l'effort d'équilibre disparaît et la connaissance générale s'effondre, avec des effets très néfastes sur le bien-être

Dans le régime à états d'équilibre multiples, $\alpha - 1 < \frac{1}{4}$, l'IA agentique produit un effet *de sélection* supplémentaire : elle peut réduire (et, à terme, éliminer) le bassin d'attraction du scénario à haut niveau de connaissance. Nous exposons d'abord une analyse statique comparative du bien-être similaire pour l'état d'équilibre à haut niveau de connaissance.

Proposition 11. Supposons que $\alpha - 1 < \frac{1}{4}$ et que l'hypothèse 2 soit vérifiée. Il existe un seuil fini τ^* satisfaisant $0 \leq \tau^*_A < \tau^c_A$ tel que :

A

- (i) U^+_{-} est strictement croissante en τ_A sur l'intervalle $(0, \tau^*_A)$.
- (ii) U^+_{-} est strictement décroissante en τ_A sur (τ^*_A, τ^c_A) .
- (iii) $U^+_{-} = 0$ lorsque $\tau_A > \tau^c_A$.

Par rapport au régime d'équilibre unique, les conséquences de l'IA agentique sur le bien-être sont ici plus marquées, car une IA suffisamment précise peut induire un changement de régime discontinu : dès que τ_A dépasse τ^c , l'équilibre à haut niveau de connaissance disparaît et l'effondrement à long terme des connaissances devient inévitable. De plus, même pour $\tau_A < \tau^c$, le bien-être réalisé ne suit pas nécessairement U^+_{-} car l'économie peut ne pas parvenir à se coordonner sur le bassin d'attraction à haut niveau de connaissance, et une valeur plus élevée de τ_A peut accroître la probabilité d'un échec de coordination en augmentant le seuil du bassin d'attraction, X^-_m .

4.4 Asymptotiques pour $I \rightarrow +\infty$

Il est également instructif d'examiner le cas où la capacité de l'économie à agréger l'information devient très grande, c'est-à-dire $I \rightarrow \infty$.

Proposition 12. Supposons que l'hypothèse 2 soit vérifiée. Alors, lorsque la capacité d'agrégation I tend vers l'infini (tous les autres paramètres étant fixes), on a : ⁷

$$\lim_{I \rightarrow +\infty} \frac{\tau_A^c}{\log I} = 2(\alpha - 1),$$

et

$$\lim_{I \rightarrow +\infty} \frac{\tau_A^*}{\log I} = \frac{2(\alpha - 1)}{\alpha}.$$

La proposition 12 aboutit à un résultat très net de type « loi d'échelle ». Tant le seuil d'effondrement τ_A^c que la précision maximisant le bien-être τ_A^* n'augmentent que *de manière logarithmique* en I , reflétant une forte diminution des retours à l'agrégation. La raison économique est que, à mesure que τ_A augmente, la valeur marginale de l'effort humain dans la production d'une précision idiosyncratique diminue très rapidement (car $g(\cdot)$ décroît de manière exponentielle avec la précision), de sorte que l'effort d'équilibre diminue fortement. Par conséquent, le maintien d'un niveau donné de connaissance publique nécessite que l'échelle d'agrégation croisse très rapidement, ce qui se traduit par une évolution de τ_A proportionnelle au logarithme de I . Enfin, ce résultat asymptotique implique que même lorsque l'agrégation devient arbitrairement puissante,

la précision maximisant le bien-être reste à une fraction fixe de la limite d'effondrement :

$$\frac{\tau_A^*}{\tau_A^c} \rightarrow \frac{1}{\alpha}.$$

Une autre conclusion importante est donc que le planificateur choisit de manière optimale de rester à distance du point de basculement.

4.5 Conception de l'information

Motivés par les effets négatifs potentiels de l'IA agentique et le risque d'effondrement des connaissances, nous examinons maintenant s'il existe des principes simples de conception de l'information ou des politiques de régulation susceptibles d'améliorer le bien-être.

Formellement, nous modélisons la réglementation comme une *altération* de la recommandation de l'IA agentique par un bruit gaussien additif indépendant (ce qui implique que l'organisme public ne peut pas produire l'information, mais peut partiellement supprimer l'information disponible). Une politique de conception de l'information peut alors être représentée par la séquence $\kappa = \{\kappa_t\}_{t \geq 1}$, telle que $\kappa_t \in [0, +\infty]$, où κ_t est la précision du bruit ajouté à la période t .⁸ Sous la politique κ , le signal transmis à l'agent i à la période t est

$$\tilde{s}_{i,t}^A = s_{i,t}^A + \eta_{i,t}, \quad \eta_{i,t} \sim \mathbf{N}(0, \kappa_t^{-1}),$$

$\{\eta_{i,t}\}$ étant indépendant d'un agent à l'autre et dans le temps. Les signaux étant gaussiens, le brouillage est

⁷ Ici, τ^* désigne la précision agentique (unique) maximisant le bien-être, caractérisée dans les propositions 10 et 11.

⁸ $\kappa_t = +\infty$ correspond à l'absence de brouillage (révélation complète), tandis que $\kappa_t = 0$ correspond à une suppression totale.

ce qui revient à remplacer la précision agentique brute τ_A par une précision *effective*

$$\tau_A^{(t)} = \tau_A(\kappa) := \tau_A^{-1} + \kappa_t^{-1} \in [0, \tau_A].$$

Par conséquent, le choix de κ_t par le régulateur est, sans perte de généralité, lié à un plafond variable dans le temps

$\tau_A^{(t)} \leq \tau_A$ sur l'informativité du signal agentique.

Pour une politique κ donnée, l'équilibre dynamique induit est caractérisé exactement comme dans la proposition 1, en remplaçant τ_A par $\tau_A^{(t)}$ période par période. Soit $U_t(\kappa)$ l'utilité attendue de la cohorte *de la période t* dans l'équilibre induit par κ . Nous évaluons les politiques en fonction du bien-être moyen à long terme :

$$U(\kappa) := \liminf_{T \rightarrow \infty} \frac{\sum_{t=1}^T U_t(\kappa)}{T}.$$

Étant donné que les transitions finies ont un poids nul selon ce critère, il est naturel de se concentrer sur les politiques qui finissent par se stabiliser à un plafond constant à long terme. Cette classe de politiques correspond à un régime réglementaire permanent, éventuellement précédé d'une phase temporaire de « moratoire ».

Appelons une politique d'information κ *finalement constante* s'il existe $M < \infty$ et $\kappa^- \in [0, +\infty]$ tels que $\kappa_t = \kappa^-$ pour tout $t \geq M$; notons K^{EC} l'ensemble de ces politiques.

Nous considérons le cas à régimes multiples $\alpha - 1 < 1$. La proposition suivante caractérise la stratégie optimale de brouillage gaussien dans ce régime.

Proposition 13. *Supposons que l'hypothèse 2 soit vérifiée et que $\alpha - 1 < 1$, de sorte que (i) l'effondrement complet seuil τ^c et les objets en régime permanent $\{X_m^-(\tau_A), X_h^-(\tau_A)\}_{\tau_A \leq \tau^c}$ sont tels que dans la proposition 5, et (ii) le bien-être en régime permanent à haut niveau de connaissance $U^+(\tau_A)$ est bien défini et est maximisé pour une valeur unique de $\tau^* \in [0, \tau^c]$*

comme dans la proposition 11.

Fixons une précision publique initiale $X_1 > 0$ telle que $X_1 > X_m^-(0)$, de sorte que, sous suppression totale ($\kappa_t = 0, \forall t$), la dynamique d'équilibre converge vers l'état stationnaire à haute connaissance $X_h^-(0)$. Définissons $\tau^* := \min\{\tau_A, \tau^*\}$ et choisissons κ^* tel que $\tau_A(\kappa^*) = \tau^*$. Alors :

(i) *Il existe un M^* fini tel que la politique à deux phases*

$$\kappa_t^{2ph} := \begin{cases} 0, & t < M^*, \\ \kappa^*, & t \geq M^*, \end{cases}$$

satisfait

$$U(\kappa^{2ph}) = U^+(\tau^*).$$

(ii) *Cette politique en deux phases est optimale parmi toutes les politiques finalement constantes. Autrement dit, $\sup_{\kappa \in K^{EC}} U(\kappa) =$*

$$U(\kappa^{2ph}) = U^+(\tau^*).$$

La logique économique de la politique en deux phases est instructive et peut être comprise en considérant

chaque phase séparément :

- **Phase 1 (reconstruction du stock).** En fixant $\kappa_t = 0$, on supprime totalement les recommandations des agents, ce qui les oblige à s'appuyer sur leur propre apprentissage. L'effort étant décroissant par rapport à la précision effective des agents, cette phase maximise l'effort d'équilibre et accélère la reconstruction de la connaissance générale publique, poussant l'état vers le bassin d'attraction de l'état stationnaire à haut niveau de connaissance sous le plafond final.
- **Phase 2 (plafonnement au niveau optimal)** Une fois que le stock de connaissances générales s'est reconstitué, le régulateur assouplit le moratoire mais plafonne de manière permanente la précision effective à τ^* , le niveau à long terme maximisant le bien-être (sous réserve de faisabilité). Si $\tau_A \leq \tau_A^*$, alors $\tau^* = \tau_A$ et la phase 2 correspond à l'absence de brouillage ; si $\tau_A > \tau_A^*$, la phase 2 met en œuvre un plafond de précision strict.

Intuitivement, la phase 1 s'attaque au problème de *la dépendance vis-à-vis du cheminement* : elle garantit que l'économie échappe au bassin d'attraction de l'état stationnaire caractérisé par un effondrement des connaissances. La phase 2 fixe ensuite un plafond optimal à long terme afin de constituer un stock suffisant de connaissances générales. Si le stock initial de connaissances générales est suffisamment élevé, la phase 1 devient alors superflue.

5 Extensions

Dans cette section, nous abordons plusieurs extensions importantes.

5.1 Communauté contre agrégation par l'IA

Dans le modèle de référence, l'IA agentique n'affecte que les informations spécifiques au contexte, mais ne modifie pas directement le stock de connaissances générales de l'économie (ou n'améliore pas la capacité d'agrégation I). Cette sous-section examine une extension dans laquelle la même technologie d'IA (i) fournit simultanément des recommandations spécifiques à chaque agent et (ii) améliore le stock de connaissances générales.

L'IA agentique fournit un signal de recommandation à chaque agent, comme dans le modèle de base :

$$s_{i,t}^A \sim \mathcal{N}(\vartheta_{i,t}, \tau_A^{-1}).$$

Nous supposons en outre qu'un système d'IA *présentant un τ_A plus élevé* augmente également la capacité d'agrégation effective. Concrètement, nous remplaçons la taille fixe des îlots I par une échelle d'agrégation *effective* dépendante de l'IA

$$I(\tau_A) := I_0 + \exp(\eta \tau_A), \quad (11)$$

où $I_0 > 0$ est la capacité d'agrégation pré-IA (de référence) et $\eta \geq 0$ détermine dans quelle mesure l'amélioration des capacités des agents se traduit par une meilleure agrégation.

La technologie de signal public reste inchangée, à l'exception de ce remplacement. Dans le cadre d'un effort symétrique e_t : à la période t , l'effort global devient $E_t = l(\tau_A) e_t$, de sorte que le signal public au niveau de l'île est

$$s_t^C \sim N \left(\vartheta_t \frac{1}{\lambda_G l(\tau_A) e_t} \right).$$

Toutes les autres parties du modèle restent inchangées.

Contrairement au modèle de référence, τ_A regroupe désormais deux forces opposées :

- **Substitution (canal agentique)** : une valeur plus élevée de τ_A réduit l'effort $e(X, \tau_A)$, ce qui diminue la création de connaissances générales au niveau individuel ;
- **Complémentarité (canal d'agrégation)** : une valeur plus élevée de τ_A augmente $l(\tau_A)$, rendant tout effort humain donné plus informatif sur le plan social.

La force qui domine dépend des valeurs des paramètres, et plus particulièrement de η , qui mesure la réactivité de l'agrégation face à la capacité de l'IA agentique.

Proposition 14. *Supposons que l'agrégation pilotée par l'IA ne croisse pas trop rapidement, c'est-à-dire que $\eta < \eta^{\text{crit}} := \frac{1}{2(\alpha-1)}$. Alors l'état d'équilibre à haut niveau de connaissance (lorsqu'il existe) satisfait $\bar{X}(\tau) = 0$ lorsque $\tau \rightarrow \infty$. Il en résulte que le niveau d'IA maximisant le bien-être est fini. Autrement dit, $\arg \max_{\tau_A \geq 0} U^{\bar{\pi}} + \bar{A}(\tau_A) < +\infty$.*

La proposition 14 souligne que le fait de permettre à l'IA d'améliorer simultanément le stock de connaissances générales ne modifie pas nos principales conclusions, tant que cet effet n'est pas trop prononcé. Comme d'habitude, les recommandations générées par l'IA se substituent à l'effort humain ; ainsi, lorsque τ_A augmente, l'effort fourni diminue. La meilleure agrégation fournie par l'IA agit dans le sens inverse. À condition que $\eta < \eta^{\text{crit}} = 1/[2(\alpha - 1)]$, ce deuxième effet n'est pas trop fort, ce qui garantit que le stock de connaissances générales continue de s'amenuiser lorsque $\tau_A \rightarrow \infty$.

5.2 Données synthétiques

Le modèle de référence part du principe que les nouvelles connaissances *générales* ne sont générées que par l'effort humain, puis intégrées au corpus de connaissances générales. Il s'agit d'une approximation raisonnable dans les domaines où la vérification est lente ou coûteuse, de sorte que le contenu généré par le modèle risque de devenir autoréférentiel plutôt que véritablement informatif. En revanche, dans des environnements régis par des règles ou vérifiables mécaniquement (tels que l'apprentissage par renforcement appliqué aux jeux, la recherche algorithmique ou la vérification formelle de preuves), l'IA peut générer des « données synthétiques » associées à un signal de correction objectif et peut donc créer une véritable nouveauté (comme dans *AlphaGo Zero* via l'auto-jeu Silver et al., 2017 ; la recherche par multiplication matricielle rapide Fawzi et al., 2022 ; et les systèmes récents de démonstration de théorèmes qui s'appuient sur des assistants de preuve Xin et al., 2024, Lin et al., 2025, Dong et Ma, 2025).⁹

⁹ Nous notons également que l'intuition selon laquelle « les données synthétiques permettent de surmonter le manque de données » est particulièrement forte lorsque les résultats du modèle peuvent être directement vérifiés par rapport à une vérité de référence. Pour la génération de texte en monde ouvert ou dans les problèmes où il n'existe pas de

Pour modéliser la possibilité de données synthétiques de manière simple, supposons qu'à chaque période t , un signal exogène supplémentaire concernant l'état commun ϑ_t soit produit avec une précision $\tau_{\text{syn}} \geq 0$, et qu'il soit intégré au même signal public observé par les cohortes futures. Dans le cadre de notre modèle, ces données synthétiques ajoute alors τ_{syn} unités de précision au signal public de fin de période, de sorte que la loi de variation de la précision de la connaissance générale devient

$$\mathbf{v}_{t+1}^{-1} = (\mathbf{X}_t + \lambda_G \mathbf{E}_t + \tau_{\text{syn}})^{-1} + \Sigma^2, \quad \mathbf{E}_t = I e(\mathbf{X}_t, \tau_A). \quad (12)$$

De manière équivalente, la carte de transition d'état devient

$$F_{\text{syn}}(\mathbf{X}) = \Sigma^2 + (\mathbf{X} + \lambda_G I e(\mathbf{X}, \tau_A) + \tau_{\text{syn}})^{-1}. \quad (13)$$

Cette équation montre déjà que nos mécanismes clés restent inchangés : une plus grande précision de l'agent τ_A évince l'effort humain $e(\mathbf{X}, \tau_A)$ et réduit ainsi le flux de connaissances générales générées par l'humain. La différence réside dans le fait que la découverte synthétique rend le système moins fragile : si $\tau_{\text{syn}} > 0$, alors $F_{\text{syn}}(0) > 0$, de sorte que l'état stationnaire dégénéré de « connaissance nulle » n'est plus un point fixe. Dans la région des paramètres $\alpha - 1 < 1$ où le modèle de référence présente un état d'équilibre stable de « collapse », les données synthétiques remplacent le piège de « collapse » par un état d'équilibre de « faible connaissance » qui reste strictement positif. Comme le terme ajouté rompt l'argument de convexité utilisé pour caractériser l'ensemble complet des états d'équilibre, nous nous concentrons sur les états d'équilibre extrêmes.¹⁰

Proposition 15. *Supposons que $0 < \tau_{\text{syn}} < +\infty$. Soient (X_ℓ^-, e_ℓ^-) et (X_h^-, e_h^-) respectivement le plus petit et les états stationnaires maximaux du système dynamique $\mathbf{X}_{t+1} = F_{\text{syn}}(\mathbf{X}_t)$.*

1. *L'état stationnaire minimal est strictement positif : $X_\ell^- > 0$ et $e_\ell^- > 0$.*
2. *Les quatre grandeurs X_h^-, e_h^-, X_ℓ^- et e_ℓ^- sont strictement croissantes en fonction de la capacité d'agrégation I , strictement décroissantes en précision agentique τ_A et strictement croissantes en précision synthétique τ_{syn} .*

Cette proposition établit donc que nos principaux résultats se généralisent à cet environnement avec des données synthétiques, à condition que ces données ne puissent pas remplacer entièrement les connaissances générées par l'humain.

