STONE CENTER | INEQUALITY & SHAPING THE FUTURE OF WORK

AI, Human Cognition and Knowledge Collapse

Daron Acemoglu
Dingwen Kong
Asuman Ozdaglar

May 2026

Stone Center on Inequality and Shaping the Future of Work
70 Memorial Drive, Building E51-355 - Cambridge, MA 02142

shapingwork.mit.edu

AI, Human Cognition and Knowledge Collapse*

Daron Acemoglu† Dingwen Kong‡ Asuman Ozdaglar§

May 5, 2026

Abstract

We study how generative AI, and in particular agentic AI, shapes human learning incentives and the long-run evolution of society's information ecosystem. We build a dynamic model of learning and decision-making in which successful decisions require combining shared, community-level general knowledge with individual-level, context-specific knowledge; these two inputs are complements. Learning exhibits *economies of scope*: costly human effort jointly produces a private signal about their own context and a "thin" public signal that accumulates into the community's stock of general knowledge, generating a learning externality. Agentic AI delivers context-specific recommendations that *substitute* for human effort. By contrast, a richer stock of general knowledge *complements* human effort by raising its marginal return.

The model highlights a sharp dynamic tension: while agentic AI can improve contemporaneous decision quality, it can also erode learning incentives that sustain long-run collective knowledge. When human effort is sufficiently elastic and agentic recommendations exceed an accuracy threshold, the economy can tip into a *knowledge-collapse* steady state in which general knowledge vanishes ultimately, despite high-quality personalized advice. Welfare is generally non-monotone in agentic accuracy, implying an interior, welfare-maximizing level of agentic precision and motivating information-design regulations. In contrast, greater aggregation capacity for general knowledge—meaning more effective sharing and pooling of human-generated general knowledge—unambiguously raises welfare and increases resilience to knowledge collapse.

Keywords: artificial intelligence, cognition, collective knowledge, community, social learning.
JEL Classification: D80, D83.

*We are grateful to the Hewlett Foundation, the Stone Foundation and the MIT Gen-AI Consortium for financial support.

†Massachusetts Institute of Technology, NBER, and CEPR, daron@mit.edu

‡Massachusetts Institute of Technology, dingwenk@mit.edu

§Massachusetts Institute of Technology, asuman@mit.edu

1 Introduction

Rapid advances in (generative) AI since 2022 have raised hopes of widespread benefits from superior AI-provided information provision and task services (Chui et al., 2023; Maslej et al., 2025). AI models can now analyze and write text on an impressive range of topics on a par with specialized human experts and can sift through vast data sets to recognize patterns and locate relevant useful information (OpenAI, 2023). Most importantly, they can use these capabilities to provide context-specific information and recommendations. Nevertheless, growing use of generative AI tools has also raised fears, for instance, about how reliance on AI will impact human cognition, knowledge and creativity (Holmes et al., 2023).

Opinions are divergent on this question. On one side are commentators and experts who believe that with the use of AI tools human knowledge will expand greatly, for example, in the form of faster and better scientific discoveries, and much better decision-making by professionals with access to AI. In this spirit, Demis Hassabis has described AI as "the ultimate tool to help scientists, help us explore the universe around us".[1] At the other end of the spectrum are those concerned with the effects from growing reliance on AI for classroom work on children's ability to "learn how to learn" or from task automation on novice workers acquiring relevant occupation- or task-specific skills.[2]

The disagreements are in part about whether AI-provided information is a *complement* or a *substitute* to human learning. If the former, then expanding of AI will make humans put their effort and attention in where it matters and use AI's inputs with growing effectiveness. In the substitutes case, however, better and better AI will increasingly discourage human effort and learning—because most relevant information comes to be served to humans on a platter.

Proponents of both views can find evidence to support their position. AI's critical role in helping scientists design new proteins provides a well-known example of how these tools have already moved the dial in advancing human scientific knowledge (Jumper et al., 2021; Watson et al., 2023). It is not only in the scientific domain that AI is proving capable of complementing human decision-making. Using data from a leading firm in the industry, Brynjolfsson et al. (2025) have documented how AI has successfully helped customer service agents by providing them with relevant background information on the problem they are dealing with. At the other end of the ledger, several studies show how the use of AI is negatively impacting human learning and sometimes even broader capabilities. Lei et al. (2026) track 26,811 Chinese K–12 students' engagement with generative AI for 30 months and find that AI use significantly raises homework scores, but the same students have much worse scores in closed-book classroom exams and show significantly poorer performance in the high-stakes college and high-school entrance exams. In a smaller study, Kosmyna et al. (2025) look at the effects of using ChatGPT for writing assistance on mental capabilities and brain connectivity. They find that AI use reduces users' ability to memorize and accurately

[1] See, `https://www.nobelprize.org/prizes/chemistry/2024/hassabis/interview/`

[2] See, e.g., `https://www.edweek.org/technology/rising-use-of-ai-in-schools-comes-with-big-downsides-for-students/2025/10` and `https://www.weforum.org/stories/2025/04/ai-jobs-international-workers-day/`

record their own arguments, and remarkably, it appears to do so by changing neural connectivity in users' brains. The authors also document that the resulting output now looks much more similar across writers, raising concerns about the erosion of individual creativity and agency. Relatedly, Gerlich (2025) finds that large language models appear to reduce the creativity of users, especially among younger users. Budiyono et al. (2025) report similar findings in longer-term assessments, while Jakesch et al. (2023) report that collaboration with ChatGPT for writing affects people's opinions and beliefs significantly. Relatedly, several studies have found that AI recommendations are often misinterpreted by experts, for example, in radiology (Agarwal et al., 2023).

These differences may be an unavoidable byproduct of the reality that generative AI tools will be used in many fields, with potentially different effects. But they may also reflect the fact that some of the "equilibrium effects" of generative AI cannot be learned from a case-by-case empirical approach without an appropriate theoretical structure. For example, even in the substitutes case, individual decision-making could still improve from AI use, since the main reason human decision-makers put less effort is because they can now make good decisions without paying the costs. Hence, in the examples mentioned in the previous paragraph, each individual decision-maker could improve his or her performance in some well-defined tasks, even if this comes at the cost of the acquisition of general useful knowledge. Even more importantly, AI reliance may also lead to the depletion of collective knowledge. In fact, there is evidence that large language models have already adversely affected collective knowledge-building in a number of settings. For instance, in the context of the coding Q&A platform Stack Overflow, there is now less activity, less human engagement with questions, and arguably less generation of new knowledge as experienced coders grapple with new questions. The evidence indicates that this is in response to people turning to generative AI tools to get that same help (del Rio-Chanona et al., 2024). The pattern seems to be similar for Wikipedia, with less article reading and generation in areas where ChatGPT is an effective substitute to Wikipedia (Lyu et al., 2025).