5.3 Séparabilité (imparfaite) de l'effort

Dans le modèle de référence, une unité d'effort d'apprentissage produit conjointement (i) des connaissances spécifiques au contexte et

(ii) une contribution « limitée » à la connaissance générale. Ici, nous permettons aux agents de *dissocier partiellement* ces deux

vérité de base, la vérification n'est généralement pas automatique, de sorte que l'auto-apprentissage récursif peut devenir autoréférentiel et dégrader l'environnement informationnel, comme le montrent, par exemple, Shumailov et al. (2024). Dans de tels environnements, les données synthétiques doivent être considérées comme un *complément* aux données réelles (souvent via une reformulation ou une augmentation) qui améliore principalement l'efficacité de l'apprentissage, tandis que les corpus entièrement générés par des modèles peuvent détériorer les performances en aval (Maini et al., 2024 ; Kang et al., 2025). Nous réservons une analyse plus systématique de ce type d'utilisation des données synthétiques à des travaux futurs.

¹⁰ Voir la démonstration des propositions 3 et 5, et en particulier le lemme A-2, pour plus de détails.

sorties. Nous modélisons cette séparation partielle en conservant la structure de base suivante

$$s_{i,t}^H \sim N \left(\vartheta_{i,t}, \frac{1}{\lambda l e_{i,t}} \right), \quad s_{i,t}^A \sim N \left(\vartheta_{i,t}, \frac{1}{\tau_A} \right), \quad \text{coût} = \frac{1}{\alpha} e_{i,t}^\alpha, \quad \alpha > 1,$$

mais supposons maintenant que l'effort de l'agent i fournisse un signal concernant l'état commun ϑ_t avec une précision proportionnelle à

$$\lambda e_{i,t}^\beta, \quad \beta \in [0, +\infty).$$

Comme précédemment, l'effort effectif global pour la connaissance générale sur l'île m peut être calculé comme suit

$$E_{m,t}^{(\beta)} := \int_{\text{Im}} e_{i,t}^\beta di, \text{ et donc } s_{m,t}^C \sim N \left(\vartheta_t, \frac{1}{\lambda G E_{m,t}^{(\beta)}} \right). \quad (14)$$

Par conséquent, cette généralisation nous permet de paramétrer avec β le degré de « regroupement » entre l'effort d'apprentissage individuel et la construction collective de connaissances. En particulier,

$$\frac{d \log(e^\beta)}{d \log e} = \beta,$$

de sorte qu'un β plus petit signifie une « plus grande séparabilité », car les deux dimensions de la connaissance ne varient plus parfaitement de concert. Le cas limite $\beta = 0$ correspond à une séparabilité parfaite, en ce sens

que les variations de $e_{i,t}$ ne modifient pas l'entrée d'apprentissage public, puisque $e^\beta = 1$ est constant. La proposition suivante constitue notre résultat principal dans cette sous-section :

Proposition 16. *Dans le cadre de cette extension, si $\beta > 0$ (modèle non parfaitement séparable), tous les résultats du modèle de base restent valables, à l'exception de la condition de stabilité $\alpha - 1 < \frac{(1)}{4}$ qui est remplacée par $\alpha - 1 < \frac{\beta}{4}$, et les résultats asymptotiques de la proposition 12 étant remplacés par*

$$\lim_{l \rightarrow \infty} \tau_{\Delta=}^c = \frac{2(\alpha - 1)}{\beta} \quad \text{et} \quad \lim_{l \rightarrow \infty} \tau_{\Delta=}^* = \frac{2(\alpha - 1)}{\alpha - 1 + \beta}.$$

Nous examinons ici comment β modifie la stabilité du système à l'approche de l'effondrement des connaissances. Comme le précise le lemme 2, la stabilité locale de l'état d'équilibre d'effondrement des connaissances est déterminée par la force de l'apprentissage public à proximité de $X = 0$. Dans cette région, le stock de connaissances générales est proportionnel à e^β ; ce qui importe donc, c'est la rapidité avec laquelle e^β s'annule lorsque $e \downarrow 0$. Si β est grand, e^β diminue très rapidement lorsque $e \downarrow 0$, renforçant ainsi la logique d'auto-renforcement qui sous-tend l'effondrement. À l'inverse, si β est petit, un effort humain modeste suffira à produire des améliorations significatives de la connaissance générale, rendant l'état d'équilibre d'effondrement des connaissances localement instable.

6 Conclusion

Dans cet article, nous avons présenté un cadre simple pour étudier les implications des nouvelles technologies d'IA générative qui promettent de fournir des informations et des recommandations spécifiques au contexte aux décideurs humains. Notre cadre repose sur trois idées fondamentales :

1. Les bonnes décisions humaines combinent les connaissances générales et les informations spécifiques au contexte.
2. Les efforts humains visant à améliorer les capacités cognitives génèrent ces deux types d'informations, le principal bénéfice privé provenant des informations spécifiques au contexte.
3. Les contributions individuelles à la connaissance générale créent des externalités pour les autres qui s'appuient sur cette connaissance générale.

Ces trois observations impliquent ensemble que la principale motivation de l'effort individuel est souvent l'acquisition d'informations spécifiques au contexte, tandis que la connaissance générale générée par un individu est principalement une externalité. Par conséquent, une meilleure connaissance générale dans la société est un *complément* à l'effort d'apprentissage humain, tandis que de meilleures recommandations spécifiques au contexte en sont *des substituts*. Comme l'IA générative promet de fournir ce type d'informations spécifiques au contexte, elle peut constituer un tel substitut et réduire l'effort humain. Mais comme une réduction de l'effort humain diminue la construction de connaissances générales, l'IA générative (en particulier l'IA agentique) peut pousser *dynamiquement* une société vers une baisse de l'information effective et même conduire à un état d'équilibre caractérisé par un effondrement des connaissances.

Notre analyse est purement théorique et montre qu'un tel cadre est gérable et permet d'obtenir un certain nombre de résultats de statique comparative nouveaux et intuitifs. Elle clarifie également les conditions dans lesquelles un état d'équilibre de « effondrement des connaissances » émerge et ce qui détermine l'étendue de son bassin d'attraction. La gérabilité de notre modèle nous permet également d'envisager plusieurs extensions, visant à démontrer la robustesse de nos principales conclusions qualitatives.

Plusieurs domaines de recherche futurs s'avèrent intéressants, notamment l'intégration de données synthétiques et d'autres nouvelles capacités de l'IA, ainsi que l'étude de leur capacité à rompre la forte substitution entre les recommandations de l'IA spécifiques au contexte et l'effort humain. La maniabilité de notre modèle rend possible toute une série d'autres applications et extensions théoriques, par exemple l'examen de différents types d'efforts du côté des humains et les implications de différentes technologies d'IA qui fournissent des combinaisons variables de recommandations générales et spécifiques au contexte.

Notre cadre fournit également des lignes directrices sur les différents types d'effets qui doivent être mesurés empiriquement pour évaluer les impacts globaux des nouvelles avancées en IA sur le bien-être.

Références

Daron Acemoglu et Pascual Restrepo. La course entre l'homme et la machine : implications de la technologie sur la croissance, la répartition des facteurs de production et l'emploi. *American Economic Review*, 108(6) :

- 1488–1542, 2018.
- Daron Acemoglu et Pascual Restrepo. Automatisation et nouvelles tâches : comment la technologie remplace et réintègre la main-d'œuvre. *Journal of Economic Perspectives*, 33(2):3–30, 2019.
- Daron Acemoglu, Munther A Dahleh, Ilan Lobel et Asuman Ozdaglar. L'apprentissage bayésien dans les réseaux sociaux. *The Review of Economic Studies*, 78(4):1201–1236, 2011.
- Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian et Asu Ozdaglar. Trop de données : prix et inefficacités sur les marchés des données. *American Economic Journal: Microeconomics*, 14(4):218–256, 2022.
- Daron Acemoglu, Asuman Ozdaglar et James Siderius. Un modèle de désinformation en ligne. *Review of Economic Studies*, 91(6):3117–3150, 2024.
- Nikhil Agarwal, Alex Moehring, Pranav Rajpurkar et Tobias Salz. Combiner l'expertise humaine et l'intelligence artificielle : données expérimentales issues de la radiologie. Rapport technique, National Bureau of Economic Research, 2023.
- Nikhil Agarwal, Alex Moehring et Alexander Wolitzky. Concevoir la collaboration entre l'humain et l'IA : une approche par statistiques suffisantes. Rapport technique, Bureau national de recherche économique, 2025.
- Ajay K Agrawal, Joshua S Gans et Avi Goldfarb. La transformation de Turing : intelligence artificielle, augmentation de l'intelligence et primes de compétence. Rapport technique, Bureau national de recherche économique, 2023.
- Kenneth J Arrow. Les implications économiques de l'apprentissage par la pratique. *The Review of Economic Studies*, 29(3):155–173, 1962a.
- Kenneth Joseph Arrow. Bien-être économique et allocation des ressources pour l'invention. Dans *Readings in industrial economics : Volume deux : Entreprise privée et intervention de l'État*, pages 219–236. Springer, 1962b.
- Abhijit V Banerjee. Un modèle simple du comportement grégaire. *The Quarterly Journal of Economics*, 107(3) : 797–817, 1992.
- Dirk Bergemann, Alessandro Bonatti et Tan Gan. L'économie des données sociales. *The RAND Journal of Economics*, 53(2):263–296, 2022.
- Sushil Bikhchandani, David Hirshleifer et Ivo Welch. Une théorie des modes, de la mode, des coutumes et des changements culturels sous forme de cascades informationnelles. *Journal of Political Economy*, 100(5) : 992-1026, 1992.
- Patrick Bolton et Christopher Harris. Expérimentation stratégique. *Econometrica*, 67(2):349–374, 1999.

- Tilman Börschers, Angel Hernando-Veciana et Daniel Krämer. Quand les signaux sont-ils complémentaires ou substituables ? *Journal of Economic Theory*, 148(1):165–195, 2013.
- Benjamin Brooks, Alexander Frankel et Emir Kamenica. Comparaisons de signaux. *American Economic Review*, 114(9):2981–3006, 2024.
- Erik Brynjolfsson, Danielle Li et Lindsey Raymond. L'IA générative au travail. *The Quarterly Journal of Economics*, 140(2):889–942, 2025.
- Herman Budiyo, M. Pudjaningsih, B. Prastio et A. Maulidina. Étude de l'impact à long terme des outils d'écriture basés sur l'IA sur les compétences en écriture autonome : une étude de cas portant sur des étudiants en formation pédagogique en langue indonésienne. *International Journal of Information and Education Technology*, 15(5) : 1003-1013, 2025.
- Michael Chui, Eric Hazan, Roger Roberts, Alex Singla et Kate Smaje. Le potentiel économique de l'IA générative. 2023.
- Zoe Cullen, Danielle Li et Shengwu Li. Le travail en tant que capital : l'IA et la propriété de l'expertise. Version préliminaire et incomplète. Première version du 12 août 2025 ; cette version du 2 septembre 2025. Manuscrit non publié, 2025. URL <https://danielle.li/assets/docs/LaborAsCapital.pdf>.
- Krishna Dasaratha et Kevin He. Efficacité agrégative de l'apprentissage bayésien dans les réseaux. *Prépublication arXiv arXiv:1911.10116*, 2019.
- Krishna Dasaratha et Kevin He. Apprendre à partir de contenus viraux. *Prépublication arXiv arXiv:2210.01267*, 2022.
- R. Maria del Rio-Chanona, Nadzeya Laurentsyeva et Johannes Wachs. Les grands modèles linguistiques limitent le partage public des connaissances sur les plateformes de questions-réponses en ligne. *PNAS Nexus*, 3(9):p. 400, 2024.
- Kefan Dong et Tengyu Ma. Stp : démonstrateurs de théorèmes LLM en auto-jeu avec conjecture et démonstration itératives. *Prépublication arXiv arXiv:2502.00212*, 2025.
- Maryam Farboodi, Andrew Koh et Bryant Xia. Automatisation basée sur les données, 2025.
- Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Moham-madamin Barekatin, Alexander Novikov, Francisco J R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. Découverte d'algorithmes de multiplication matricielle plus rapides grâce à l'apprentissage par renforcement. *Nature*, 610(7930):47–53, 2022.
- Michael Gerlich. Les outils d'IA dans la société : impacts sur le déchargement cognitif et l'avenir de la pensée critique. *Societies*, 15(1):6, 2025.
- Gillian K Hadfield et Andrew Koh. Une économie d'agents IA. *Prépublication arXiv arXiv:2509.01063*, 2025.

- Wayne Holmes, Fengchun Miao, et al. *Lignes directrices pour l'IA générative dans l'éducation et la recherche*. Éditions de l'Unesco, 2023.
- Enrique Ide. Automatisation, IA et transmission intergénérationnelle des connaissances. *Prépublication arXiv arXiv:2507.16078*, 2025.
- Enrique Ide et Eduard Talamás. L'intelligence artificielle dans l'économie de la connaissance. *Journal of Political Economy*, 133(12):3762–3800, 2025.
- Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson et Mor Naaman. La co-rédaction avec des modèles linguistiques ayant des opinions influence les opinions des utilisateurs. Dans *les Actes de la conférence CHI 2023 sur les facteurs humains dans les systèmes informatiques*, pages 1–15, 2023.
- John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, et al. Prédiction hautement précise de la structure des protéines avec alphafold. *Nature*, 596(7873):583–589, 2021.
- Feiyang Kang, Newsha Ardalani, Michael Kuchnik, Youssef Emad, Mostafa Elhoushi, Shubhabrata Sengupta, Shang-Wen Li, Ramya Raghavendra, Ruoxi Jia et Carole-Jean Wu. Démystifier les données synthétiques dans le pré-entraînement des grands modèles de langage : une étude systématique des lois d'échelle, des avantages et des pièges. Dans *les Actes de la Conférence 2025 sur les méthodes empiriques en traitement du langage naturel*, pages 10750–10769, 2025.
- Godfrey Keller et Sven Rady. Expérimentation stratégique avec les bandits de Poisson. *Theoretical Economics*, 5(2):275–311, 2010.
- Godfrey Keller, Sven Rady et Martin Cripps. Expérimentation stratégique avec les bandits exponentiels. *Econometrica*, 73(1):39–68, 2005.
- Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein et Pattie Maes. Votre cerveau face à ChatGPT : accumulation de dette cognitive lors de l'utilisation d'un assistant IA pour la rédaction d'un essai. *Prépublication arXiv arXiv:2506.08872*, 2025.
- Yong Lin, Shange Tang, Bohan Lyu, Jiayun Wu, Hongzhou Lin, Kaiyu Yang, Jia Li, Mengzhou Xia, Danqi Chen, Sanjeev Arora, et al. Goedel-prover : un modèle de pointe pour la démonstration automatisée de théorèmes en open source. *Prépublication arXiv arXiv:2502.07640*, 2025.
- Liang Lyu, James Siderius, Hannah Li, Daron Acemoglu, Daniel Huttenlocher et Asuman Ozdaglar. Contributions à Wikipédia à la suite de ChatGPT. Dans *les actes complémentaires de la conférence ACM on Web 2025*, pages 1176–1179, 2025.
- Pratyush Maini, Skyler Seto, Richard Bai, David Grangier, Yizhe Zhang et Navdeep Jaitly. Reformuler le Web : une recette pour une modélisation linguistique efficace en termes de calcul et de données. Dans *les actes de la 62e réunion annuelle de l'Association for Computational Linguistics (Volume 1 : Articles longs)*, pages 14044–14072, 2024.

- Nestor Maslej, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, et al. Rapport sur l'indice de l'intelligence artificielle 2025. *Prépublication arXiv arXiv:2504.07139*, 2025.
- Richard R Nelson. L'économie simple de la recherche scientifique fondamentale. *Journal of political economy*, 67(3):297–306, 1959.
- OpenAI. GPT-4, mars 2023. URL <https://openai.com/index/gpt-4-research/>. Consulté en janvier 2026.
- Ilya Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson et Yarin Gal. Les modèles d'IA s'effondrent lorsqu'ils sont entraînés sur des données générées de manière réursive. *Nature*, 631(8022):755–759, 2024.
- David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Maîtriser le jeu de go sans connaissances humaines. *Nature*, 550(7676):354–359, 2017.
- Lones Smith et Peter Sørensen. Conséquences pathologiques de l'apprentissage par observation. *Econometrica*, 68(2):371–398, 2000.
- Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. Conception de novo de la structure et de la fonction des protéines avec rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.
- Huajian Xin, Daya Guo, Zhihong Shao, Zhizhou Ren, Qihao Zhu, Bo Liu, Chong Ruan, Wenda Li et Xiaodan Liang. Deepseek-prover : faire progresser la démonstration de théorèmes dans les LLM grâce à des données synthétiques à grande échelle. *Prépublication arXiv arXiv:2405.14333*, 2024.

Annexe A

Annexe principale

A.1 Preuve du lemme 1

Fixons $\tau_A \geq 0$ et posons $e(X) = e(X, \tau_A)$, où $e(X, \tau_A)$ est l'effort de meilleure réponse (unique) qui résout

$$e(X, \tau_A) := \{ e \geq 0 \mid \Delta_X G(X) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) = e^{\alpha-1} \}.$$

Définiss
ons

$$\Psi(e, X) := \Delta_X G(X) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) - e^{\alpha-1}$$

Pour tout e et $X > 0$, on a $\partial_e \Psi(e, X) < 0$ et $\partial_X \Psi(e, X) > 0$; d'après le théorème de la fonction implicite, $e(X)$ est donc continue et strictement croissante sur $(0, \infty)$. Lorsque X tend vers 0, $G(X)$ tend vers 0, ce qui implique que $e(X)$ tend vers 0; ainsi, $e(X)$ est continue sur $[0, \infty)$.

Par conséquent, $F(X) = h\Sigma^2 + X + \lambda_G e(X)^{-1}$ est également continue et strictement croissante sur $[0, \infty)$, avec $F(0) = 0$ et $\lim_{X \rightarrow +\infty} F(X) = \Sigma^{-2}$.

A.2 Preuve de la proposition 1

Considérons une période fixe t et posons (X_t, τ_A) comme étant données. Nous avons montré que l'utilité attendue d'un agent i est strictement concave par rapport à l'effort $e_{i,t}$, de sorte qu'il existe une réponse optimale unique $e(X_t, \tau_A)$; les prédictions optimales sont des moyennes a posteriori. Dans un équilibre symétrique, $e_{i,t} = e_t = e(X_t, \tau_A)$ pour tout i , de sorte que l'effort global sur un

La fonction d'état est $E_t = \int e_{i,t} di = I e_t$. Étant donné E_t , le signal public a une précision de $\lambda_G E_t$ et la mise à jour de X_t est $X_{t+1} = (X_t + \lambda_G E_t)^{-1} + \Sigma^2$, ce qui revient à dire que $X_{t+1} = \Sigma^2 + I e_t)^{-1}$. Puisque $e(X_t, \tau_A)$ est à valeur unique, la récursion détermine un chemin d'équilibre symétrique unique.

A.3 Preuve de la proposition 2

Fixons $X > 0$ et définissons

$$u := X + \lambda_G I e(X, \tau_A),$$

de sorte que F s'écrit $F = \Sigma^2 + u^{-1}$. Il est évident que F est (strictement) croissante en u sur $\mathbb{R}_{\geq 0}$.

Puisque $X > 0$, on a $e(X, \tau_A) > 0$, de sorte que u est strictement croissante sur I . D'après l'observation 2, $e(X, \tau_A)$ est strictement décroissante en τ_A , donc u est également strictement décroissante en τ_A . En rassemblant ces résultats, on en conclut que F est strictement croissante sur I et strictement décroissante en τ_A .

A.4 Démonstration des propositions 3 et 5

Dans le corps du texte, nous définissons la précision sociale en régime permanent X^- comme un point fixe de la fonction à une étape F , et l'effort correspondant en régime permanent comme la meilleure réponse $e^- = e(X^-, \tau_A)$. Ici, nous fournissons d'abord

une caractérisation pratique « axée sur l'effort » : nous dérivons une équation qui détermine directement e^- .

En réorganisant les termes de la condition de point fixe $F(X^-) = X^-$, on obtient la fonction de liaison entre l'effort global et la précision publique en régime permanent : $(X^{-1} - \Sigma^2)^{-1} - X^- = \lambda_G \cdot (Ie^-)$.

En résolvant pour X^- , on obtient $X^- = W(Ie^-)$, où W est donné par la racine (positive) de l'équation du second degré qui en résulte :

$$W(E) := \frac{\lambda_G E}{2} \left(1 + \sqrt{1 + \frac{4}{\Sigma^2 \cdot \lambda_G E}} \right) - 1.$$

En substituant $X^- = W(Ie^-)$ dans la condition du premier ordre pour l'effort d'équilibre, on obtient l'équation que l'effort en régime permanent doit satisfaire : $D(e^-) = 0$, où¹¹

$$D(e) := \underbrace{\Delta_X \cdot G_W(Ie) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) - e^{\alpha-1}}_{=: MB(e)} \cdot \underbrace{X}_{=: MC(e)}. \quad (A-1)$$

Ici, $G_W(\cdot)$ est le lien implicite entre l'effort global et la probabilité de « réussite » en régime permanent pour la tâche publique :

$$G_W(E) := G(W(E)). \quad (A-2)$$

Nous étudierons les états stationnaires à l'aide du signe et des racines de $D(e)$, puis nous reviendrons à la précision publique en régime permanent en utilisant $X^- = W(Ie^-)$.

Le lemme de liaison des signes suivant est utile pour étudier la dynamique globale.

Lemme A-1. Soit W la fonction de liaison effort-précision, et soit $D(e)$ donné par l'équation (A-1). Pour tout $e \geq 0$, on définit $X(e) := W(Ie)$. Alors

$$\text{sign } D(e) = \text{sign}(F(X(e)) - X(e)),$$

où $\text{sign}(\cdot)$ désigne la fonction de signe habituelle.

Preuve. Pour toute précision publique $X \geq 0$, définissons une fonction auxiliaire D_X :

$$D_X(e) := \Delta_X G(X) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) - e^{\alpha-1}, \quad e \geq 0.$$

C'est exactement la condition du premier ordre pour l'effort optimal de l'agent à la précision publique X . Par définition, $e(X)$ est la racine unique de D_X : $D_X(e(X)) = 0$. Puisque $D_X(e)$ est strictement décroissante en e , passe par zéro exactement une fois en $e = e(X)$, donc

$$\text{sign}(D_X(e)) = \text{sign}(e(X) - e) \quad \text{pour tout } e \geq 0. \quad (A-3)$$

¹¹ Lorsque l'on examine la statique comparative par rapport à τ_A et I , on écrit parfois $D(e)$ sous la forme $D(e; \tau_A, I)$ afin de mettre en évidence sa dépendance vis-à-vis de τ_A et I .

En particulier, l'application de (A-3) avec $X = X(e)$ donne

$$\text{sign}(D(e)) = \text{sign}(e(X(e)) - e) \quad \text{pour tout } e \geq 0. \quad (\text{A-4})$$

Définissons, pour un effort arbitraire e ,

$$T(X, e) := [\Sigma^2 + (X + \lambda_G Ie)^{-1}]^{-1}, \quad X \geq 0, e \geq 0.$$

La fonction dynamique effective est $F(X) = T(X, e(X))$. On a donc

$$F(X(e)) - X(e) = T(X(e), e(X(e))) - T(X(e), e). \quad (\text{A-5})$$

Étant donné que $T(X(e), \cdot)$ est strictement croissante en effort, le signe du membre de droite de (A-5) coïncide avec celui de $e(X(e)) - e$, c'est-à-dire

$$\text{sign}(F(X(e)) - X(e)) = \text{sign}(e(X(e)) - e). \quad (\text{A-6})$$

En comparant (A-4) et (A-6), on obtient pour tout $e > 0$,

$$\text{signe } D(e) = \text{signe}(e(X(e)) - e) = \text{signe}(F(X(e)) - X(e)) = \text{signe}(F(W(Ie)) - W(Ie)).$$

□

Posons $x := \log e$ afin de pouvoir réécrire (A-1) dans l'espace logarithmique :

$$D(x) := \underbrace{\Delta_X \cdot G_W(I \exp(x)) \cdot \lambda_I g(\sigma^{-2} + \lambda_I \exp(x) + \tau_A)}_{\text{MB}(x)} - \underbrace{\exp((\alpha - 1)x)}_{M \cdot C(x)}.$$

Chaque racine réelle de $D(x)$ correspond à un effort positif en régime permanent. Le lemme suivant est la clé de la caractérisation des racines de $D(x)$.

Lemme A-2. $\log(\text{MB}(x))$ est une fonction strictement concave de x . *Démonstration.*

On a la décomposition suivante :

$$\log(\text{MB}(x)) = \underbrace{\log(\lambda_I \Delta_X)}_{=: t_1(x)} + \underbrace{\log G_W(Ie^x) + \log g(\sigma^{-2} + \lambda_I e^x + \tau_A)}_{=: t_2(x)}$$

Nous montrons que $t_1(x)$ et $t_2(x)$ sont toutes deux strictement concaves. Pour $t_1(x)$, nous avons

$$t_1(x) = \frac{G'_W(Ie^x)Ie^x}{G_W(Ie^x)} = R(Ie^x),$$

qui est strictement décroissante en x d'après le lemme A-4.

Pour $t_2(x)$, on a

$$t_2(x) = \frac{g'(\sigma^{-2} + \lambda_1 e^x + \tau_A)}{g(\sigma^{-2} + \lambda_1 e^x + \tau_A)} \cdot \lambda_1 e^x$$

$$= -\frac{\lambda_1 e^x}{2} - \frac{\lambda_1 e^x}{2(\sigma^{-2} + \lambda_1 e^x + \tau_A)},$$

qui est également strictement décroissante en x . □

Preuve de la proposition 3. Rappelons que

$$D(e) := \underbrace{\Delta_X \cdot G(W(le)) \cdot \lambda_1 g(\sigma^{-2} + \lambda_1 e + \tau_A) - e^{\alpha-1}}_{MB(e)} \cdot \underbrace{X}_{MC(e)}$$

D'après le lemme A-3, lorsque E tend vers $0^{(+)}$ on a $G(W(E)) = \kappa_1 E^{1/4} + o(E^{1/4})$ pour un certain $\kappa_1 > 0$. En posant $E = le$

et en utilisant la continuité de $g(\cdot)$, on obtient, lorsque $e \rightarrow 0^+$,

$$D(e) = Ce^{1/4} - e^{\alpha-1} + o(e^{1/4}), \quad (A-7)$$

pour un certain $C > 0$. Ainsi, $D(e)$ est strictement positif près de 0. Il est également facile de vérifier que $D(e)$ est strictement négatif pour de grandes valeurs de e .

Ensuite, le lemme A-2 montre que $\log(MB(x))$ est strictement concave en $x = \log e$, tandis que $\log(MC(x)) = (\alpha - 1)x$ est linéaire. De plus, le comportement limite de $D(e)$ évoqué précédemment implique que la courbe MB est plus grande en $x = -\infty$, tandis que la courbe MC est plus grande en $x = +\infty$. Par conséquent, ces deux courbes doivent se croiser *exactement une fois*. Par conséquent, $D(e) = 0$ possède exactement une racine positive, notée $e_h^- > 0$. D'après la structure de concavité, on obtient que $D(e) > 0$ pour $0 < e < e_h^-$ et $D(e) < 0$ pour $e > e_h^-$.

Soit $X_h^- := W(le_h^-) > 0$. En plus de l'état stationnaire trivial $X_e^- = 0$, on obtient exactement deux précisions publiques d'état stationnaire : $0 = X_l^- < X_h^- < \infty$. D'après le lemme A-1, pour tout $e \geq 0$,

$$\text{le signe de } D(e) = \text{le signe de } (F(X(e)) - X(e)), \quad X(e) := W(le).$$

Ainsi

$$F(X) - X \begin{cases} > 0, & 0 < X < X_h^- \\ < 0, & X > X_h^- \end{cases}$$

Considérons le système dynamique $X_{t+1} = F(X_t)$ sur $[0, \infty)$. Le lemme 1 donne que F est continue, strictement croissante, $F(0) = 0$ et $F(+\infty) = \Sigma^{-2}$. Si $X_1 \in (0, X_h^-)$, alors $\{X_t\}$ est croissante et bornée supérieurement par X_h^- , et converge donc vers X_h^- . Le même argument s'applique à $X_1 \in (X_h^-, +\infty)$. Ainsi, toute trajectoire avec $X_1 > 0$ converge vers X_h^- , tandis que $X_1 = 0$ reste à 0. □

Démonstration de la proposition 5. Pour tout $\tau \geq 0$, désignons par $F_\tau(\cdot)$ la fonction de transition d'états $F(\cdot)$ lorsque $\tau_A = \tau$,

et
posons

$$D_{\tau}(e) := \underbrace{\Delta_X \cdot G_W(1e) \cdot \lambda_1 g(\sigma^{-2} + \lambda_1 e + \tau)}_{=MB_{\tau}(e)} - \underbrace{e^{\alpha-1}}_{=MC(e)}. \quad (A-8)$$

Soit l'équation de l'effort en régime permanent (A-1) lorsque $\tau_A = \tau$.

D'après le lemme A-3, lorsque $E \rightarrow 0^{(+)}$ on a $G_W(E) = \kappa_1 E^{1/4} + o(E^{1/4})$ pour un certain $\kappa_1 > 0$. En posant $E = 1e$ et en utilisant la continuité de $g(\cdot)$, on obtient, lorsque $e \rightarrow 0^+$,

$$D_{\tau}(e) = C(\tau) e^{1/4} - e^{\alpha-1} + o(e^{1/4}),$$

pour un certain $C(\tau) > 0$ tel que $C(\tau) < 0$. Puisque $\alpha - 1 < \frac{1}{4}$, on a $D_{\tau}(e)/e^{1/4} \rightarrow -\infty$ lorsque $e \rightarrow 0^+$, uniformément en $\tau \in [0, \infty)$. Il existe donc un $\delta > 0$ (indépendant de τ) tel que

$$D_{\tau}(e) < 0 \quad \text{pour tout } 0 < e \leq \delta. \quad (A-9)$$

Pour un effort important, notons que $0 \leq G_W(1e) \leq G(\Sigma^{-2})$ et que $g(\sigma^{-2} + \lambda_1 e + \tau) \rightarrow 0$ lorsque $e \rightarrow \infty$. Ainsi, le terme de bénéfice marginal dans $D_{\tau}(e)$ est borné, alors que $e^{\alpha-1} \rightarrow \infty$. Il en résulte qu'il existe $R > 0$ (là encore uniforme sur τ) tel que

$$D_{\tau}(e) < 0 \quad \text{pour tout } e \geq R. \quad (A-10)$$

Par conséquent, pour chaque τ fixe, $D_{\tau}(e)$ est continue, strictement négative près de 0 et pour les grands e , et toute racine positive doit appartenir à (δ, R) .