This paper is an attempt to contribute to a better theoretical understanding of how AI tools impact human cognition and knowledge. We build a dynamic model of learning and decision-making where AI inputs can be either complementary or substitutable to human effort. At the center of our approach is a distinction between two types of information: general and individual- (or context-) specific. To perform any task, individuals require general knowledge. For example, for investment decisions one needs a basic understanding of different financial instruments such as treasury bonds, corporate bonds, stocks, options, etc., as well as information on how world stock markets and economies have been performing, some relevant aspects of their institutional structure, an understanding of macroeconomic risks etc. But one also needs information related to an individual's *context*: what is the risk tolerance and planning horizon of the individual in question? What correlation is there between their other income sources and different asset returns? Do they have information, hunches, preferences or beliefs affecting how they should invest and what types of risks they should take? And so on. Notably, human decision-makers often acquire both general and specific knowledge jointly. For example, most individuals will learn about general

financial knowledge in a finance course or reading relevant financial literature, and they will come to recognize their own needs and form their preferences and beliefs relevant for investment during the same process. Put differently, often there is *economies of scope* in learning, with the same efforts generating both general and individual- or *context-specific knowledge*.

Like humans, generative AI tools can also acquire both general and context-specific knowledge. But it is their ability—especially the promise of the much-anticipated *agentic AI models*—to develop and provide context-specific information and decision support that is most promising. Building on their architecture that enables the storage and rapid inspection of vast amounts of data and inference on connections between context and relevant information, these models are promising to find patterns that are relevant to a specific context and uniquely useful for individual decision-makers. Of course, AI models are often capable of doing many things, including the aggregation of existing collective knowledge. At some level, however, the Internet already aggregated a huge amount of general knowledge that was out there, and current generative AI tools are already improving on this aggregation role. The big next step—with both great potential and danger—is the individual and context-specific aid from AI models. It is this type of context-specific recommendation from agentic AI that we focus on in our analysis.[3] Returning to the investment example, textbooks and online resources can teach the mechanics of a broad range of financial instruments, but a future agentic system might translate an individual's particular context into a concrete portfolio choice or even autonomously execute trades subject to that person's conditions and constraints.

A key premise underlying our approach is that good prediction or performance and tasks typically requires *both* general knowledge and context-specific knowledge, and that these inputs are *complements* in the production of successful decisions. General knowledge makes context-specific evidence interpretable and valuable; conversely, context-specific knowledge pinpoints where the decision-maker sits within that general framework. Another important feature of general knowledge is that it builds on an entire community's learning efforts—by its nature general knowledge can be shared. This implies that most of the general knowledge an individual generates is an externality that he or she does not internalize. Consequently, the main individual motivation for exerting effort comes from context-specific learning. These elements imply that additional general knowledge raises the marginal return to an individual's learning effort: with more general knowledge, the same unit of effort is more beneficial for understanding and usefully acting on an individual's specific context, and this motivates more effort when there is more abundant general knowledge. Thus general knowledge is *complementary* to human learning effort. By the same token, however, with context-specific recommendations from agentic AI, there will be less impetus to exert costly effort, because one of the objectives of this effort is already well served by agentic AI. Consequently, agentic recommendations are a *substitute* for human effort.

Our model embeds this relationship between different types of knowledge and individual learning

[3]Agentic AI models, too, have many capabilities and uses, and some of these are very different than our focus here. Nevertheless, many of the things that agentic AI targets differ from other uses of generative AI in enabling greater customization and engagement with context-specific conditions and needs.

effort into a dynamic model of community learning. An additional important element is that community-level knowledge is itself an input into the AI models—without sufficient human effort, experiments and discovery there would not be enough valuable information for AI models to aggregate and sift through for either distilling general knowledge or for making individual-specific recommendations.[4]

We model the decision problem of a collection of human agents as a prediction problem—each agent's payoff depends on the distance of a common state representing general knowledge and their prediction about this state, and on the distance of their prediction about the context- state from its true value. In making these predictions, agents use their own learning effort, which generates two signals, one correlated with their context-specific state and another one correlated with the common state. By virtue of being about the common state, the latter signal is useful to all decision-makers, thus generating learning externalities. Agentic AI also provides context-specific recommendations, again modeled as predictions, which agents optimally combine with their own signals and priors.

The main result of the paper is a cautionary one: a powerful agentic AI model can *statically* help human decision-makers, but it can *dynamically* harm collective knowledge building. In fact, it can lead to what we call "knowledge collapse" whereby in the long-run equilibrium all (useful) human knowledge is ultimately destroyed. Our analysis provides a sharp characterization of the conditions for these negative outcomes.

Understanding the intuition for why agentic AI can harm human learning helps clarify our main contribution and the interpretation of our results. It is not surprising that, statically, individuals gain from additional information, in the absence of any misspecification in their model of the world or other biases. They only reduce their effort because they already receive a fairly good recommendation from agentic AI, which is naturally a substitute to their effort. This substitution when done optimally cannot harm their utility statically. However, human effort dynamically feeds into collective knowledge, and this externality is not internalized by the agents. As people reduce their learning effort, the amount of information that the community and AI models can aggregate starts diminishing. We assume that useful collective knowledge needs to be replenished constantly (because human environment and needs change over time). As a consequence, if there is not enough human learning effort, (useful) collective knowledge can collapse. Whether this happens in equilibrium depends on whether the cost of human learning from a small collective knowledge base becomes very low rapidly or not. In the former case, knowledge collapse can be averted, because humans will continue to put enough effort to generate some amount of collective knowledge. In the latter case, society may ultimately end up heading toward zero collective knowledge.

Even when knowledge collapse is averted, there is under-provision of human effort because of the learning externality. Long-run welfare depends on the balance of better static decision-making and the dynamic effects of under-provision of human learning effort on collective knowledge. We

[4]We discuss later the possibility of "synthetic data," whereby AI models generate their own data and learning, without any need for human effort, knowledge or data-generation.

show that the negative dynamic effect is likely to dominate when agentic AI is very accurate. On the other hand, the ability of agents to learn from the general knowledge of others—either via community aggregation or more traditional AI or Internet-type tools—always improves welfare.

These same comparative statics also apply to the likelihood that the system converges to the knowledge-collapse steady state. Specifically, in domains where learning effort is sufficiently elastic, the system can exhibit multiple steady states: a high-knowledge one and a knowledge-collapse trap with zero general knowledge. As agentic AI improves, the basin of attraction of the knowledge-collapse steady state expands. Moreover, once agentic AI is accurate enough, a "complete collapse" occurs in which the high-knowledge steady state disappears and the system converges to the knowledge-collapse trap regardless of initial conditions. On the other hand, better aggregation of human-produced general knowledge has the opposite effect: it shrinks the basin of attraction of the knowledge-collapse steady state and can offset part of the discouragement effect of agentic AI.

Because agentic recommendations operate purely through information provision, these results motivate regulation policies using information design. We show that limiting the effective precision of agentic recommendations via deliberate garbling can preserve learning incentives and prevent knowledge collapse. Specifically, we characterize an optimal two-phase garbling policy that maximizes long-run welfare. The first phase puts severe restrictions on the use of agentic AI in order to push society out of the basin of attraction of knowledge collapse, while the second phase garbles part of the information or the capabilities of these AI models in order to increase human learning effort.

We also consider several extensions to show both the flexibility of the model and the robustness of our main findings. First, in our baseline model, we assume that even without AI, human-generated general knowledge is relatively well aggregated and thus AI does not lead to a significant improvement in the aggregation of general knowledge. In our first extension we relax this assumption and establish that similar results apply even when AI simultaneously improves the aggregation of general knowledge relative to what the community was able to do and introduces the agentic element of providing individual specific recommendations to human decision-makers. Our results readily generalize to this case. Second, we investigate the extent to which "synthetic data," generated by AI models without relying on human learning and experimentation can substitute for human effort. We show that the same qualitative insights apply even with synthetic data, provided that such data is not a perfect substitute to human effort. However, in this case the knowledge-collapse steady state still features some positive amount of general information about the common state, since even without human effort synthetic data generates new knowledge. Third, we study a version of the model in which individuals can decide the direction of their learning effort, determining the balance between acquiring general knowledge and context-specific knowledge. Provided that general and context-specific knowledge cannot be perfectly separated, our results continue to apply in this case.

The rest of the paper is organized as follows. In the next section we provide a brief overview of the related literature. Section 3 introduces our basic model and contains our main analysis and most important results. Section 4 characterizes the welfare effects of AI. Section 5 presents the

extensions discussed above, while Section 6 concludes. The Appendix contains the proofs of the results stated in the text as well as further extensions.

2 Related Literature

Our main contribution is to provide, to the best of our knowledge, the first theoretical analysis of collective learning dynamics in the presence of AI agents and establish the possibility that AI, even as it helps static decision-making, could harm long-run collective knowledge building. Nevertheless, there are several literatures related to our work.

Most closely related to our paper are a few works studying AI's impact on human information acquisition. Ide (2025) develops a model in which firms can decide to automate entry-level tasks but this can negatively affect long-term growth because it hampers the intergenerational transmission of tacit knowledge, typically taking place within firms via novice-expert interactions. Agarwal et al. (2025) consider human-AI collaboration in classification tasks and emphasize how AI input is a substitute to human effort. They discuss how optimal collaboration schemes can be constructed via a mechanism design approach. In a complementary principal-agent framework, Bastani et al. (2024) show that more reliable AI can reduce human oversight effort and harm welfare in human-AI teams, because motivating costly human inspection of rare AI errors becomes prohibitively expensive, leading the principal to abandon human-AI collaboration or to prefer a less reliable AI tool. None of these papers nor any other ones that we are aware of consider the effects of AI on collective knowledge acquisition when individual learning depends on the efforts of others. Nor do they feature the key element of our approach, based on the possibility that AI input can be either a complement or substitute to human learning efforts.

Also related are a few works that discuss the impact of recent digital technologies on collective information or behavior, though in very different settings. Acemoglu et al. (2024) and Dasaratha and He (2022) analyze the implications of social media engagement on learning. Agrawal et al. (2023), Ide and Talamàs (2025) and Cullen et al. (2025) study the effects of AI on knowledge sharing, organizational structure and productivity in firms. Recent work by Farboodi et al. (2025) extend the task-based automation framework in Acemoglu and Restrepo (2018, 2019) so that data generated by both AI and human workers becomes an input into further AI-based automation. Hadfield and Koh (2025) provide a general discussion of the economic effects of AI agents.

The emphasis on the collective nature of human learning goes back to the early classic work by Nelson (1959) and Arrow (1962a,b). These ideas are also at the root of the recent literature on collective experimentation, for example, Bolton and Harris (1999), Keller et al. (2005) and Keller and Rady (2010), and the literature on social learning (from others' actions), including Banerjee (1992), Bikhchandani et al. (1992) and Smith and Sørensen (2000). More recent contributions in this literature, such as Acemoglu et al. (2011) and Dasaratha and He (2019) also note how the observation structure determines individual and collective learning. None of these papers discuss

the role of outside aggregators, such as AI, in collective knowledge-building.

We are also related to the literature on the conditions under which signals are complements or substitutes and the broader data externalities that arise within that context. See, among others, Börgers et al. (2013) and Brooks et al. (2024) on comparisons of signals, and Acemoglu et al. (2022) and Bergemann et al. (2022) on data and informational externalities.

Finally, there is an emerging literature within computer science investigating the individual and social effects of AI tools, and some of the relevant papers within this literature were already cited in the Introduction.

3 Model

In this section, we introduce our model, provide the characterization of equilibria, establishing how a knowledge-collapse steady state may emerge as a result of agentic AI, and present several comparative statics.

3.1 Environment

Time is discrete, $t = 1, 2, \ldots$. Each period t, there is a continuum of short-lived agents of total mass $M > 0$, indexed by $i \in [0, M)$. The population is partitioned into several islands such that *general knowledge* is only aggregated and shared within each island. Specifically, partition $[0, M)$ into islands of equal mass I: for $m = 0, 1, \ldots, K-1$ with $K := M/I$, island I_m is

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

For each island I_m, general knowledge learned by agents within this island is aggregated and shared with all future cohorts within this island. Here, I proxies for the ability of agents to learn from the general knowledge of others—either via community aggregation, more traditional AI, or Internet-type tools. The island structure enables us to have a tractable parameterization of the extent of general knowledge aggregation. The relevant knowledge circulates within island communities, approximating professional networks, language groups, or users of a particular platform. A greater I (holding the total populations M fixed) corresponds to larger communities and better aggregation of general knowledge. In particular, when $I = M$, society consists of a single island and all general knowledge is perfectly shared across the entire population, while with small I, there is much less sharing of general knowledge.

There are two latent states of the world, representing two types of knowledge of the society:

- A common state θ_t evolves as a random walk:

$$\theta_{t+1} = \theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma^2), \qquad \theta_1 \sim \mathcal{N}(0, \Sigma_0^2).$$

  We interpret this common state as general knowledge that is broadly useful across the community. For example, in the context of investment decisions it may capture a basic understanding of financial instruments and market institutions, while in the context of medical decisions it may represent basic knowledge on anatomy, bacteria, viruses, diseases and standard treatments. We model the common state as *evolving* over time because the frontier of general knowledge is not static: technological innovation can introduce new financial instruments and practices, and environmental or epidemiological change can lead to new diseases and therapies. The random-walk specification provides a parsimonious way to capture this continual evolution. This feature implies that in the absence of constant learning effort, effective collective knowledge will diminish, because past learning about θ_t is only imperfectly informative about θ_{t+1} and future values.

- An idiosyncratic state for each agent i, independently drawn each period,

$$\theta_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2).$$

  We interpret $\theta_{i,t}$ as context-specific knowledge that is unique to agent i in period t. For example, in the context of investment decisions it may capture the individual's risk tolerance, planning horizon, and exposure to nonfinancial income risk; in medical decision-making it may represent patient-specific characteristics such as symptoms and physical condition. Because this information is inherently individualized and (in this stylized formulation) does not persist in a systematic way from one period to the next, we model $\theta_{i,t}$ as an independent draw from a fixed distribution.

For agent i in period t, we model her task as predicting both θ_t (the common state in that period) and $\theta_{i,t}$ (her idiosyncratic state in the same period). Denote these predictions by $x_{i,t}$ and $y_{i,t}$, respectively. Output depends on the two prediction errors $|x_{i,t} - \theta_t|$ and $|y_{i,t} - \theta_{i,t}|$. For simplicity, we assume output depends only on whether each prediction falls within a unit tolerance band:

$$f(\mathbf{1}\{|x_{i,t} - \theta_t| \leq 1\},\ \mathbf{1}\{|y_{i,t} - \theta_{i,t}| \leq 1\}),$$

where $f : \{0,1\}^2 \to \mathbb{R}$ is a weakly increasing function, normalized so that $f(1,1) - f(0,0) = 1$. The fact that payoff depends on whether the prediction errors are less than 1 is entirely for simplicity. Our qualitative results continue to hold when payoffs are defined over other metrics of prediction errors and using other functional forms.