Ensuite, le lemme A-2 montre que $\log(MB_{\tau}(x))$ est strictement concave en $x = \log e$, tandis que $\log(MC(x)) = (\alpha - 1)x$ est linéaire. Par conséquent, ces deux courbes se croisent au plus deux fois, de sorte que $D_{\tau}(e)$ possède au plus deux racines positives. En combinant cela avec (A-9) et (A-10), nous concluons :

- (i) pour chaque τ , D_{τ} possède 0, 1 ou 2 racines positives ;
- (ii) si $0 < e_m(\tau) < e_h(\tau)$ sont les deux racines, alors

$$D_{\tau}(e) < 0 \text{ sur } (0, e_m(\tau)) \cup (e_h(\tau), \infty), \quad D_{\tau}(e) > 0 \text{ sur } (e_m(\tau), e_h(\tau)).$$

Comme $g(\cdot)$ est strictement décroissante, $D_{\tau}(e)$ est strictement décroissante en τ pour tout $e > 0$. Soit

$$S := \{\tau \geq 0 : \text{il existe un } e > 0 \text{ tel que } D_{\tau}(e) = 0\}$$

soit l'ensemble des τ pour lesquels il existe au moins un effort positif en régime permanent. Si $S = \emptyset$, alors il n'y a jamais de régime permanent positif et la proposition 5(ii) est vérifiée avec $\tau^c = 0$; dans ce cas, la partie (i) est vide de sens.

Désormais, nous supposons que $S \neq \emptyset$.

Comme $D_{\tau}(e)$ est (point par point) strictement décroissante en τ pour $e > 0$, l'ensemble S est fermé vers le bas : si $\tau_2 \in S$ et $\tau_1 < \tau_2$, alors $\tau_1 \in S$. Étant donné que, pour *tout* τ suffisamment grand, on a $D_{\tau}(e) < 0$ pour tout $e > \delta$ (par

la borne uniforme $g(\sigma^{-2} + \lambda_1 e + \tau) \leq g(\sigma^{-2} + \tau) \rightarrow 0$, l'ensemble S est borné vers le haut.

Le seuil d'effondrement complet correspond alors à $\tau^c := \sup S < \infty$.

Nous affirmons maintenant que :

- (i) Pour tout $\tau < \tau^c$, D_τ possède exactement deux racines positives $0 < e_m(\tau) < e_h(\tau)$.
- (ii) Pour tout $\tau > \tau^c$, $D_\tau(e) < 0$ pour tout $e > 0$ (c'est-à-dire qu'il n'y a pas de racine positive).
- (iii) si $\tau^c > 0$, alors pour $\tau = \tau^c$, il existe exactement une racine positive $e^- > 0$, et c'est un point de tangence ($D_\tau(e^-) = D'_\tau(e^-) = 0$).

Le point (ii) découle immédiatement de la définition de τ^c en tant que suprême. Pour (i), fixons $\tau < \tau^c$ et choisissons $\tilde{\tau} \in (\tau, \tau^c)$ de sorte que $\tilde{\tau} \in S$. Alors $D_{\tilde{\tau}}$ possède au moins une racine positive $e^* > 0$; d'après la monotonie $D_{\tilde{\tau}}(e^*) > 0$. Conjointement avec (A-9) et (A-10), la continuité implique que D_τ doit passer deux fois par zéro sur $(0, \infty)$, et possède donc exactement deux racines positives $e_m(\tau)$ et $e_h(\tau)$ d'après la propriété « au plus deux racines » ci-dessus.

Pour (iii), prenons une suite $\tau_n \uparrow \tau^c$ telle que $\tau_n \in S$, et posons $e_m(\tau_n)$, $e_h(\tau_n)$ comme les deux racines de D_{τ_n} . Nous avons des bornes uniformes $\delta \leq e_m(\tau_n) \leq e_h(\tau_n) \leq R$, donc après passage à une sous-séquence, nous pouvons supposer que $e_m(\tau_n) \rightarrow e^-_1$ et $e_h(\tau_n) \rightarrow e^-_2$ avec $\delta < e^-_1 \leq e^-_2 \leq R$. Par continuité, e^-_1 et e^-_2 sont toutes deux des racines de D_τ lorsque $\tau = \tau^c$. Si $e^-_1 < e^-_2$, le théorème de la fonction implicite implique que pour tout τ proche de τ^c , il existe deux racines proches distinctes, ce qui contredit la définition de τ^c en tant que suprême de S .

Par conséquent, $e^-_1 = e^-_2 =: e^- > 0$, et D_τ a une racine *double* en e^- lorsque $\tau = \tau^c$.

Les efforts positifs en régime permanent correspondent biunément aux précisions sociales positives en régime permanent via

$X = W(Ie)$. Ainsi, pour chaque $\tau < \tau^c$, il existe exactement trois états stationnaires dans l'espace X :

$$X^-_l(\tau) = 0, \quad X^-_m(\tau) = W(Ie_m(\tau)), \quad X^-_h(\tau) = W(Ie_h(\tau)),$$

avec $0 < X^-_m(\tau) < X^-_h(\tau) < \infty$. Pour $\tau_A > \tau^c$, on n'a qu'un seul état stationnaire $X^-_l(\tau) = 0$. Il en résulte l'existence et le nombre d'états stationnaires énoncés aux parties (i) et (ii).

Fixons $\tau < \tau^c$ et considérons le système dynamique unidimensionnel $X_{t+1} = F_\tau(X_t)$ sur $[0, \infty)$.

Le lemme 1 implique que F_τ est continue et strictement croissante, avec $F_\tau(0) = 0$ et $F_\tau(+\infty) = \Sigma^{-2} < \infty$.

De plus, F possède exactement les trois points fixes $0 < X^-_m(\tau) < X^-_h(\tau)$, et il s'ensuit du lemme A-1 qu

$$\begin{aligned} F_\tau(X) - X &< 0, & X \in (0, X^-_m(\tau)), \\ F_\tau(X) - X &> 0, & X \in (X^-_m(\tau), X^-_h(\tau)), \\ F_\tau(X) - X &< 0, & X \in (X^-_h(\tau), \infty). \end{aligned}$$

Soit maintenant $X_1 \geq 0$, quel qu'il soit. Si $X_1 \in (0, X^-_m(\tau))$, alors $\{X_t\}$ restera dans $(0, X^-_m(\tau))$ et sera strictement décroissante, et converge donc vers un point fixe dans $[0, X^-_m(\tau)]$, nécessairement $X^-_l(\tau) = 0$. Si

si $X_1 \in (X_m^-(\tau), X_h^-(\tau))$, alors la suite est croissante et reste dans $(X_m^-(\tau), X_h^-(\tau))$; on a donc $X_t \rightarrow X_h^-(\tau)$. Des arguments similaires s'appliquent au cas où $X_1 > X_h^-(\tau)$.

Cela prouve la partition du bassin dans la partie (i) :

$$X_1 < X_m^-(\tau) \Rightarrow X_t \rightarrow 0, \quad X_1 > X_h^-(\tau) \Rightarrow X_t \rightarrow X_h^-(\tau).$$

Lorsque $\tau_A > \tau_c$, nous avons démontré que 0 est le seul état d'équilibre. De plus, pour tout $X > 0$, $F_\tau(X) < X$ d'après le lemme A-1. Par conséquent, pour tout $X_1 > 0$, la suite $X_{t+1} = F_\tau(X_t)$ est strictement décroissante et bornée inférieurement par 0, et doit donc converger vers 0. Cela donne la partition du bassin dans la partie (ii). $\square$

A.5 Preuve du lemme 2

D'après le lemme 1, la fonction de transition d'états F est continue, strictement croissante sur $\mathbb{R}_{\geq 0}$, et satisfait $F(0) = 0$. Rappelons que d'après (A-7), on a, pour un certain $C > 0$, $D(e) = Ce^{1/4} - e^{\alpha-1} + o(e^{1/4})$.

Par conséquent, lorsque $\alpha - 1 > \frac{1}{4}$, il existe un $\delta_e > 0$ tel que $D(e) > 0$ pour tout $e \in (0, \delta_e)$. D'après le lemme A-1, il existe $\delta_X > 0$ tel que $F(X) > X$ pour $X \in (0, \delta_X)$, c'est-à-dire que $F(X) > X$ dans un voisinage de 0. En combinant cela avec la propriété ci-dessus de F , on en conclut que l'état stationnaire zéro est localement instable.

De même, lorsque $\alpha - 1 < \frac{1}{4}$, il existe un $\delta_e' > 0$ tel que $D(e) < 0$ pour tout $e \in (0, \delta_e')$. D'après le lemme A-1, il existe $\delta_X' > 0$, tel que $F(X) < X$ pour $X \in (0, \delta_X')$. Dans ce cas, l'état stationnaire zéro est localement stable.

A.6 Lemmes techniques supplémentaires

Lemme A-3. Pour la fonction $G_W(\cdot)$ définie dans (A-2), les propositions suivantes sont vraies :

- $G_W(\cdot)$ est strictement croissante, avec $\lim_{E \rightarrow 0} G_W(E) = 0$, et $\lim_{E \rightarrow +\infty} G_W(E) = G(\Sigma^{-2})$.

- Sous le régime $E \rightarrow 0^+$, on a

$$\begin{aligned} -G_W(E) &= \kappa_1 E^{\frac{1}{4}} + o(E^{\frac{1}{4}}) \\ -G'_W(E) &= \frac{1}{4} \kappa_1 E^{-\frac{3}{4}} + o(E^{-\frac{3}{4}}) \end{aligned}$$

$$\text{où } \kappa_1 = \frac{1}{22} \lambda^2 a^{\frac{1}{4}} \phi(0) > 0.$$

- Sous le régime $E \rightarrow +\infty$,

$$-G_W(E) = G(\Sigma^{-2}) - \frac{\kappa_2}{E} + o\left(\frac{1}{E}\right)$$

$$-G'_W(E) = \frac{\kappa_2}{E^2} + o\left(\frac{1}{E^2}\right)$$

$$\text{où } \kappa_2 = \phi(\Sigma^{-1}) \lambda^{-1} \Sigma_G^{-3} > 0.$$

Démonstration. Étant donné que $G(\cdot)$ et $W(\cdot)$ sont toutes deux strictement croissantes, $G_W(\cdot)$ est également strictement croissante.

Soit

$$a := \frac{4}{\lambda_G \Sigma^2}, \quad S(E) := \frac{1}{1 + \frac{a}{E}}.$$

On a alors que $\frac{2}{\lambda_G} W(E) = E(S(E) - 1)$. Notons que

$$S'(E) = \frac{1}{2} \cdot \frac{1}{1 + \frac{a}{E}} \cdot \frac{a}{E^2} = -\frac{a}{2E^2 S(E)}.$$

Ainsi

$$\frac{2}{\lambda_G} W'(E) = S(E) - 1 - \frac{a}{2ES(E)} = \frac{(S(E) - 1)^2}{2S(E)}. \quad (\text{A-11})$$

Régime $E \rightarrow 0^+$: Nous avons les asymptotiques suivantes :

$$W(E) = \frac{\lambda_G}{2} \cdot E(S(E) - 1) \sim \frac{\lambda_G a^{\frac{1}{2}}}{2} \cdot E^{\frac{1}{2}},$$

$$G_W(E) = G(W(E)) \sim 2 \sqrt{\frac{1}{W(E)\phi(0)}} \sim 2^{\frac{1}{2}} \lambda^{\frac{1}{2}} a^{\frac{1}{4}} \phi(0) \cdot E^{\frac{1}{4}},$$

et

$$W(E) = \frac{\lambda_G}{2} \cdot \frac{(S(E) - 1)^2}{2S(E)} \sim \frac{\lambda_G a^{\frac{1}{2}}}{4} \cdot E^{-\frac{1}{2}}$$

Par conséquent,

$$\begin{aligned} G'_W(E) &= G'(W(E))W'(E) \\ &= \frac{\phi(W(E))}{W(E)} \\ &\sim 2^{-\frac{3}{2}} \lambda^{\frac{1}{2}} a^{\frac{1}{4}} \phi(0) \cdot E^{-\frac{3}{4}} \end{aligned}$$

Régime $E \rightarrow +\infty$: Dans ce cas, nous avons les asymptotiques suivantes :

$$\begin{aligned} -1 &\sim \frac{a S(E)}{2E}, \\ \Sigma^{-2} - W(E) &= \Sigma^{-2} \cdot \left(1 - \frac{2}{S(E) + 1}\right) = \Sigma^{-2} \cdot \frac{S(E) - 1}{S(E) + 1} \sim \Sigma^{-2} \cdot \frac{a}{4E} = \lambda^{-1} \Sigma_G^{-4} \cdot E^{-1}, \end{aligned}$$

$$\begin{aligned} G_W(E) - G(\Sigma^{-2}) &= G(W(E)) - G(\Sigma^{-2}) \\ &= G'(\Sigma^{-2}) \cdot (W(E) - \Sigma^{-2}) + o(W(E) - \Sigma^{-2}) \\ &\sim -G'(\Sigma^{-2}) \lambda^{-1} \Sigma_G^{-4} \cdot E^{-1} \\ &= -\phi(\Sigma^{-1}) \lambda^{-1} \Sigma_G^{-3} \cdot E^{-1}, \end{aligned}$$

et

$$\begin{aligned} W'(E) &= \frac{\lambda_G}{2} \frac{(S(E) - 1)^2}{2S(E)} \\ &= \frac{\lambda_G}{4} \frac{(S(E)^2 - 1)^2}{S(E)(S(E) + 1)^2} \\ &\sim \frac{\lambda_G}{4} E^{-2}. \end{aligned}$$

Par conséquent,

$$\begin{aligned} G'_W(E) &= \frac{\phi(\sqrt{W(E)})}{\phi(\frac{W(E)}{\Sigma})} W'(E) \\ &\sim \frac{\sqrt{\lambda_G \Sigma}}{\Sigma^{-2}} E^{-2} \\ &= \phi(\Sigma^{-1}) \lambda_G^{-1} \Sigma^{-3} E^{-2}. \end{aligned}$$

□

Lemme A-4. Soit

$$R(E) := \frac{d \log G_W(E)}{d \log E} = \frac{E G'_W(E)}{G_W(E)}.$$

Nous avons que $R(E)$ est strictement décroissante sur $(0, +\infty)$, avec $\lim_{E \rightarrow 0^+} R(E) = \frac{1}{4}$.

Démonstration. Rappelons que $G(\tau) = 2\Phi(\sqrt{\tau}) - 1$, et $W(E) = \frac{\lambda_G E^2}{1 + \frac{4}{\lambda_G \Sigma^2 E^2}} - 1$.

Nous avons la factorisation suivante de $R(E)$:

$$R(E) = \frac{E G'(W(E)) W'(E)}{G(W(E))} = \underbrace{\frac{E W'(E)}{W(E)}}_{L(E)} \cdot \underbrace{\frac{W(E) G'(W(E))}{G(W(E))}}_{Q(W(E))},$$

de sorte que $R(E) = L(E) \cdot Q(W(E))$. Pour démontrer que $R(E)$ est strictement décroissante, il suffit de prouver que $L(\cdot)$ et $Q(\cdot)$ sont strictement décroissantes.

En suivant la même notation que dans la preuve du lemme A-3, on a

$$L(E) = \frac{E W'(E)}{W(E)} = E \cdot \frac{S(E) - 1}{2S(E)E} = \frac{S(E) - 1}{2S(E)},$$

où $a := \frac{4}{\lambda_G \Sigma^2}$ et $S(E) := \sqrt{1 + a \frac{1}{E^2}}$. Puisque $S(E)$ est strictement décroissante en E , $L(E)$ l'est également.

Nous prouvons ensuite que $Q(\tau) = \frac{\tau G'(\tau)}{G(\tau)}$ est strictement décroissante. Notons que

$$Q(\tau) = \frac{\sqrt{-\tau} \cdot \phi(\sqrt{\tau})}{2\Phi(\sqrt{\tau}) - 1} = \frac{z \cdot \phi(z)}{2\Phi(z) - 1} =: P(z),$$

où $z = \sqrt{\tau}$. Il suffit de prouver que P est strictement décroissante.

Notons que

$$\begin{aligned} P'(z) &= \frac{(z\phi'(z) + \phi(z))(2\Phi(z) - 1) - 2\phi(z) \cdot z\phi(z)}{(2\Phi(z) - 1)^2} \\ &= \frac{(-z^2\phi(z) + \phi(z))(2\Phi(z) - 1) - 2\phi(z) \cdot z\phi(z)}{(2\Phi(z) - 1)^2} \\ &= \frac{\phi(z)}{(2\Phi(z) - 1)^2} \cdot (1 - z^2)(2\Phi(z) - 1) - 2z\phi(z). \end{aligned}$$

De plus, pour $z > 0$, on a

$$\begin{aligned} \frac{d(1 - z^2)(2\Phi(z) - 1) - 2z\phi(z)}{dz} &= -2z(2\Phi(z) - 1) + (1 - z^2)2\phi(z) - 2\phi(z) + 2z^2\phi(z) \\ &= -2z(2\Phi(z) - 1) < 0, \end{aligned}$$

ce qui implique que pour $z \in (0, +\infty)$,

$$(1 - z^2)(2\Phi(z) - 1) - 2z\phi(z) < (1 - z^2)(2\Phi(z) - 1) - 2z\phi(z) \Big|_{z=0} = 0.$$

On a donc que, pour $z \in (0, +\infty)$, $P'(z) < 0$. □

A.7 Preuve de la proposition 4

Nous divisons la preuve en plusieurs parties :

1. e_h^- dans / et τ_A .