The following re-parameterization helps isolate the effects of the two ingredients—general and

context-specific knowledge—in production:

$$\begin{aligned}\Delta_{\mathrm{G}} &:= f(1,0) - f(0,0),\\ \Delta_{\mathrm{I}} &:= f(0,1) - f(0,0),\\ \Delta_{\mathrm{X}} &:= f(1,1) - f(1,0) - f(0,1) + f(0,0).\end{aligned}$$

These objects represent, respectively, the gains from general knowledge, from context-specific knowledge, and the additional gain from their complementarity. The normalization $f(1,1)-f(0,0) = 1$ implies

$$\Delta_{\mathrm{G}} + \Delta_{\mathrm{I}} + \Delta_{\mathrm{X}} = 1.$$

Throughout, we impose the following assumption on the production function:

Assumption 1. $\Delta_{\mathrm{I}} = 0$ *and* $\Delta_{\mathrm{X}} > 0$.

Here, $\Delta_{\mathrm{I}} = 0$ captures environments where context-specific knowledge alone, without general knowledge, creates no additional value. For example, a doctor who observes a patient's symptoms may still be unable to provide effective treatment without a correct understanding of the disease mechanisms. Put differently, $\Delta_{\mathrm{I}} = 0$ captures a strong complementarity between general and context-specific knowledge, as motivated the Introduction. Indeed, since $\Delta_{\mathrm{I}} + \Delta_{\mathrm{X}} = f(1,1) - f(1,0) \geq 0$ (by monotonicity of f), it follows that $\Delta_{\mathrm{X}} \geq 0$ whenever $\Delta_{\mathrm{I}} = 0$. To rule out the trivial case $\Delta_{\mathrm{I}} = \Delta_{\mathrm{X}} = 0$ (under which agents would have no incentive to learn), we additionally impose $\Delta_{\mathrm{X}} > 0$. Additionally, we allow for $\Delta_{\mathrm{G}} > 0$: general knowledge alone may sometimes create value; for example, a doctor with correct basic medical knowledge may be able to give some rudimentary treatment even without a careful examination of the patient.

A classical example satisfying Assumption 1 is the Leontief production function

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

for which

$$\Delta_{\mathrm{I}} = \Delta_{\mathrm{G}} = 0, \Delta_{\mathrm{X}} = 1.$$

Before undertaking the tasks, each agent chooses an effort level $e_{i,t} \geq 0$ and incurs a cost

$$\frac{\varepsilon}{\varepsilon+1} e_{i,t}^{\frac{\varepsilon+1}{\varepsilon}},$$

where $\varepsilon > 0$ denotes the constant (Frisch) elasticity of effort supply. The agent's payoff is given by task output minus effort cost:

$$u_{i,t} = f\Big(\mathbf{1}\{|x_{i,t} - \theta_t| \leq 1\},\ \mathbf{1}\{|y_{i,t} - \theta_{i,t}| \leq 1\}\Big) - \frac{\varepsilon}{\varepsilon+1} e_{i,t}^{\frac{\varepsilon+1}{\varepsilon}}.$$

We first describe the information environment in the world *without* agentic AI. As discussed above,

an agent's learning effort generates two types of signals:

- **Private Learning.** Agent i's effort produces a private signal about her idiosyncratic state $\theta_{i,t}$, with precision proportional to her effort:

$$s_{i,t}^{H} \sim \mathcal{N}\left(\theta_{i,t},\ \frac{1}{\lambda_{\mathrm{I}} e_{i,t}}\right),$$

  where $\lambda_{\mathrm{I}} > 0$ is a technological parameter.

- **Public Learning.** An agent's effort also generates general knowledge—information about the common state θ_t. In the baseline model, we assume that such general knowledge is already well aggregated within islands. We also simplify the analysis by assuming that each agent is atomistic, so individual contributions to the (public) stock of general knowledge are infinitesimal, but the aggregation of many such contributions is still finite.[5] Specifically, agent i's effort yields a signal about θ_t with precision proportional to $\lambda_{\mathrm{G}} e_{i,t}$, where $\lambda_{\mathrm{G}} > 0$ is a technology parameter. Signals produced within island m are aggregated—through community observation, the Internet, or more traditional AI—into a public social signal

$$s_{m,t}^{C} \sim \mathcal{N}\left(\theta_t, \frac{1}{\lambda_{\mathrm{G}} E_{m,t}}\right), \tag{1}$$

  where $E_{m,t}$ is aggregate effort on island m:

$$E_{m,t} := \int_{I_m} e_{i,t} \mathrm{d}i. \tag{2}$$

  This public signal $s_{m,t}^{C}$ is shared with all future cohorts on island m. The specification implies constant informativeness per unit of aggregate effort.

Because each individual's contribution is infinitesimal, the public-learning component of effort does not generate private incentives, and individual incentives come entirely from context-specific learning. We view this as a natural benchmark since learning externalities from general knowledge are typically much larger than an individual's utility from that general knowledge. For instance, a scientist's contribution to basic knowledge is typically used by hundreds of thousands and sometimes millions of people. The same is true in more applied domains as well. For example, when a software engineer diagnoses a rare production bug, the main private payoff comes from restoring their own system (a highly context-specific fix), whereas turning the insight into a clear, maintained public write-up (e.g., on Stack Overflow) would benefit many future developers, but yields only a small private return.

[5] These assumptions are relaxed in Section 5.1, where we allow AI advances to also improve aggregation of general knowledge.

Let $m(i)$ denote the island of agent i. In the pre-AI world, each agent observes her private signal about the idiosyncratic state as well as the history of public signals previously aggregated on her island. Formally, agent i's pre-AI information set and the public history on island m up to time t are

$$\mathcal{I}_{i,t}^{\text{pre-AI}} := \left\{ s_{i,t}^{H},\ \{s_{m(i),t'}^{C}\}_{1\le t'<t} \right\}, \qquad \mathcal{I}_{m,t}^{\text{pre-AI}} := \{s_{m,t'}^{C}\}_{1\le t'<t}.$$

Thus, even though the underlying common state θ_t is shared across islands, agents on different islands condition on different public histories.

We now introduce agentic AI. As discussed in the Introduction, we model agentic AI as an information technology that provides individual, context-specific information. For example, such an agentic system might provide personalized medical recommendations or investment suggestions tailored to the user. Accordingly, we model agentic AI as providing each agent a signal about her idiosyncratic state:

$$s_{i,t}^{A} \sim \mathcal{N}\left(\theta_{i,t},\ \frac{1}{\tau_{\text{A}}}\right),$$

where $\tau_A \ge 0$ captures the accuracy (precision) of agentic AI.

In the baseline model, agentic AI neither provides information about the common state directly nor improves the technology by which dispersed human efforts are pooled into the public knowledge. This assumption is an approximation to a world in which general knowledge is already well aggregated via in-person social learning and other pre-AI tools. As a result, agentic AI does not improve the available general knowledge for decision-making. In the medical example, for instance, this interpretation would be aligned with the case in which agentic AI does not invent new treatments or discover new diseases, and does not inform specialists about important general knowledge. In Section 5.1, we allow improvements in AI capabilities to also raise the effective aggregation capacity of society, while in Section 5.2, we consider AI-generated synthetic data as a substitute for human effort in building collective knowledge. Provided these additional channels are not too strong, the forces emphasized in the baseline model continue to govern the long-run dynamics of collective knowledge, as we will see.

Formally, the post-AI individual and public information sets are

$$\mathcal{I}_{i,t}^{\text{post-AI}} := \left\{ s_{i,t}^{H},\ s_{i,t}^{A}, \{s_{m(i),t'}^{C}\}_{1\le t'<t} \right\}, \quad \text{and} \quad \mathcal{I}_{m,t}^{\text{post-AI}} := \{s_{m,t'}^{C}\}_{1\le t'<t}.$$

Notably, the public information available on each island is unchanged by the introduction of agentic AI in the baseline model.

The within-period timing is as follows: in period t,

(i) Each agent i observes public history $\{s_{m(i),t'}^{C}\}_{1\le t'<t}$ on their island and chooses $e_{i,t}$. (Since past cohorts' effort profiles are uniquely pinned down in equilibrium, whether the current cohort observes past efforts is irrelevant.)

(ii) Each agent i observes private signal $s_{i,t}^H$.

(iii) *(Post-AI only)* Each agent i receives agentic AI recommendation $s_{i,t}^A$.

(iv) Each agent i forms prediction $(x_{i,t}, y_{i,t})$ and receives utility equal to output minus effort cost.

(v) The community aggregates general-knowledge signals within each island m and produces the public signal $s_{m,t}^C$, which becomes available for all members of island m from cohort-t.

We will show that the unique dynamic equilibrium is symmetric across islands and agents. In that symmetric equilibrium, effort $e_{i,t}$ is identical across individuals and islands (which are of equal size), so $E_{m,t} = E_t$ and the induced precision in (1) is the same for all m. Therefore, from this point onward we focus, without loss of any generality, on a representative island (say $m = 0$) and drop the island index. We therefore write E_t for $E_{m=0,t}$, s_t^C for $s_{m=0,t}^C$, and $\mathcal{I}_t$ for $\mathcal{I}_{m=0,t}$. We also focus on the behavior of a single agent, which represents, by symmetry, the behavior all other agent on the same island.

It is also useful to observe that the pre-AI environment is equivalent to a post-AI one with agentic AI precision $\tau_{\mathrm{A}} = 0$. Therefore, in the subsequent sections we will focus on the post-AI environment, treat pre-AI world as a special example where $\tau_{\mathrm{A}} = 0$, and drop the post-AI superscript for information sets.

3.2 Belief Updates and Knowledge

Let us define

$$V_t := \mathrm{Var}(\theta_t \mid \mathcal{I}_t)$$

as the posterior variance of the common state θ_t given the public history up to time t. The corresponding *public precision*

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

summarizes the stock of general knowledge (which is available to all individuals in society). Because signals are Gaussian, the path $\{X_t\}_{t\geq 1}$ does not depend on the realizations of the public signals, and is instead pinned down by the path of aggregate efforts $\{E_t\}_{t\geq 1}$.

In what follows, we focus on this public precision X_t, since it fully summarizes the relevant state of collective knowledge. Greater values of X_t correspond to more collective knowledge, and knowledge collapse results when $X_t \downarrow 0$.

Standard Kalman filtering implies the following law of motion for the precision of the public signal, which measures the stock of general knowledge. At the end of period t, aggregate effort E_t generates the public signal $s_t^C \sim \mathcal{N}\left(\theta_t, (\lambda_{\mathrm{G}} E_t)^{-1}\right)$. Combining this signal with the prior variance V_t yields posterior variance $\left(V_t^{-1} + \lambda_{\mathrm{G}} E_t\right)^{-1}$ for θ_t, and to this, the state drift $\theta_{t+1} = \theta_t + \epsilon_t$ adds

variance Σ^2. Hence,

$$V_{t+1} = \mathrm{Var}(\theta_{t+1} \mid \mathcal{I}_{t+1}) = \Big(V_t^{-1} + \lambda_{\mathrm{G}} E_t\Big)^{-1} + \Sigma^2.$$

This cohort's aggregate effort E_t raises next cohort's public precision V_{t+1}^{-1}; the innovation variance Σ^2 dilutes it between periods.

Equivalently, public precision evolves according to

$$X_{t+1}^{-1} = \Big(X_t + \lambda_{\mathrm{G}} E_t\Big)^{-1} + \Sigma^2. \tag{3}$$

Next, define agent i's *idiosyncratic precision* as

$$Y_{i,t} := \mathrm{Var}(\theta_{i,t} \mid \mathcal{I}_{i,t})^{-1} = \sigma^{-2} + \lambda_{\mathrm{I}} e_{i,t} + \tau_{\mathrm{A}}, \tag{4}$$

which is the sum of (i) prior precision σ^{-2}, (ii) precision coming from private learning $\lambda_{\mathrm{I}} e_{i,t}$, and (iii) precision from agentic AI recommendation τ_{A}.

3.3 Definition of Equilibrium

A (Perfect Bayesian) equilibrium consists of effort choices $\{e_{i,t}\}$ and predictions $\{(x_{i,t}, y_{i,t})\}$ of all agents such that for every agent i and period t, the chosen $e_{i,t}$ and $(x_{i,t}, y_{i,t})$ maximize expected utility $\mathbb{E}[u_{i,t}]$, where

$$u_{i,t} = f\Big(\mathbf{1}\{|x_{i,t} - \theta_t| \le 1\},\ \mathbf{1}\{|y_{i,t} - \theta_{i,t}| \le 1\}\Big) - \frac{\varepsilon}{\varepsilon + 1} e_{i,t}^{\frac{\varepsilon+1}{\varepsilon}}.$$

Before characterizing the dynamic equilibrium, we zoom in on a single agent's within-period behavior. Because agents are short-lived and f is weakly increasing in each of its argument, in equilibrium agents make one-step ahead *Bayesian* predictions, choosing the posterior means:

$$x_{i,t} = \mathbb{E}[\theta_t \mid \mathcal{I}_{i,t}], \qquad y_{i,t} = \mathbb{E}[\theta_{i,t} \mid \mathcal{I}_{i,t}]. \tag{5}$$

Given (5), expected utility is pinned down by the agent's posterior beliefs, summarized by the public precision X_t and the idiosyncratic precision $Y_{i,t}$.

For a zero-mean Gaussian random variable $Z \sim \mathcal{N}(0, \tau^{-1})$ with precision $\tau > 0$, let us define

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

Under (5), we have $x_{i,t} - \theta_t \sim \mathcal{N}(0, \mathrm{Var}(\theta_t \mid \mathcal{I}_t)) = \mathcal{N}(0, X_t^{-1})$, which implies $\mathbb{P}\left(|x_{i,t} - \theta_t| \le 1\right) = G(X_t)$. Similarly, $\mathbb{P}\left(|y_{i,t} - \theta_{i,t}| \le 1\right) = G(Y_{i,t})$. Therefore, using the fact that $\Delta_{\mathrm{I}} = 0$ (individual knowledge generates no value without general knowledge), agent i's expected period-t utility can be

written as:

$$U_{i,t} := \mathbb{E}\big[u_{i,t}\big] = f(0,0) + \underbrace{G(X_t)\cdot\Delta_{\mathrm{G}}}_{\text{General Knowledge}} + \underbrace{G(X_t)G(Y_{i,t})\cdot\Delta_{\mathrm{X}}}_{\text{Complementarity}} - \underbrace{\frac{\varepsilon}{\varepsilon+1}e_{i,t}^{\frac{\varepsilon+1}{\varepsilon}}}_{\text{Cost}}. \tag{6}$$

3.4 Substitutes and Complements

As discussed in the Introduction, a central theme of the paper is that information provided by generative AI can either *complement* human learning (by making human effort more valuable) or *substitute* for it (by directly providing the information that effort would have produced), depending on whether the provided information is general or context-specific. In our model, these two types of information map into two distinct objects: the inherited stock of public precision X_t, measuring general knowledge, and the accuracy of context-specific, agentic recommendation τ_A. The expected payoff in (6) clarifies these two forces.