D'après la proposition 3, $D(e; \tau_A, I)$ possède une racine positive unique $e_h^- > 0$ telle que $\partial_e D(e_h^-; \tau_A, I) < 0$. De plus, pour tout $e > 0$, on a $\partial_\tau D(e; \tau_A, I) < 0$ et $\partial_I D(e; \tau_A, I) > 0$. Ainsi, le théorème de la fonction implicite implique que e_h^- est C^1 dans (τ_A, I) , et

$$\frac{\partial e_h^-}{\partial \tau_A} = -\frac{\partial_\tau D(e_h^-; \tau_A, I)}{\partial_e D(e_h^-; \tau_A, I)}, \quad \frac{\partial e_h^-}{\partial I} = -\frac{\partial_I D(e_h^-; \tau_A, I)}{\partial_e D(e_h^-; \tau_A, I)}.$$

Par conséquent, nous avons $\frac{\partial e_h^-}{\partial I} > 0$ et $\frac{\partial e_h^-}{\partial \tau_A} < 0$.

2. Y_h^- dans / et τ_A .

Pour $Y_h^- = \sigma^{-2} + \lambda \bar{e}_h^- + \tau_A$, on a $\frac{\partial Y_h^-}{\partial I} = \lambda \cdot \frac{\partial e_h^-}{\partial I} > 0$.

La statique comparative de Y_h^- par rapport à τ_A nécessite une analyse plus fine. Nous démontrons d'abord un lemme.

Lemme A-5. *On a :*

$$-\frac{\partial e_h^-}{\partial \tau_A} = \frac{\lambda(1 + \frac{1}{Y_h^-}) e_h^-}{\lambda(1 + \frac{1}{Y_h^-}) e_h^- + 2(\alpha - 1 - R(l e_h^-))} \cdot \quad (A-12)$$

$$\text{où } R(E) := \frac{EG'_W(E)}{G_W(E)}$$

Démonstration. Des calculs directs montrent que

$$\frac{\partial D(e_h^-; \tau_A, l)}{\partial e} = A - B - C, \quad \frac{\partial D(e_h^-; \tau_A, l)}{\partial \tau_A} = -\frac{1}{\lambda_I} B$$

où A, B, C représentent

$$A := \Delta_X \cdot l G'_W(l e_h^-) \cdot \lambda_I g \sigma^{-2} + \lambda_I e_h^- + \tau_A$$

$$B := -\Delta_X G_W(l e_h^-) \cdot \lambda^2 g' \sigma^{-2} + \lambda_I e_h^- + \tau_A$$

$$C := (\alpha - 1)(e_h^-)^{\alpha-2}$$

Par conséquent, on a

$$\frac{\partial e_h^-}{\partial \tau_A} = -\frac{\frac{\partial}{\partial \tau_A} D(e_h^-; \tau_A, l)}{\frac{\partial}{\partial e} D(e_h^-; \tau_A, l)} = -\frac{B}{\lambda_I (B + C - A)}$$

De plus, les quantités peuvent être simplifiées comme suit :

$$\begin{aligned} A &= \Delta_X \cdot l G'_W(l e_h^-) \cdot \lambda_I g Y_h^- \\ &= \Delta_X \cdot G_W(l e_h^-) \cdot \lambda_I g Y_h^- \cdot R(l e_h^-) \cdot (e_h^-)^{-1} \\ &= (e_h^-)^{\alpha-1} \cdot R(l e_h^-) \cdot (e_h^-)^{-1} \\ &= R(l e_h^-) \cdot (e_h^-)^{\alpha-2} \\ B &= -\Delta_X G_W(l e_h^-) \cdot \lambda^2 g' Y_h^- \\ &= \Delta_X G_W(l e_h^-) \cdot \lambda g Y_h^- \cdot \frac{\lambda_I}{2} (1 + \frac{1}{Y_h^-}) \\ &= \frac{\lambda_I}{2} (1 + \frac{1}{Y_h^-}) \cdot (e_h^-)^{\alpha-1} \end{aligned}$$

On a donc

$$\begin{aligned} \frac{\partial e_h^-}{\partial \tau_A} &= -\frac{B}{\lambda_I (B + C - A)} \\ &= -\frac{\frac{\lambda_I}{2} (1 + \frac{1}{Y_h^-}) \cdot (e_h^-)^{\alpha-1}}{\lambda_I \left(\frac{\lambda_I}{2} (1 + \frac{1}{Y_h^-}) \cdot (e_h^-)^{\alpha-1} + (\alpha - 1)(e_h^-)^{\alpha-2} - R(l e_h^-) \cdot (e_h^-)^{\alpha-2} \right)} \\ &= -\frac{(1 + \frac{1}{Y_h^-}) \cdot e_h^-}{\lambda_I (1 + \frac{1}{Y_h^-}) \cdot e_h^- + 2(\alpha - 1 - R(l e_h^-))} \end{aligned}$$

comme souhaité. □

Le lemme A-4 montre que $R(E)$ est strictement décroissante avec $\lim_{E \downarrow 0} R(E) = \frac{1}{4}$, donc pour tout $E > 0$, $R(E) \leq$

$\frac{1}{4} < \alpha - 1$. Par conséquent, dans (A-12), le dénominateur est strictement supérieur à $\lambda_l (1 + 1/Y_h^-) e_h^-$, ce qui implique $0 < -\frac{\partial e_h^-}{\partial \tau_A} < \frac{\lambda_l}{4}$. Par conséquent, $\frac{\partial \tau_A^-}{\partial Y_h^-} = 1 + \lambda_l \frac{\partial e_h^-}{\partial \tau_A} > 0$, de sorte que Y_h^- augmente également en τ_A .

3. X_h^- dans l et τ_A .

Puisque $X_h^- = W'(le_h^-)$, on a

$$\frac{\partial \bar{X}_h}{\partial \tau_A} = W'(le_h^-) l \frac{\partial e_h^-}{\partial \tau_A} < 0, \quad \frac{\partial \bar{X}_h}{\partial l} = W'(le_h^-) e_h^- + l \frac{\partial e_h^-}{\partial l} > 0,$$

comme souhaité.

Annexe B

Annexe en ligne pour l'IA, la cognition humaine et le effondrement des connaissances

B.1 Démonstration de la proposition 6

Nous démontrons d'abord la partie (i). Rappelons que dans la démonstration de la proposition 5, nous avons déduit la formule du seuil de « complete-collapse » comme suit (ici, nous définissons $\sup \emptyset = 0$)

$$\tau_A^c(I) = \sup S(I), \quad S(I) := \{\tau \geq 0 : \exists e > 0 \text{ tel que } D(e; \tau_A, I) = 0\}.$$

Ici, nous écrivons τ_A^c sous la forme $\tau^c(I)$ pour souligner sa dépendance vis-à-vis de I .

Si $\tau_A^c(I) > 0$, il y a tangence au seuil :

$$\exists e^- > 0: \quad D(e^-; \tau_A^c(I), I) = 0, \quad \partial_e D(e^-; \tau_A^c(I), I) = 0. \quad (B-1)$$

On notera que $D(e; \tau_A, I)$ est strictement croissante en I point par point dans (τ_A, e) . Soit $I_2 > I_1$. Si $\tau \in S(I_1)$, il existe un $e^- > 0$ tel que $D(e^-; \tau, I_1) = 0$, d'où $D(e^-; \tau, I_2) > 0$. Rappelons que nous avons montré que $D(\cdot; \tau, I_2)$ est continue, strictement négative près de 0 et de $+\infty$. Ainsi, $D(\cdot; \tau, I_2)$ doit avoir deux racines positives. Par conséquent, $\tau \in S(I_2)$, ce qui implique que $S(I_1) \subseteq S(I_2)$ et

$$\tau_A^c(I_2) = \sup S(I_2) \geq \sup S(I_1) = \tau^c(I_1), \text{ donc } \tau^c(I)$$

est faiblement croissante en I .

Si $\tau_A^c(I_1) > 0$, alors d'après (B-1), il existe $e^- > 0$ tel que $D(e^-; \tau^c(I_1), I_1) = 0$. Monotonie stricte

Le fait que $D(e^-; \tau^c(I_1), I_2)$ soit à nouveau dans I implique que $D(\cdot; \tau^c(I_1), I_2) > 0$. Comme $D(\cdot; \tau^c(I_1), I_2)$ est continue et négative près de

0 et $+\infty$, une petite augmentation de τ au-dessus de $\tau^c(I_1)$ et une petite perturbation de e^- rétablissent une racine ; il y a donc

il existe un $\tau' > \tau^c(I_1)$ tel que $D(e^-; \tau', I_2) = 0$ pour un certain $e^- > 0$. Ainsi, $\tau^c(I_2) \geq \tau' > \tau^c(I_1)$, ce qui prouve que

monotonie dès que $\tau^c(I_1) \neq 0$.

La démonstration de la partie (ii) (celle concernant Δ_X) est identique ; il suffit de noter que $D(e; \tau_A, I)$ est également strictement croissante en Δ_X point par point sur (τ_A, e) .

B.2 Preuve des propositions 7 et 8

Nous divisons la preuve en plusieurs parties :

1. e_m^-, e_h^- dans I et τ_A .

Pour tout $\alpha \in [0, 1]$ et tout $\tau_A < \tau^c$, la proposition 5 implique que $D(e; \tau_A, I)$ possède exactement deux éléments positifs

racines $0 < e_m^- < e_h^-$ avec

$$D_e(e_m^-; \tau_A, I) > 0, \quad D_e(e_h^-; \tau_A, I) < 0.$$

Le théorème de la fonction implicite implique que $\bar{e}_k$ ($k \in \{m, h\}$) est une fonction C^1 dans (τ_A, l) et

$$\frac{\partial \bar{e}_k}{\partial \tau} = - \frac{\partial_\tau D(\bar{e}_k; \tau_A, l)}{\partial_e D(\bar{e}_k; \tau_A, l)}, \quad \frac{\partial \bar{e}_k}{\partial l} = - \frac{\partial_l D(\bar{e}_k; \tau_A, l)}{\partial_e D(\bar{e}_k; \tau_A, l)},$$

ce qui implique

$$\frac{\partial \bar{e}_h}{\partial \tau_A} < 0, \quad \frac{\partial \bar{e}_m}{\partial \tau_A} > 0, \quad \frac{\partial \bar{e}_h}{\partial l} > 0, \quad \frac{\partial \bar{e}_m}{\partial l} < 0.$$

Ainsi, l'effort en régime permanent élevé diminue en τ et augmente en l , tandis que la racine intermédiaire évolue dans la direction opposée.

2. X_m^- , X_h^- en l et τ_A .

Par définition,

$$X_k^- = W(k_k^-), \quad k \in \{m, h\},$$

et W est strictement croissante. Puisque $\frac{\partial \bar{e}_h}{\partial \tau_A} < 0$ et $\frac{\partial \bar{e}_m}{\partial \tau_A} > 0$, on a

$$\frac{\partial X_h^-}{\partial \tau_A} < 0, \quad \frac{\partial X_m^-}{\partial \tau_A} > 0.$$

Cela prouve la statique comparative τ_A de X_h^- et X_m^- .

Pour la statique comparative l de X_h^- et X_m^- , il est pratique de travailler directement avec la loi de mouvement $X_{t+1} = F(X_t; l, \tau_A)$ directement. Pour $\tau_A < \tau_A^{(e)}$ il existe trois états stationnaires

$$0 = X_l^- < X_m^- < X_h^-,$$

et le signe de $F(X; l, \tau_A) - X$ est

$$F(X; l, \tau_A) - X \begin{cases} < 0, & X \in (0, X_m^-), \\ > 0, & X \in (X_m^-, X_h^-), \\ = 0, & X \in (X_h^-, \infty). \end{cases}$$

De plus, la proposition 2 stipule que $F(X; l, \tau_A)$ est strictement croissante en fonction de l pour tout $X > 0$.

Fixons τ_A et prenons $l_2 > l_1$. Écrivons $F_j(X) = F(X; l_j, \tau_A)$ et $X_{m,j}^- = X_m^-(l_j, \tau_A)$, $j = 1, 2$. Pour $X = X_{m,(1)}^-$, on a $F_1(X_{m,(1)}^-) = X_{m,(1)}^-$, donc $F_2(X_{m,(1)}^-) - X_{m,(1)}^- > 0$. D'autre part, pour

Pour tout $X > 0$ suffisamment petit, on a $F_2(X) - X < 0$. Par continuité, $F_2(X) - X$ passe par zéro pour un certain X positif minimal strictement inférieur à $X_{m,1}^-$; cette racine est précisément $X_{m,2}^-$. Par conséquent, on a $X_m^-(l_2, \tau) < X_m^-(l_1, \tau)$, ce qui montre que X_m^- est strictement décroissante sur l . De même, X_h^- est strictement croissante sur l .

3. Y_h^- sur l .

En régime permanent, on a $Y_h^- = \sigma^{-2} + \lambda_1 e_h^- + \tau_A$, donc

$$\frac{\partial Y_h^-}{\partial I} = \lambda_1 \frac{\partial e_h^-}{\partial I} > 0.$$

4. Y_h^- dans τ_A .

Le lemme A-5 s'applique ici à l'équilibre supérieur : on a

$$\frac{\partial e_h^-}{\partial \tau} = - \frac{(1 + 1/Y_h^-) e_h^-}{\lambda_1 (1 + 1/Y_h^-) e_h^- + 2(\alpha - 1 - R(le_h^-))}.$$

où le numérateur et le dénominateur sont tous deux strictement positifs.

Par conséquent

$$\frac{\partial Y_h^-}{\partial \tau_A} = 1 + \lambda_1 \frac{\partial e_h^-}{\partial \tau} = \frac{2(\alpha - 1 - R(le_h^-))}{\lambda_1 (1 + 1/Y_h^-) e_h^- + 2(\alpha - 1 - R(le_h^-))}.$$

Le dénominateur est positif, donc le signe $\frac{d(e_h^-)}{d\tau_A}$ est égal au signe de $(\alpha - 1 - R(le_h^-))$.

Le lemme A-4 montre que $R(E)$ est strictement décroissante sur $(0, \infty)$ avec $\lim_{E \rightarrow 0} R(E) = 1$ et $\lim_{E \rightarrow \infty} R(E) = 0$. Puisque $\alpha - 1 \in (0, 1)$, il existe un unique $E_* > 0$ tel que $R(E_*) = \alpha - 1$.

Nous avons montré que e_h^- est continue et strictement décroissante en τ_A sur $[0, \tau^c]$. Au niveau de l'ensemble complet seuil de collapse $\tau_A = \tau^c$ les deux racines positives se confondent en un point de tangence $e_*^- > 0$ tel que $D(e_*^-, \tau^c) = 0$ et $D_e(e_*^-, \tau^c) = 0$. La condition de tangence implique

$$\lambda_1 (1 + 1/Y_*^-) e_*^- + 2(\alpha - 1 - R(le_*^-)) = 0,$$

d'où $R(le_*^-) > \alpha - 1$ et $le_*^- < E_*$. Par continuité de e_h^- sur τ_A , pour *tout* τ_A suffisamment proche de τ^c , on a $le_h^- < E_*$.

Définissons

$$\hat{\tau}_A := \inf\{\tau_A \in [0, \tau^c] : le_h^- < E_*\}.$$

Lorsque $\tau_A^c > 0$, l'ensemble entre accolades est un intervalle non vide, et $\hat{\tau}_A \in [0, \tau^c]$ est bien défini. Par construction,

$$\alpha - 1 - R(E(\tau_A)) \begin{cases} > 0, & \text{si } \tau_A \in [0, \hat{\tau}_A), \\ < 0, & \text{si } \tau_A \in (\hat{\tau}_A, \tau_A^c). \end{cases}$$

En utilisant l'identité des signes ci-dessus, on en conclut que

$$\frac{\partial Y_h^-}{\partial \tau_A} > 0 \quad \text{pour } \tau_A \in [0, \hat{\tau}_A) \quad \frac{\partial Y_h^-}{\partial \tau_A} < 0 \quad \text{pour } \tau_A \in (\hat{\tau}_A, \tau_A^c),$$

comme souhaité.