Observation 1. *Public precision X_t complements human effort, while agentic-AI precision τ_A substitutes for human effort. That is:*

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

Here, we first explain why general knowledge, represented by public precision X_t, is a complement to human effort. Under Assumption 1 ($\Delta_I = 0$ and $\Delta_X > 0$), an agent's private return to effort comes entirely through the complementarity term

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

Learning effort raises $Y_{i,t}$, and its marginal benefit is proportional to $G(X_t)$: when the stock of general knowledge is higher, improving context-specific precision is more valuable because the two inputs are complements in task production, and achieving better context-specific precision is the main reason why individuals are exerting learning effort (the general knowledge that they generate is an externality that they do not internalize). Formally,

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

This means that individual learning effort is more beneficial when there is more general knowledge (higher X_t) that can be applied more productively to one's own context.

On the other hand, agentic-AI precision enters only through $Y_{i,t}$, directly improving context-specific precision. Because the probability of success $G(Y)$ exhibits diminishing returns in precision,

higher τ_A reduces the marginal gain from additional human effort:

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

Thus, better agentic recommendations crowd out learning effort by making incremental self-acquired precision less valuable at the margin.

Because X_t is pinned down by past cohorts' equilibrium efforts (and by the island's aggregation capacity I), Observation 1 implies that:

1. Higher past effort $\{e_{t'}\}_{t'<t}$ strengthens the public signal and encourages further effort.
2. Greater aggregation capacity (a larger I) strengthens the public signal, also encouraging effort.
3. Higher agentic-AI precision τ_{A} weakens the marginal return to effort, discouraging effort.

Overall, effort choices are *strategic complements* across cohorts, and stronger aggregation capacity amplifies this complementarity. By contrast, agentic AI is as a *substitute* for human learning effort.

3.5 Existence and Characterization of Equilibrium

Fix a period t and an agent i. Taking the current public precision X_t (pinned down by past cohorts' aggregate efforts on the island) and agentic-AI precision τ_{A} as given, agent i chooses effort $e_{i,t} \geq 0$ to maximize her expected utility $U_{i,t}$ in (6). Using (4), the first-order condition is

$$\frac{\partial U_{i,t}}{\partial e_{i,t}} = \Delta_{\mathrm{X}} G(X_t) \cdot \lambda_{\mathrm{I}} g\left(\sigma^{-2} + \lambda_{\mathrm{I}} e_{i,t} + \tau_{\mathrm{A}}\right) - e_{i,t}^{1/\varepsilon}.$$

The marginal benefit term is weakly decreasing in $e_{i,t}$ because $g(\cdot)$ is decreasing, while the marginal cost $e_{i,t}^{1/\varepsilon}$ is strictly increasing. Consequently, $\frac{\partial U_{i,t}}{\partial e_{i,t}}$ is strictly decreasing in $e_{i,t}$. Moreover, we have

$$\left.\frac{\partial U_{i,t}}{\partial e_{i,t}}\right|_{e_{i,t}=0} = \Delta_{\mathrm{X}} G(X_t) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \tau_{\mathrm{A}}) \geq 0, \qquad \lim_{e_{i,t} \to +\infty} \frac{\partial U_{i,t}}{\partial e_{i,t}} < 0,$$

so $U_{i,t}$ is strictly concave in effort and admits a unique and finite maximizer.[6]

This maximizer is common across all agents in cohort t: $e_{i,t} = e_t$ for all i. Let us then define the *best-response effort* $e(X, \tau_{\mathrm{A}})$ as the unique solution to the first-order condition:

$$e(X, \tau_{\mathrm{A}}) := \left\{ e \geq 0 \mid \Delta_{\mathrm{X}} G(X) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}}) = e^{1/\varepsilon} \right\},$$

[6] At the boundary $X_t = 0$, we have $G(X_t) = 0$, so effort yields no benefit and the unique best response is $e_{i,t} = 0$. This is because an individual cannot individually increase the stock of general knowledge and without general knowledge, context-specific knowledge is not useful. On the contrary, starting from any $X_1 > 0$ the equilibrium recursion ensures $X_t > 0$ for all finite t, and the best response effort is interior in every period along the equilibrium path.

so that in equilibrium, $e_{i,t} = e_t = e(X_t, \tau_A)$.

Observation 1 together with standard monotone comparative statics arguments (e.g., Topkis' theorem) yields:

Observation 2. *The best-response effort $e(X, \tau_A)$ is increasing in X and decreasing in τ_A, strictly so for all $X > 0$.*

Given the symmetric choice $e_t = e(X_t, \tau_A)$, aggregate effort on the island is

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

and the Kalman recursion (3) pins down next period's public precision uniquely. We summarize the equilibrium dynamics as follows. The proof of all the results stated in the text is presented in the Appendix.

Proposition 1 (Dynamic equilibrium)**.** *Fix an initial public precision $X_1 \geq 0$ implied by the prior. There exists a unique symmetric (Perfect Bayesian) equilibrium path with $e_{i,t} = e_t \geq 0$, where $\{X_t, e_t\}_{t=1}^{\infty}$ is the unique solution to the recursion*

$$\forall t, \quad \begin{cases} e_t = e(X_t, \tau_A), \\ X_{t+1} = \left[\Sigma^2 + \left(X_t + \lambda_G I e_t\right)^{-1}\right]^{-1}. \end{cases} \tag{7}$$

Substituting the best response into the law of motion in (7) yields a one-dimensional (one-step ahead) state transition map:

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

where

$$F(X_t) := \left[\Sigma^2 + \left(X_t + \lambda_G I e(X_t, \tau_A)\right)^{-1}\right]^{-1}. \tag{8}$$

The dynamic equilibrium thus reduces to iterating the map F on the state variable X_t, with effort e_t adjusting endogenously via the best response $e(X_t, \tau_A)$. This representation highlights two forces shaping the evolution of general knowledge.

First, there is knowledge accumulation: higher current effort e_t increases the precision of the public signal generated at the end of period t, raising next period's stock of public precision. This is captured by the $\lambda_G I e_t$ term inside F, which represents both the productivity of public learning (λ_G) and the island's aggregation capacity (I).

Second, there is knowledge depreciation: because the common state follows a random walk, what was learned in the past becomes less relevant and the magnitude of this effect is captured by the innovation variance Σ^2. For example, even if the current state were known with arbitrarily high precision, next period's variance would be no less than Σ^2. This is reflected in the fact that public

precision can never exceed Σ^{-2} and thus F maps into a bounded interval:

$$0 \leq F(X) < \Sigma^{-2} \quad \text{for all finite } X.$$

3.6 Steady-State Equilibria

A *steady-state equilibrium* can be represented by a scalar $\bar{X}$ satisfying the fixed-point condition of the state transition map F:

$$\bar{X} = F(\bar{X}).$$

In other words, a steady-state equilibrium can be represented by a level of public precision that reproduces itself from one cohort to the next.

The associated steady-state effort $\bar{e}$ and idiosyncratic precision $\bar{Y}$ are defined by

$$\bar{e} := e(\bar{X}, \tau_{\mathrm{A}}), \quad \bar{Y} := \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e} + \tau_{\mathrm{A}}.$$

For our purpose, an important type of steady state is the degenerate one where

$$(\bar{X}, \bar{e}, \bar{Y}) = (0, 0, \sigma^{-2} + \tau_{\mathrm{A}}). \tag{9}$$

This steady state corresponds to a situation in which the community accumulates essentially no usable general knowledge in the long run. Even though agentic AI may still deliver relevant individual- or context-specific information, this is not socially useful because over time general knowledge that is an essential input (from our assumption that $\Delta_{\mathrm{I}} = 0$) disappears and makes the context-specific information also useless. This is the reason why we refer to the steady state represented in (9) as the *knowledge-collapse* steady state (see Sections 5.1 and 5.2 for generalizations).

We next derive a number of basic properties of the state transition map F, which enables us to characterize the set of steady states and long-run dynamics.

Lemma 1. *The state transition map F is continuous and strictly increasing on $\mathbb{R}_{\geq 0}$, with*

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

Here, the monotonicity of F stems from the fundamental complementarity between general knowledge and returns to learning effort: a higher inherited stock of general knowledge X raises current effort $e(X, \tau_{\mathrm{A}})$, which strengthens the public signal and increases next period's public precision.

The mapping F depends on the aggregation scale I and agentic AI precision τ_{A} through the endogenous aggregate effort $I \cdot e(X, \tau_{\mathrm{A}})$. Observation 2 enables us to establish monotone comparative statics of F.

Proposition 2 (Comparative Statics of F)**.** *The function F is pointwise increasing in I and pointwise decreasing in τ_{A}, strictly so for all $X > 0$.*

Proposition 2 has a direct economic interpretation. A larger aggregation scale I makes each cohort's effort more productive because the same individual effort is pooled into a more informative public signal, and this translates into greater general knowledge in the next period. By contrast, a higher agentic-AI precision τ_{A} depresses effort today (since recommendations substitute for private learning), reducing the flow of new human-generated knowledge into the stock of general knowledge. It is important to recall that in our baseline model, agentic AI affects the evolution of collective knowledge only through human behavior—it does not directly generate general knowledge, but it can indirectly erode it by discouraging the effort that feeds public learning.

We next study the local stability of the knowledge-collapse steady state (the zero fixed point of F). Intuitively, when X is very small, the marginal return to effort is also small (since the complementarity term is scaled by $G(X)$), so a key question is whether agents still exert *enough* effort to prevent X from drifting toward zero. The answer depends on the elasticity of effort supply.

Lemma 2. *(i) If $\varepsilon < 4$, then $F(X) > X$ in a neighborhood of 0, so that the knowledge-collapse steady state* (9) *is locally unstable.*

(ii) If $\varepsilon > 4$, then $F(X) < X$ in a neighborhood of 0, so that the knowledge-collapse steady state(9) *is locally stable.*

Whether the zero steady state is locally stable comes down to comparing the effort needed to keep a tiny amount of public knowledge alive with the effort agents actually choose in equilibrium. For a small value of collective knowledge (that is, as $X \downarrow 0$), maintaining the same level of X in the next period requires effort of order $\Theta(X^2)$. By contrast, the best response yields effort $e(X, \tau_{\mathrm{A}}) = \Theta(X^{\frac{\varepsilon}{2}})$. Hence, when $\varepsilon > 4$, this exponent is larger than 2, so equilibrium effort falls faster than X^2 because in the vicinity of zero agents invest less than the "maintenance" level, and consequently any small positive X drifts back to zero. This implies that the knowledge-collapse (zero) steady state is locally stable. Conversely, when $\varepsilon < 4$, the exponent is below 2, so effort decays more slowly than X^2. In that case, for X close to zero, agents exert more than enough effort to offset drift, a small positive X tends to grow, and the zero state becomes locally unstable.

Lemma 2 identifies the elasticity of effort as a key determinant of whether knowledge collapse is a self-reinforcing trap. This is intuitive: when effort is not too elastic ($\varepsilon < 4$), agents do not cut effort too sharply as incentives weaken, which stabilizes collective knowledge accumulation and rules out collapse as a locally stable outcome. In contrast, when $\varepsilon > 4$, the system becomes fragile: a lower stock of general knowledge sharply reduces incentives, which then reduces effort and further depresses the stock of general knowledge. This opens the door to coordination failure and path dependence, as we will describe below.

The elasticity of effort can be influenced by technological and institutional factors. For example,

effort tends to be less elastic when institutions or policies impose a baseline level of learning and make it difficult for the amount of knowledge to exceed a certain upper bound. In the medical context, this may be the case because (i) high professional standards, such as minimum years of medical training, ensure such a baseline, and (ii) the enormity and difficulty of the relevant advanced material makes very high levels of effective effort challenging.

3.7 Unique Steady State Regime

We first provide a more detailed analysis of the regime in which effort is not too elastic, meaning that

$$\varepsilon < 4,$$

which (as we will show) ensures a unique steady state.

Because the state transition map $X_{t+1} = F(X_t)$ is strictly increasing and continuous (Lemma 1), steady states correspond to intersections of F with the 45-degree line. Lemma 2 implies that when $\varepsilon < 4$, the transition map F lies *above* the 45-degree line for X sufficiently close to zero, ensuring the zero fixed point is locally unstable. But also, $F(X)$ is bounded above by Σ^{-2} (Lemma 1), while the 45-degree line is unbounded, so $F(X) < X$ for all sufficiently large X. Therefore, by continuity, F must cross the 45-degree line at least once on $(0, \infty)$. In this regime, the state transition map also has a slope less than one whenever it crosses the 45° line so that the (non-trivial) steady state is *unique*.

Proposition 3. *Suppose $\varepsilon < 4$. Then there exist exactly two steady-state equilibria, $0 = \bar{X}_\ell < \bar{X}_h < +\infty$, where $\bar{X}_h$ is locally stable and $\bar{X}_\ell$ is locally unstable. Moreover, there is a single basin of attraction on $\mathbb{R}_+ \setminus \{0\}$. That is, for the sequence $\{X_t\}$ defined by $X_{t+1} = F(X_t)$ with $X_1 > 0$, we have $X_t \to \bar{X}_h$.*

Proposition 3 describes a "robust learning regime": although $X = 0$ is always a steady state (social knowledge can always remain at zero if it ever reaches it), it is a knife-edge outcome: any strictly positive stock of general knowledge, no matter how small, triggers enough learning effort to more than offset the depreciation coming from changes, and the economy converges to the unique high-knowledge steady state $\bar{X}_h$. As a result, there is no coordination failure: long-run outcomes do not depend on history (except for the degenerate initial condition $X_1 = 0$).