B.3 Preuve de la proposition 9

Fixons τ_A et les autres paramètres, et supposons que l'état stationnaire à haute connaissance existe. Soit $X_h^-(I) > 0$ la précision publique d'état stationnaire correspondante. Rappelons la fonction de valeur (au sein de la cohorte)

$$U^-(X^-, \tau_A) = G(X^-) \cdot \Delta_G + \max_{e' \geq 0} G(\sigma^{-2} + \lambda_1 e' + \tau_A) \cdot G(X^-) \Delta_X - \frac{1}{\alpha} (e')^\alpha,$$

de sorte que $U^{+}(I) = U^-(X_h^-(I), \tau_A)$.

Tout d'abord, notons que $U^-(X, \tau_A)$ est strictement croissante en X pour $X > 0$. Ensuite, les résultats de statique comparative des propositions 4 (ii) et 8 (ii) impliquent que $X_h^-(I)$ est strictement croissante en I . Par conséquent, $U^{+}(I)$ est strictement croissante en I .

B.4 Preuve de la proposition 10

D'après les propositions 3, lorsque $\alpha - \frac{1}{4} > \frac{1}{4}$, l'état d'équilibre à haute connaissance est toujours unique pour tout τ_A et varie de manière continue en fonction de τ_A . Ainsi, d'après le théorème de l'enveloppe,

$$\begin{aligned} \frac{\partial U^{+}}{\partial \tau_A} &= g(Y_h^-) \cdot G_W(l_h^-) \Delta_X + \frac{\partial G_W(l_h^-)}{\partial \tau_A} \cdot (\Delta_G + G(Y_h^-) \Delta_X) \\ &= \underbrace{g(Y_h^-) \cdot G_W(l_h^-) \Delta_X}_{DE} - \underbrace{G'_W(l_h^-) \cdot I \cdot \frac{\partial \bar{e}_h}{\partial \tau_A}}_{IE} \cdot (\Delta_G + G(Y_h^-) \Delta_X). \end{aligned}$$

Étudier le signe de $\frac{\partial U^{+}}{\partial \tau_A}$ il est donc pratique de travailler avec le rapport $\frac{IE}{DE}$. Nous avons l'expression suivante :

$$\begin{aligned} \frac{IE}{DE} &= \frac{G'_W(l_h^-) \cdot I}{G_W(l_h^-)} \cdot \frac{1}{g(Y_h^-)} - \frac{\frac{\partial \bar{e}_h}{\partial \tau_A}}{\Delta_X} \cdot \frac{\Delta_G + G(Y_h^-) \Delta_X}{\Delta_X} \\ &= \frac{R(l_h^-)}{e_h^-} \cdot \frac{1}{g(Y_h^-)} \cdot \frac{(1 + \frac{1}{Y_h^-}) \cdot \bar{e}_h}{\lambda_1 (1 + \frac{1}{Y_h^-}) \cdot e_h^- + 2(\alpha - 1 - R(l_h^-))} \cdot \frac{\Delta_G + G(Y_h^-) \Delta_X}{\Delta_X}. \end{aligned}$$

En réorganisant les termes, nous avons exprimé $\frac{IE}{DE}$ comme une fonction de e_h^- et de Y_h^- :

$$\frac{IE}{DE} = \frac{(1 + 1/Y_h^-) \cdot R(l_h^-)}{\lambda_1 (1 + 1/Y_h^-) \cdot e_h^- + 2(\alpha - 1 - R(l_h^-))} \cdot \frac{\Delta_G + G(Y_h^-) \Delta_X}{g(Y_h^-) \Delta_X}. \quad (B-2)$$

Lemme B-1. Lorsque $Y_h^- \geq \sqrt{2} - 1$, l'application

$$(e_h^-, Y_h^-) \rightarrow \frac{IE}{DE}$$

donnée par (B-2) est strictement décroissante en e_h^- et strictement croissante en Y_h^- .

Preuve. Puisque $R(\cdot)$ est une fonction strictement décroissante, la monotonie par rapport à e_h^- est facile à vérifier. Pour la partie Y_h^- , il suffit de prouver que $\frac{1+Y_h^-}{g(Y_h^-)}$ augmente en fonction de Y_h^- . En effet,

$$\frac{1+Y_h^-}{g(Y_h^-)} = \frac{\sqrt{Y_h^-}}{2\pi e^2} \cdot \frac{\sqrt{Y_h^-}}{Y_h^- + \sqrt{Y_h^-}},$$

qui croît effectivement en Y_h^- en raison de l'hypothèse selon laquelle $Y_h^- \geq \sqrt{2} - 1$. □

La proposition 4 montre que lorsque τ_A augmente, e_h^- diminue tandis que Y_h^- augmente, ce qui fait que le rapport $\frac{IE}{DE}$ strictement croissant en τ_A sur $(0, \infty)$. Nous distinguons deux cas :

Cas I : $\frac{IE}{DE} \geq 1$. Dans ce cas, les arguments ci-dessus garantissent que pour tout $\tau_A > 0$, on a $\frac{IE}{DE} > 1$, de sorte que $\frac{\partial U}{\partial \tau_A} < 0$. Par conséquent, en choisissant $\tau^* = 0$ et les énoncés (i) et (ii) de la proposition s'ensuivent.

Cas II : $\frac{IE}{DE} < 1$. D'après (B-2), on a que pour tout $\tau_A \in (0, +\infty)$, $\frac{IE}{DE}$ est strictement positive, continue et strictement croissante en τ_A . De plus, on a que $\lim_{\tau_A \rightarrow +\infty} \frac{IE}{DE} = +\infty$. En combinant avec notre hypothèse de cas, il existe un seuil unique et fini τ^* pour lequel $\frac{IE}{DE} = 1$.

D'après la monotonie stricte de $\frac{IE}{DE}$ par rapport à τ_A , nous concluons que :

$$\frac{\partial U}{\partial \tau_A} > 0 \quad \text{pour } 0 < \tau_A < \tau^*, \quad \frac{\partial U}{\partial \tau_A} < 0 \quad \text{si } \tau_A > \tau^*.$$

Démontrons maintenant la partie (iii). À partir de la condition du premier ordre, il est facile de montrer que e_h^- , X_h^- tendent vers 0, car $\tau_A \rightarrow +\infty$, ce qui donne également $U^+(\tau_A) \rightarrow 0$.

B.5 Preuve de la proposition 11

Nous restreignons τ_A à l'intervalle $(0, \tau_A^c)$ où existe l'état stationnaire à haute connaissance : $X_h^- > 0$. Par Proposition 5 : l'état stationnaire supérieur est unique dans cette région et dépend de manière continue de τ_A , de sorte que l'on peut dériver U^+ par rapport à τ_A . D'après le théorème de l'enveloppe,

$$\frac{\partial U}{\partial \tau_A} = \underbrace{g(Y_h^-) \cdot G_W(l_h^-) \Delta_X}_{DE} - \underbrace{G'_W(l_h^-) \cdot I \cdot \frac{\partial e_h^-}{\partial \tau_A} \cdot (\Delta_G + G(Y_h^-) \Delta_X)}_{IE}.$$

De manière similaire à la démonstration de la proposition 10, on peut exprimer $\frac{IE}{DE}$ comme une fonction de e_h^- et de Y_h^- :

$$\frac{IE}{DE} = \frac{(1 + 1/Y_h^-) \cdot R(l_h^-)}{\lambda_1(1 + 1/Y_h^-) \cdot e_h^- + 2(\alpha - 1 - R(l_h^-))} \cdot \frac{\Delta_G + G(Y_h^-) \Delta_X}{g(Y_h^-) \Delta_X}, \quad (B-3)$$

et le lemme B-1 continue de s'appliquer ici.

D'après la proposition 8(i), la statique comparative des états d'équilibre supérieurs par rapport à τ_A est la suivante :

$$X_h^- \downarrow, \quad e_h^- \downarrow \quad \text{pour tout } \tau_A < \tau_A^c,$$

tandis que Y_h^- est croissant pour les petits $\tau_A < \hat{\tau}_A$ et finit par diminuer lorsque $\hat{\tau}_A < \tau_A < \tau^c$. A

Ici, nous démontrons d'abord les deux arguments suivants :

- (a) Dans la région où $0 \leq \tau_A < \hat{\tau}_A$ (de sorte que $R(l e_h^-) < \alpha - 1$), e_h^- est décroissante en τ_A , et Y_h^- est croissant en τ_A , de sorte que $\frac{IE}{DE}$ est croissant en τ_A .
- (b) Dans la région où $\hat{\tau}_A \leq \tau_A < \tau^c$ (de sorte que $R(l e_h^-) \geq \alpha - 1$), même si Y_h^- est décroissante en τ_A , $\frac{IE}{DE}$ reste croissant en τ_A .

Prouvons (b). Lorsque $R(l e_h^-) > \alpha - 1$, rappelons que

$$\frac{IE}{DE} = \frac{1}{g(Y_h^-)} \cdot \frac{(1 + \frac{1}{Y_h^-}) \cdot R(l e_h^-)}{\lambda_I (1 + \frac{1}{Y_h^-}) \cdot e_h^- + 2(\alpha - 1 - R(l e_h^-))} \cdot \frac{\Delta_G + G(Y_h^-) \Delta_X}{\Delta_X}.$$

Lemme B-2. Lorsque $\hat{\tau}_A \leq \tau_A < \tau^c$, $\lambda_I (1 + \frac{1}{Y_h^-}) \cdot e_h^- + 2(\alpha - 1 - R(l e_h^-))$ est décroissante en τ_A ; de plus,

$$\frac{d \log(\lambda_I (1 + \frac{1}{Y_h^-}) \cdot e_h^- + 2(\alpha - 1 - R(l e_h^-)))}{d \tau_A} \leq \frac{d \log(\lambda_I (1 + \frac{1}{Y_h^-}) \cdot e_{Y_h}^-)}{d \tau_A}$$

Démonstration

n. Soit

$$T_1 := \lambda_I \left(1 + \frac{1}{Y_h^-}\right) \cdot e_h^-, \quad T_2 := 2(\alpha - 1 - R(l e_h^-)),$$

et posons $T = T_1 + T_2$.

Nous montrons d'abord que T_1 et T_2 sont toutes deux strictement décroissantes en τ_A . Notons que

$$T_1 = \lambda_I \cdot e_h^- + \frac{e_h^-}{\sigma - 2 + \lambda_I e_h^- + \tau_A}.$$

Comme e_h^- est décroissante, T_1 et T_2 sont toutes deux décroissantes. T est donc également décroissante.

Par conséquent, nous avons

$$\frac{T'}{T} = \frac{T_1' + T_2'}{T_1 + T_2} \leq \frac{T_1'}{T_1 + T_2}.$$

Sur $[\hat{\tau}_A, \tau_A^c]$, on a également $T_2 \leq 0$. Comme $T_1' < 0$ et $T_1 + T_2 > 0$, en divisant par le plus petit dénominateur positif, on obtient

$$\frac{T_1'}{T_1 + T_2} \leq \frac{T_1'}{T_1}.$$

En combinant les deux dernières inégalités, on obtient $\tau' \leq \tau'_1$, c'est-à-dire

$$\frac{d}{d\tau_A} \log T \leq \frac{d}{d\tau_A} \log T_1,$$

comme affirmé. □

D'après ce lemme, pour montrer que $\frac{IE}{DE}$ est croissante en τ_A , il suffit de prouver que

$$g\left(\frac{Y^-}{h}\right) \cdot \frac{(1 + \frac{1}{Y^-}) \cdot R(l e^-_h)}{\lambda_I (1 + \frac{1}{Y^-}) \cdot e^-} \cdot \frac{\Delta_G + G(Y_h) \Delta_X}{X} = g\left(\frac{Y^-}{h}\right) \cdot \frac{R(l e^-_h)}{\lambda_I e^-} \cdot \frac{\Delta_G + G(Y_h) \Delta_X}{X}$$

est croissant en τ_A .

En simplifiant et en omettant les facteurs qui ne dépendent pas de τ_A , il suffit de montrer que

$$\frac{G'_W(l e^-_h)}{G_W(l e^-_h)} \cdot \frac{\Delta_G + G(Y_h) \Delta_X}{g(Y_h)}$$

est croissant en τ_A .

Ici, le dénominateur $G_W(l e^-_h) g(Y_h) = \frac{e^{-\alpha-1}}{\lambda_I \Delta_X}$ est décroissant en τ_A . Pour terminer la preuve de (b), il Cela suffit pour démontrer le lemme suivant.

Lemme B-3. Lorsque $\tau_A \leq \tau^c$, $G'_W(l e^-_h) G(Y_h)$ est croissant en τ_A .

Démonstration. Le long de la trajectoire d'équilibre haute, définissons

$$E(\tau_A) := l e^-_h(\tau_A), \quad Y(\tau_A) := Y^-_h(\tau_A) = \sigma^{-2} + \lambda_I e^-_h(\tau_A) + \tau_A.$$

En dérivant l'objectif, on obtient

$$\frac{d}{d\tau_A} \left(G'_W(E(\tau)) G(Y(\tau)) \right) = G''_W(E(\tau)) G(Y(\tau)) + G'_W(E(\tau)) G'(Y(\tau)) \tau' + G'_W(E(\tau)) g(Y(\tau)) Y'(\tau).$$

Il suffit donc de montrer que

$$G''_W(E) G(Y) E'(\tau_A) + G'_W(E) g(Y) Y'(\tau_A) \geq 0. \quad (B-4)$$

À l'état stationnaire élevé, e^-_h satisfait l'équation d'effort en régime permanent (A-1) :

$$G_W(l e^-_h) g(Y_h) = \frac{e^{-\alpha-1}}{\lambda_I \Delta_X} \quad (B-5)$$

Étant donné que le membre de droite de (B-5) est strictement décroissant en τ_A , il en va de même pour le membre de gauche. Dérivée

En calculant, on obtient

$$G'_W(E)g(Y)E'(\tau_A) + G_W(E)g'(Y)Y'(\tau_A) \leq 0. \quad (B-6)$$

En divisant (B-6) par $E'(\tau_A) < 0$ puis par $G_W(E)g'(Y) < 0$, on obtient la borne supérieure

$$\frac{Y'(\tau_A)}{E'(\tau_A)} \leq - \frac{G'_W(E)g(Y)}{G_W(E)g'(Y)}. \quad (B-7)$$

Puisque $E'(\tau_A) < 0$ et $G'_W(E)g(Y) > 0$, l'inégalité (B-4) équivaut à

$$\frac{Y'(\tau_A)}{E'(\tau_A)} \leq - \frac{G''_W(E)G(Y)}{G'_W(E)g(Y)}. \quad (B-8)$$

Par conséquent, d'après (B-7), il suffit de prouver que

$$\frac{G''_W(E)G(Y)}{G'_W(E)g(Y)} \geq - \frac{G'_W(E)g(Y)}{G_W(E)g'(Y)},$$

ce qui équivaut à

$$- \frac{G_W(E)G'_W(E)}{(G'_W(E))^2} \geq - \frac{g(Y)^2}{G(Y)g'(Y)}. \quad (B-9)$$

Dans ce qui suit, nous démontrons que le membre de droite de (B-9) est faiblement inférieur à 1, tandis que le membre de gauche est faiblement supérieur à 1.

Rappelons que $g'(Y) = -\frac{Y+1}{2Y}g(Y)$, donc

$$- \frac{g(Y)^2}{G(Y)g'(Y)} = \frac{2Yg(Y)}{(Y+1)G(Y)}.$$

Soit $z = \sqrt{Y}$. En utilisant $g(Y) = \phi(z)/z$ et $G(Y) = 2\Phi(z) - 1$, on obtient

$$\begin{aligned} \frac{2Yg(Y)}{(Y+1)G(Y)} &= \frac{2z\phi(z)}{(1+z^2)(2\Phi(z)-1)} \\ &\leq \frac{2z\phi(z)}{(1+z^2)^2z\phi(z)} \\ &\leq 1 \end{aligned}$$

Pour le côté gauche de (B-9), notons que

$$G'_W(E) = g(W)W'(E), \quad G''_W(E) = g'(W)(W'(E))^2 + g(W)W''(E).$$

En substituant dans le membre de gauche, on obtient

$$- \frac{G_W(E)G'_W(E)}{(G'_W(E))^2} = - \frac{G(W)g'(W)}{g(W)^2} - \frac{G(W)}{g(W)} \frac{W''(E)}{(W'(E))^2}. \quad (B-10)$$

Nous affirmons que $W''(E) < 0$ pour tout $E > 0$. En effet, le lemme A-3 implique

$$W'(E) = \frac{\lambda_G}{4} \frac{(S(E) - 1)^2}{S(E)}, \quad S(E) := \frac{1}{1 + \frac{a}{E}}, \quad a := \frac{4}{\lambda^G \Sigma^2}.$$

Puisque $S'(E) < 0$ et que $W'(E)$ est strictement croissante en $S(E)$, on a $W''(E) < 0$.

Par conséquent, le deuxième terme de (B-10) est non négatif :

$$\frac{G(W)}{g(W)} - \frac{W''(E)}{(W'(E))^2} \geq 0.$$

De plus, par analogie avec l'analyse du côté droit de (B-9), on a –
le premier terme de (B-10) est au moins égal à 1. Par conséquent

$$\frac{g(W)^2}{G(W)g'(W)} \leq 1, \text{ donc}$$

$$\frac{G_W(E) G'_W(E)}{(G_W(E))^2} \geq 1.$$

Par conséquent, (B-9) est vrai, de sorte que

$$\frac{d}{d\tau_A} W G'(E(\tau)) G(Y(\tau)) \geq 0,$$

comme souhaité. □

En combinant (a) et (b), on en conclut que le rapport $\frac{IE}{DE}$ est strictement croissant en τ_A tant que $0 \leq \tau_A < \tau_A^c$.

On distingue deux cas :

Cas I : $\frac{IE}{DE} \Big|_{\tau_A=0} \geq 1$. Dans ce cas, les arguments ci-dessus garantissent que, pour tout $0 < \tau_A < \tau^c$, on a $\frac{IE}{DE} > 1$, de sorte que $\frac{\partial U^+}{\partial \tau_A} < 0$. Par conséquent, en choisissant $\tau^* = 0$, les propositions (i) et (ii) de la proposition s'ensuivent.

Cas II : $\frac{IE}{DE} \Big|_{\tau_A=0} < 1$. D'après (B-3), on a que pour tout $\tau_A \in (0, \tau_A^c)$, $\frac{IE}{DE}$ est strictement positif, continue et strictement croissante en τ_A . De plus, il est facile de vérifier que $\lim_{\tau \uparrow \tau^c} \frac{c'est\text{-à-dire}}{A \quad A \quad DE} = +\infty$. En combinant cela avec notre hypothèse de cas, il existe un seuil fini et unique $\tau^* < \tau_A^c$ pour lequel $\frac{IE}{DE} = 1$. Nous en concluons que :

$$\frac{\partial U^+}{\partial \tau_A} > 0 \quad \text{pour } 0 < \tau_A < \tau^*, \quad \frac{\partial U^+}{\partial \tau_A} < 0 \quad \text{pour } \tau_A < \tau^* < \tau_A^c.$$

La partie (iii) découle directement de la définition de τ^c .

B.6 Preuve de la proposition 12

Nous considérons d'abord l'asymptotique de $\tau_A^c(I)$. Lorsque $I \rightarrow +\infty$, il faut que $\tau_A^c(I) > 0$, de sorte que

La condition de tangence (B-1) s'applique. Notons par e_I^- et Y_I^- l'effort et la précision idiosyncratique au point de tangence correspondant à I et à $\tau_A = \tau_A^c(I)$. En utilisant les formules dérivées de la démonstration de proposition 4, (B-1) implique que

$$0 = \Delta_X \cdot G_W(E_I^-) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e_I^- + \tau_A) - e_I^{-\alpha-1} \quad (\text{B-11})$$

$$0 = \lambda_I (1 + 1/Y_I^-) e_I^- + 2(\alpha - 1 - R(E_I^-)) \quad (\text{B-12})$$

où

$$E_I^- := I e_I^-, \quad Y_I^- := \sigma^{-2} + \lambda_I e_I^- + \tau_A^c(I).$$

Le lemme A-4 montre que $R(E)$ est strictement décroissante sur $(0, \infty)$ avec $\lim_{E \downarrow 0} R(E) = \frac{1}{4}$ et $\lim_{E \rightarrow \infty} R(E) = 0$. Par conséquent, puisque $\alpha - 1 < \frac{1}{4}$, il existe un unique $E_* > 0$ tel que $R(E_*) = \alpha - 1$.

Tout d'abord, nous montrons que E_I^- converge vers E_* lorsque $I \rightarrow \infty$. D'après (B-12), nous avons $R(E_I^-) > \alpha - 1$, donc

$E_I^- \leq E_*$ pour tout I . Par conséquent, on a

$$e_I^- = E_I^- / I \leq E_* / I \rightarrow 0.$$

Il en résulte que $\lambda_I (1 + 1/Y_I^-) e_I^- \rightarrow 0$ et (B-12) donne $R(E_I^-) \rightarrow \alpha - 1$. Par conséquent, $E_I^- \rightarrow E_* \in (0, \infty)$.

Ensuite, d'après (B-11), on a

$$g(Y_I^-) = \frac{e_I^{-\alpha-1}}{\Delta_X \lambda_I G_W(E_I^-)} = \frac{\bar{E}_I^{\alpha-1}}{\Delta_X \lambda_I G_W(E_I^-)} I^{-(\alpha-1)}.$$

Puisque $E_I^- \rightarrow E_*$ et $G_W(E_*) > 0$, le préfacteur converge vers une certaine constante $C_0 > 0$, donc

$$g(Y_I^-) \sim C_0 I^{-(\alpha-1)} \quad (I \rightarrow \infty).$$

Comme g est strictement décroissante avec $g(y) \downarrow 0$ lorsque $y \rightarrow \infty$, cela implique que $Y_I^- \rightarrow \infty$. En utilisant l'expression $g(y) = (2\pi)^{-1/2} y^{-1/2} e^{-y/2}$, on obtient :

$$Y_I^- = 2(\alpha - 1) \log I + O(\log \log I) \quad (I \rightarrow \infty).$$

Enfin, à partir de l'expression de Y_I^- et du fait que $e_I^- \rightarrow 0$, on a

$$\tau_A^c(I) = 2(\alpha - 1) \log I + O(\log \log I),$$

et donc

$$\lim_{I \rightarrow \infty} \frac{\tau_A^c(I)}{\log I} = 2(\alpha - 1).$$

Ce qui achève la démonstration.

Ensuite, nous montrons l'asymptotique de $\tau_A^*(I)$. Notons par e^* et Y^* l'effort et l'idiosyncrasique à l'état stationnaire élevé correspondant à I et $\tau_A = \tau_A^*(I)$:

$$e^* := e_n^-(\tau_A^*(I), I), \quad Y^* := \sigma^{-2} + \lambda_I e^* + \tau_A^*(I),$$

et définissons également $E^* := Ie^*$ par souci de commodité.

À tout état stationnaire positif, on a la condition d'effort

$$(e^*)^{\alpha-1} = \Delta_X \lambda_I G_W(E^*) g(Y^*). \quad (B-13)$$

Si l'optimum de bien-être $\tau_A^*(I) > 0$ (ce qui est le cas lorsque $I \rightarrow +\infty$), notre analyse précédente implique que $IE/DE = 1$, ce qui est

$$1 = \frac{IE}{DE} = \frac{(1 + 1/Y^*) \cdot R(E^*)}{\lambda_I (1 + 1/Y^*) \cdot e^* + 2(\alpha - 1 - R(E^*))} \cdot \frac{\Delta_G + G(Y^*)\Delta_X}{g(Y^*)\Delta_X}.$$

En résolvant $g(Y^*)$ à partir de (B-13) et en le substituant, on obtient

$$(e^*)^{\alpha-1} [\lambda_I (1 + 1/Y^*) \cdot e^* + 2(\alpha - 1 - R(E^*))] = \lambda_I G_W(E^*) (1 + 1/Y^*) R(E^*) (\Delta_G + G(Y^*)\Delta_X). \quad (B-14)$$

Nous montrons d'abord que, le long de la suite optimale $I \rightarrow \infty$,

$$E^* \rightarrow \infty, \quad e^* \rightarrow 0.$$

Si $E_D^* \rightarrow \tilde{E} < \infty$ pour un certain $I_n \rightarrow \infty$, alors $e^* = E_D^*/I_n \rightarrow 0$, tandis que $G_W(E^*) \rightarrow G_W(\tilde{E}) > 0$ et $R(E^*) \rightarrow R(\tilde{E}) > 0$. En faisant tendre n vers l'infini dans (B-14), le membre de gauche tend vers 0 et le membre de droite vers au moins $\lambda_I G_W(\tilde{E}) R(\tilde{E}) (\Delta_G + \Delta_X G(\sigma^{-2})) > 0$, ce qui est une contradiction. Ainsi, $E^* \rightarrow \infty$.

Supposons plutôt que $e_n^* \rightarrow \tilde{e} < \infty$ sur un certain intervalle $I_n \rightarrow \infty$. Pour E grand, on a

$$G_W(E) = G_\infty + O(E^{-1}), \quad R(E) = \frac{c_R}{E} + O(E^{-2}),$$

avec les constantes $G_\infty > 0$, $c_R > 0$. En substituant dans (B-14) et en faisant tendre n vers l'infini, le membre de droite converge vers zéro, tandis que le membre de gauche converge au moins vers $\tilde{e}^{\alpha-1} (\lambda_I \tilde{e} + 2(\alpha - 1)) > 0$; là encore, une contradiction. Par conséquent, $e^* \rightarrow 0$, ce qui implique également $E^* = o(I)$. D'après (B-13), on obtient également $Y^* \rightarrow \infty$.

En substituant toutes les limites asymptotiques dans (B-14), on obtient

$$\lim_{I \rightarrow \infty} \frac{(e^*)^{\alpha-1}}{R(E^*)} = \lim_{I \rightarrow \infty} \frac{\lambda_I G_W(E^*) (1 + 1/Y^*) (\Delta_G + G(Y^*)\Delta_X)}{\lambda_I (1 + 1/Y^*) \cdot e^* + 2(\alpha - 1 - R(E^*))} = \frac{\lambda_I G_\infty (\Delta_G + \Delta_X)}{2(\alpha - 1)}. \quad (B-15)$$

En utilisant $R(E) = c_R/E + O(E^{-2})$ et $E^* = I e^*$, (B-15) devient

$$(e^*)^\alpha \sim \frac{\lambda_I G_\infty (\Delta_G + \Delta_X) c_R}{2(\alpha - 1)} I^{-1},$$

d'où

$$e^* \sim C_E \cdot I^{-1/\alpha}, \quad (\text{B-16})$$

pour une certaine constante $C_E > 0$.

À partir de l'équation d'équilibre (B-13) et de la limite *pour E très grand* $G_W(E) \rightarrow G_{(\infty)}$, on obtient

$$g(Y^*) = \frac{(e^*)^{\alpha-1}}{\Delta_X \lambda_I G_W(E^*)} \sim C_g I^{-(\alpha-1)/\alpha},$$

pour une certaine constante $C_g > 0$, d'après (B-16).

En utilisant l'expression $g(y) = (2\pi)^{-1/2} y^{-1/2} e^{-y/2}$, on en conclut :

$$Y^* = \frac{2(\alpha-1)}{\alpha} \log I + O(\log \log I).$$

En utilisant la définition de Y^* et le fait que $e^* \rightarrow 0$, on obtient

$$\tau^*(I) = \frac{2(\alpha-1)}{\alpha} \log I + O(\log \log I).$$

On a donc $\lim_{I \rightarrow \infty} \frac{\tau^*(I)}{\log I} = \frac{2(\alpha-1)}{\alpha}$, ce qui est la borne asymptotique recherchée.

B.7 Démonstration de la proposition 13

Considérons une politique d'information finalement constante $\kappa \in K^{EC}$, de sorte qu'il existe $M < \infty$ et κ^- tels que $\kappa_t = \kappa^-$ pour tout $t \geq M$. Soit $\tau^- := \tau_A(\kappa^-)$ la précision effective correspondante. À partir de la période M , la récursion d'équilibre induite coïncide avec la récursion dans une économie à précision effective constante τ^- , à partir de la condition initiale X_M :

$$X_{t+1} = F_{\tau^-}(X_t), \quad t \geq M,$$

où F_{τ^-} est la fonction de transition constante τ^- . La proposition 5 implique alors la convergence de X_t :

- Si $\tau^- \geq \tau_A^c$, alors 0 est le seul état stationnaire globalement stable ; par conséquent, $X_t \rightarrow 0$ lorsque $t \rightarrow \infty$.
- Si $\tau^- < \tau_A^c$ alors il existe trois états stationnaires $0 = X_\ell^-(\tau^-) < X_m^-(\tau^-) < X_h^-(\tau^-)$ et la partition des bassins s'applique : $X_M > X_m^-(\tau^-) \Rightarrow X_t \rightarrow X_h^-(\tau^-)$, tandis que $X_M < X_m^-(\tau^-) \Rightarrow X_t \rightarrow 0$.

Par conséquent, nous avons

$$U_t(\kappa) = \begin{cases} U^-(\tau^-), & \text{si } \tau^- < \tau_A^c \text{ et } X_M > X_m^-(\tau^-), \\ 0, & \text{sinon.} \end{cases}$$

Étant donné qu'une suite bornée convergente a la même limite de Cesàro que sa limite point par point,

$$U(\kappa) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T U_t(\kappa) = \begin{cases} U^-(\tau^-), & \text{si } \tau^- < \tau_A^c \text{ et } X_M > X_m^-(\tau^-), \\ 0, & \text{sinon.} \end{cases} \quad (B-17)$$

Nous construisons ensuite la politique biphase optimale. Considérons le régime de suppression totale $\kappa_t = 0 \ \forall t$, ce qui revient à dire $\tau_t = 0 \ \forall t$. Comme $X_1 > X_m^-(0)$ par hypothèse, la proposition 5 implique que sous $\tau_t = 0 \ \forall t$, $X_t^{\text{supp}} \rightarrow X_h^-(0)$ lorsque $t \rightarrow \infty$.

Ensuite, $\tau^* \leq \tau_A^c$, de sorte que l'état d'équilibre à haute connaissance sous τ^* existe et satisfait

$$0 < X_m^-(\tau^*) < X_h^-(\tau^*).$$

De plus, puisque l'état stationnaire à haute connaissance $X_h^-(\tau)$ est (strictement) décroissant en τ sur $[0, \tau_A^c]$, on a

$$X_h^-(0) \geq X_h^-(\tau^*) > X_m^-(\tau^*).$$

On a donc $X_h^-(0) - X_m^-(\tau^*) > 0$, et grâce à la convergence de X_t , il existe un nombre fini M^* tel que

$$X_{M^*} > X_m^-(\tau^*).$$

Définissons la politique à deux phases κ^{2ph} par $\kappa^{2ph} = 0$ pour $t < M^*$ et $\kappa^{2ph} = \kappa^*$ pour $t \geq M^*$. Par construction, la précision publique effective à la date de basculement satisfait $X_{M^*}(\kappa^{2ph}) = X^{\text{supp}} > X_m^-(\tau^*)$. L'application de (B-17) avec $\tau^- = \tau^*$ donne $U(\kappa^{2ph}) = U^-(\tau^*)$, ce qui prouve la partie (i).

Enfin, nous démontrons l'optimalité de κ^{2ph} . Soit $\kappa \in K^{EC}$ un élément arbitraire et τ^- sa précision effective de queue. (B-17) implique que soit $U(\kappa) = 0$, soit $U(\kappa) = U^-(\tau^-)$. Dans les deux cas,

$$U(\kappa) \leq \max_{\tau \in [0, \tau_A]} U^-(\tau).$$

D'après la proposition 11, $U^-(\tau)$ est croissante sur $(0, \tau_A^*)$ et décroissante sur (τ_A^*, τ_A^c) , avec $U^-(\tau) = 0$ pour $\tau \geq \tau_A^c$. Par conséquent, le point maximal de $U^-(\tau)$ sur l'ensemble des solutions admissibles $[0, \tau_A]$ est τ^* .

B.8 Preuve de la proposition 14

Ici, l'IA agentique de précision τ_A fournit à la fois (i) le signal idiosyncratique $s^A \sim N(\vartheta_{i,t}, \tau_{i,t}^{-1})$ et (ii) élève l'échelle d'agrégation effective d'une valeur de référence l_0 à

$$l(\tau_A) = l_0 + \exp(\eta \tau_A).$$

Fixons $\tau_A \geq 0$ et supposons qu'il existe un état d'équilibre positif (à haut niveau de connaissance), avec un effort par agent $\bar{e}_h(\tau_A) > 0$ et un effort agrégé (au niveau de l'îlot)

$$E_h^-(\tau_A) := l(\tau_A) \bar{e}_h(\tau_A).$$

La précision publique en régime permanent est $X_h^-(\tau_A) = W(E_h^-(\tau_A))$, où $W(\cdot)$ est continue sur $\mathbb{R}_+$ et satisfait $W(0) = 0$, et $W(E) > 0$ pour $E > 0$.

Ainsi, pour montrer que $X_h^-(\tau_A) \rightarrow 0$ lorsque $\tau_A \rightarrow \infty$, il suffit de prouver que

$$E_h^-(\tau_A) = l(\tau_A) \bar{e}_h(\tau_A) \rightarrow 0. \quad (\text{B-18})$$

Nous montrons ici que l'effort doit décroître au moins de manière exponentielle en τ_A . À tout état stationnaire positif, l'effort par agent $\bar{e}_h(\tau_A)$ satisfait la condition du premier ordre :

$$\bar{e}_h(\tau_A)^{\alpha-1} = \Delta_X G(X_h^-(\tau_A)) \lambda_I Y_h^-(\tau_A), \quad Y_h^-(\tau_A) = \sigma^{-2} + \lambda_I \bar{e}_h(\tau_A) + \tau_A.$$

Comme $g(\cdot)$ est strictement décroissante, on obtient la borne supérieure brute

$$\bar{e}_h(\tau_A)^{\alpha-1} \leq \Delta_X \lambda_I g(\sigma^{-2} + \tau_A) \leq \Delta_X \lambda_I g(\tau_A). \quad (\text{B-19})$$

En rappelant la formule gaussienne $g(\tau) = \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau}}$, on peut réécrire (B-19) sous la forme

$$\bar{e}_h(\tau_A) \leq (\Delta_X \lambda_I (2\pi)^{-1/2})^{\frac{1}{\alpha-1}} \tau_A^{-\frac{1}{2(\alpha-1)}} \exp\left(-\frac{\tau_A}{2(\alpha-1)}\right). \quad (\text{B-20})$$

Multiplier la limite (B-20) par $l(\tau_A) = l_0 + \exp(\eta \tau_A)$:

$$\begin{aligned} E_h^-(\tau_A) &= l(\tau_A) \bar{e}_h(\tau_A) \\ &\leq (\Delta_X \lambda_I (2\pi)^{-1/2})^{\frac{1}{\alpha-1}} (l_0 + e^{\eta \tau_A}) \cdot \tau_A^{-\frac{1}{2(\alpha-1)}} \exp\left(-\frac{\tau_A}{2(\alpha-1)}\right) \\ &= C \cdot \tau_A^{-\frac{1}{2(\alpha-1)}} l_0 e^{-\frac{\tau_A}{2(\alpha-1)} + \eta \tau_A} = C \cdot \tau_A^{-\frac{1}{2(\alpha-1)}} l_0 e^{(\eta - \frac{1}{2(\alpha-1)})\tau_A}, \end{aligned} \quad (\text{B-21})$$

pour une constante $C > 0$ indépendante de τ_A .

Rappelons
que

$$\eta_{\text{crit}} := \frac{1}{2(\alpha - 1)}.$$

Ainsi, si $\eta < \eta_{\text{crit}}$, alors l'exposant $(\eta - \eta_{\text{crit}})$ dans (B-21) est négatif, de sorte que le terme entre crochets converge vers 0 lorsque $\tau_A \rightarrow \infty$; le facteur polynomial $\tau^{-1/(2(\alpha-1))}$ converge également vers 0. Par conséquent, la borne supérieure (B-21) implique que $E_h^-(\tau_A) \rightarrow 0$, ce qui établit (B-18).

Nous avons ainsi démontré que $X_h^-(\tau_A) \rightarrow 0$ lorsque $\tau_A \rightarrow \infty$, ce qui implique que $\lim_{\tau_A \rightarrow \infty} U^+(\tau_A) = 0$. Par conséquent, le supremum de $U^+(\tau_A)$ sur $\tau_A \geq 0$ ne peut être atteint « à l'infini » : $\arg \max_{\tau \geq 0} U^+(\tau_A) < +\infty$.

B.9 Preuve de la proposition 15

Soit $X := [0, \Sigma^{-2}]$. Pour tout $X \geq 0$, on a $0 < F_{\text{syn}}(X) < \Sigma^{-2}$. Ainsi, $F_{\text{syn}} : X \rightarrow X$. De plus, F_{syn} est continue et strictement croissante sur X . Par conséquent, d'après le théorème du point fixe de Tarski, l'ensemble des points fixes de F_{syn} n'est pas vide et admet un élément minimal et un élément maximal; notons-les respectivement X_{ℓ}^{-*} et X_h^{-*} . Notons les efforts d'équilibre correspondants par

$$\bar{e}_k^{-*} := e(X_k^{-*}, \tau_A), \quad k \in \{\ell, h\}.$$

Comme $\tau_{\text{syn}} > 0$ et $e(0, \tau_A) = 0$,

$$F_{\text{syn}}(0) = \frac{h}{\Sigma^2 + \tau_{\text{syn}}^{-1}} > 0.$$

Puisque X_{ℓ}^{-*} est un point fixe et qu'il s'agit du point fixe *minimal* de X , on a nécessairement $X_{\ell}^{-*} \geq F_{\text{syn}}(0) > 0$, d'où $X_{\ell}^{-*} > 0$, ce qui implique également que $e_{\ell}^{-*} = e(X_{\ell}^{-*}, \tau_A) > 0$. Cela prouve la partie (1).

Nous notons tout d'abord la monotonie ponctuelle de F_{syn} par rapport aux paramètres. En particulier,

- $F_{\text{syn}}(X)$ est strictement croissante sur I pour tout $X > 0$.
- $F_{\text{syn}}(X)$ est strictement croissante en τ_{syn} pour tout $X \geq 0$.
- $F_{\text{syn}}(X)$ est strictement décroissante en τ_A pour tout $X > 0$.

Nous prouvons ces affirmations pour le point fixe minimal X_{ℓ}^{-*} ; le raisonnement pour le point fixe maximal est analogue.

Capacité d'agrégation I. Soit $I_2 > I_1$, et posons $F_j(\cdot) := F_{\text{syn}}(\cdot; I_j, \tau_A, \tau_{\text{syn}})$ et $X_{\ell,j}^{-*}$ pour le plus petit point fixe de F_j . D'après la monotonie sur I , $F_2(X) \geq F_1(X)$ pour tout $X \in X$, et $F_2(X) > F_1(X)$ pour tout $X > 0$. La statique comparative monotone standard pour les points fixes extrêmes implique $X_{\ell,2}^{-*} \geq X_{\ell,1}^{-*}$. Pour montrer que l'inégalité est stricte, supposons, par contradiction, que $X_{\ell,2}^{-*} = X_{\ell,1}^{-*} =: x$. D'après la partie (1), $x > 0$, et donc l'étape 3 donne $F_2(x) > F_1(x) = x$, ce qui contredit le fait que x soit un point fixe de F_2 . Par conséquent X_{ℓ}^{-*} est strictement croissante sur I .

La statique comparative monotone pour la précision synthétique τ_{syn} et la précision agentique τ_A est analogue.

Pour tout $k \in \{\ell, h\}$, $e_k^- = e_k(X_k^-, \tau_A)$. Étant donné que $e(\cdot, \tau_A)$ est strictement croissante en X , la monotonie stricte de X_k^- sur I et τ_{syn} implique que e_k^- est strictement croissante en I et τ_{syn} .

Pour τ_A , posons $\tau^2 > \tau^1$. Alors $X_k^-(\tau^2) < X_k^-(\tau^1)$ et $e(X, \tau_A)$ est strictement décroissante en τ_A , donc

$$e_k^-(\tau^2) = e_k(X_k^-(\tau^2), \tau^2) < e_k(X_k^-(\tau^1), \tau^2) < e_k(X_k^-(\tau^1), \tau^1) = e_k^-(\tau^1),$$

établissant que e_k^- est strictement décroissante en τ_A .

B.10 Preuve de la proposition 16

Dans le cadre de la technologie d'apprentissage public modifiée (14), un équilibre symétrique satisfait toujours $e_{it} = e_i$ pour tout i , et la meilleure réponse individuelle $e_t = e(X_t, \tau_A)$ reste inchangée par rapport au modèle de référence, car β n'influe que sur la manière dont l'effort se traduit en signal public pour la cohorte suivante.

Avec $\beta > 0$, la précision du signal public au niveau de l'île générée par la cohorte t est égale à $\lambda_G E^{(\beta)} = \lambda_G I e^\beta$.

Par conséquent, la précision du signal public évolue selon

$$v_{t+1}^{-1} = (X_t + \lambda_G I e^\beta)^{-1} + \Sigma^2, \quad e_t = e(X_t, \tau_A). \quad (\text{B-22})$$

Définissons la fonction induite à un pas

$$F_\beta(X) := \Sigma^2 + (X + \lambda_G I e(X, \tau_A)^\beta)^{-1}.$$

Alors, l'état stationnaire de effondrement de la connaissance $X = 0$ reste un point fixe de F_β lorsque $\beta > 0$ car $e(0, \tau_A) = 0$ implique que $F_\beta(0) = 0$.

La condition que l'effort en régime permanent doit satisfaire est modifiée comme suit :

$$D_\beta(e; \tau_A, I) := \Delta_X \cdot G_W(I e^\beta) \cdot \lambda_I g(\sigma^{-2} + \lambda_I e + \tau_A) - e^{\alpha-1}.$$

D'après le lemme A-3, $G_W(E) = \kappa_1 E^{1/4} + o(E^{1/4})$ lorsque E tend vers 0, d'où $G_W(I e^\beta) = \kappa e^{\beta/4} + o(e^{\beta/4})$ lorsque e tend vers 0 (pour $\kappa > 0$). Par conséquent, lorsque e tend vers 0,

$$D_\beta(e; \tau_A, I) = C(\tau_A, I) e^{\beta/4} - e^{\alpha-1} + o(e^{\beta/4}),$$

et le même raisonnement de stabilité locale que dans le lemme 2 reste valable en remplaçant $1/4$ par $\beta/4$. Les principaux arguments utilisés pour caractériser les états stationnaires restent inchangés, la condition de stabilité étant remplacée. Par conséquent, les propositions 3 et 5 sont valables lorsque la condition $\alpha - 1 > \frac{1}{4}$ (ou $\alpha - 1 < \frac{1}{4}$) est remplacée par

$\alpha - 1 > \frac{\beta}{4}$ (ou $\alpha - 1 < \frac{\beta}{4}$). Les analyses statiques comparatives s'appliquent de la même manière.

Nous examinons ensuite le comportement asymptotique de $\tau^c(I)$. Pour que ce seuil existe, nous supposons que $\alpha - 1 < \beta/4$. Par analogie avec la démonstration de la proposition 12, nous avons également que les conditions de tangence sont vérifiées en $\tau_A = \tau^c(I)$,
 qui se présentent désormais sous la forme :

$$0 = \Delta_X \cdot G_W(E_1^-) \cdot \lambda_1 g(\sigma^{-2} + \lambda_1 e_1^- + \tau^c(I)) - e_1^{-\alpha-1}, \quad (B-23)$$

$$0 = \lambda_1 (1 + 1/Y_1^-) e_1^- + 2\alpha - 1 - \theta R(E_1^-), \quad (B-24)$$

où

$$E_1^- := I e_1^{-\beta}, \quad Y_1^- := \sigma^{-2} + \lambda_1 e_1^- + \tau^c(I), \quad R(E) := \frac{EG'_W(E)}{G_W(E)}.$$

Puisque $\alpha - 1 < \beta/4$, d'après le lemme A-3, il existe un unique $E_* > 0$ tel que $R(E_*) = \frac{\alpha-1}{\beta}$.

Comme dans la démonstration initiale, on peut montrer que $e_1^- \rightarrow 0$ et $E_1^- \rightarrow E_*$ lorsque $I \rightarrow \infty$. Ensuite, d'après (B-23), on a

$$g(Y_1^-) = \frac{e_1^{-\alpha-1}}{\Delta_X \lambda_1 G_W(E_1^-)} = \frac{\bar{E}_1^{-(\alpha-1)/\beta}}{\Delta_X \lambda_1 G_W(E_1^-)} I^{-(\alpha-1)/\beta}.$$

Puisque $E_1^- \rightarrow E_*$ et que $G_W(E_*) > 0$, le préfacteur converge vers une constante $C_0 > 0$, donc

$$g(Y_1^-) \sim C_0 I^{-(\alpha-1)/\beta} \quad (I \rightarrow \infty).$$

Cela implique alors que $\hat{r}(I) = \frac{2(\alpha-1)}{\beta \log I}$ en suivant la même procédure que dans la preuve de la proposition 12.

Nous allons maintenant présenter l'asymptotique de $\tau^*(I)$. Définissons e^* et Y^* de la même manière que l'effort et l'idiosyncrasique

à l'état stationnaire élevé correspondant à I et $\tau_A = \tau^*(I)$

$$e^* := e_h^-(\tau^*(I), I), \quad Y^* := \sigma^{-2} + \lambda_1 e^* + \tau^*(I),$$

et définissons $E^* := I(e^*)^\beta$.

À tout état stationnaire positif, on a la condition d'effort

$$(e^*)^{\alpha-1} = \Delta_X \lambda_1 G_W(E^*) g(Y^*). \quad (B-25)$$

Dans l'économie β , le rapport $IE/DE = 1$ devient

$$1 = \frac{IE}{DE} = \frac{(1 + 1/Y^*) \cdot \theta R(E^*)}{\lambda_1 (1 + 1/Y_1^*) \cdot e_1^* + 2(\alpha - 1 - \theta R(E^*))} \cdot \frac{\Delta_G + G(Y^*)\Delta_X}{g(Y^*)\Delta_X}.$$

En résolvant $g(Y^*)$ à partir de (B-25) et en le substituant, on obtient

$$(e_1^*)^{\alpha-1} \lambda_1 (1 + 1/Y_1^*) \cdot e_1^* + 2(\alpha - 1 - \theta R(E^*)) = \lambda_1 G_W(E^*) (1 + 1/Y^*) \cdot \theta R(E^*) \cdot \Delta_G + G(Y^*)\Delta_X. \quad (B-26)$$

À l'instar de la démonstration de la proposition 12, on peut montrer que lorsque l tend vers l'infini,

$$E^* \text{ tend vers l'infini, } e_1^* \text{ tend vers 0.}$$

En substituant ces asymptotiques dans (B-26), on obtient

$$\lim_{l \rightarrow \infty} \frac{(e_1^*)^{\alpha-1}}{R(E^*)} = \frac{\lambda_I G_\infty \cdot \beta(\Delta_G + \Delta_X)}{2(\alpha - 1)}. \quad (\text{B-27})$$

En utilisant $R(E) = c_R/E + O(E^{-2})$ et $E^* = l(e^*)^\beta$, (B-27) devient

$$l(e^*)^{\alpha-1+\beta} \sim \frac{\lambda_I G_\infty \cdot \beta(\Delta_G + \Delta_X) c_R}{2(\alpha - 1)},$$

donc $e^* \sim C_E \cdot l^{-1/(\alpha-1+\beta)}$ pour une certaine constante $C_E > 0$.

À partir de la condition d'équilibre (B-25) et de $G_W(E^*) \rightarrow G_\infty$, on obtient

$$g(Y^*) = \frac{(e_1^*)^{\alpha-1}}{\Delta_X \lambda_I G_W(E_1^*)} \sim C_g l^{-(\alpha-1)/(\alpha-1+\beta)},$$

ce qui implique

$$\lim_{l \rightarrow \infty} \frac{\frac{\tau_A(l)}{\log l}}{l} = \frac{2(\alpha - 1)}{\alpha - 1 + \beta}.$$