Define the steady-state effort and idiosyncratic precision associated with the high steady state:

$$\bar{e}_h := e(\bar{X}_h, \tau_{\mathrm{A}}), \qquad \bar{Y}_h := \sigma^{-2} + \lambda_{\mathrm{I}} \bar{e}_h + \tau_{\mathrm{A}}.$$

Proposition 4. *Suppose $\varepsilon < 4$. The following monotonicity statements hold strictly:*

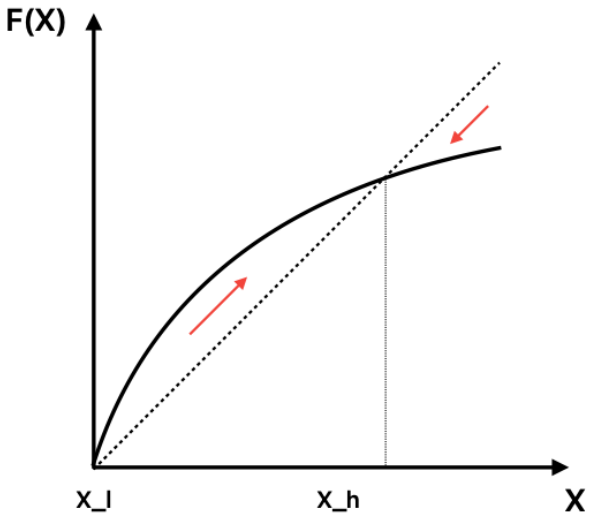


Figure 1: Equilibrium dynamics when $\varepsilon < 4$.

(i) As agentic AI precision τ_A increases:

$$\bar{X}_h \downarrow \ in\ \tau_A, \quad \bar{e}_h \downarrow \ in\ \tau_A, \qquad \bar{Y}_h \uparrow \ in\ \tau_A.$$

(ii) As the aggregation scale I increases:

$$\bar{X}_h \uparrow \ in\ I, \quad \bar{e}_h \uparrow \ in\ I, \qquad \bar{Y}_h \uparrow \ in\ I.$$

Part (i) highlights that, even in the robust-learning regime, more accurate agentic AI erodes collective knowledge. A higher τ_A reduces agents' incentive to invest in costly learning (Observation 2), lowering the flow of new human-generated information that feeds the stock of general knowledge and shifting the transition map F downward (Proposition 2). Because $\bar{X}_h$ is the unique interior fixed point, it must fall. At the same time, idiosyncratic precision $\bar{Y}_h$ rises mechanically with τ_A, even though the endogenous human component of $\bar{Y}_h$ (through $\bar{e}_h$) falls. Thus, agentic AI improves the context-specific information available at steady state while reducing the general-knowledge stock that sustains complementary human learning.

Part (ii) shows the opposite force: larger aggregation capacity strengthens the intertemporal complementarity across cohorts. Holding effort fixed, a larger I makes each cohort's contributions more informative; holding X fixed, it also raises equilibrium effort by increasing the social return to learning that agents inherit via X. Both channels shift F upward and raise the unique stable fixed point $\bar{X}_h$. In this sense, larger information-sharing communities (or better aggregation technologies) make the knowledge stock more resilient and push the economy toward a higher-knowledge steady state.

In summary, the condition $\varepsilon < 4$ implies that the long-run equilibrium is unique and the knowledge-collapse steady state is not an attractor. However, agentic AI can still reduce $\bar{X}_h$ (and thus the community's long-run stock of general knowledge), even though it cannot generically generate a self-sustaining collapse trap in this parameter range.

3.8 Multiple Steady States Regime and Knowledge Collapse

For our purposes, the more interesting part of the parameter space is the one where multiple steady states can arise, meaning

$$\varepsilon > 4,$$

in which case effort is highly elastic and, by Lemma 2, the zero-knowledge steady state $\bar{X} = 0$ is *locally stable*. This local stability opens the door to *multiple* stable long-run outcomes.

As in the case with $\varepsilon < 4$, steady states correspond to intersections of state transition map F with the 45-degree line. When $\varepsilon > 4$, the graph of F lies *below* the 45-degree line for X close to zero. Whether F crosses the 45-degree line again at higher X depends on how much effort agents exert and hence on the strength of substitution from agentic AI (captured by τ_{A}). As τ_{A} increases, the best-response effort $e(X, \tau_{\mathrm{A}})$ falls pointwise (Observation 2), shifting the entire map F downward and shrinking (or eliminating) the set of positive fixed points.

The next proposition formalizes the resulting two possibilities. Denote by τ_{A}^{c} the (endogenous) *complete-collapse threshold*, meaning the largest agentic-AI precision for which a positive steady state can be sustained.

Proposition 5. *Suppose $\varepsilon > 4$. There exists threshold $\tau_{\mathrm{A}}^{c} \geq 0$, such that*

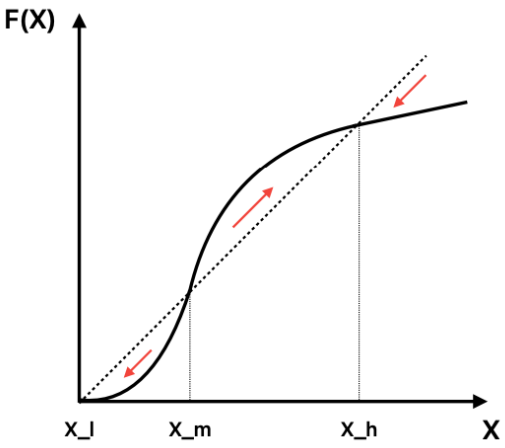
(i) If $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$, there exist three steady-state equilibria $0 = \bar{X}_{\ell} < \bar{X}_{\mathrm{m}} < \bar{X}_{\mathrm{h}} < +\infty$, where $\bar{X}_{\ell}$ and $\bar{X}_{\mathrm{h}}$ are locally stable, whereas $\bar{X}_{\mathrm{m}}$ is locally unstable.

The unstable point $\bar{X}_{\mathrm{m}}$ partitions the state space into two basins of attraction:

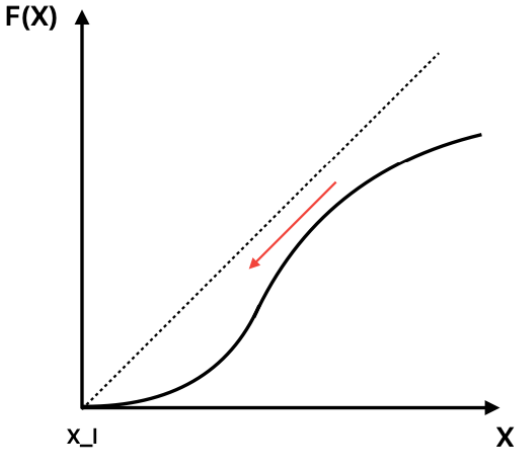
$$X_1 < \bar{X}_{\mathrm{m}} \;\Rightarrow\; X_t \to \bar{X}_{\ell} = 0$$
$$X_1 > \bar{X}_{\mathrm{m}} \;\Rightarrow\; X_t \to \bar{X}_{\mathrm{h}}$$

(ii) If $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^{c}$, then the unique steady-state equilibrium is the knowledge-collapse one, $\bar{X}_{\ell} = 0$, and this steady state is globally stable, meaning that $\forall X_1 \geq 0, \quad X_t \to \bar{X}_{\ell}$.

Figure 2 illustrates the two cases within this regime diagrammatically. In panel (a), the map F intersects the 45-degree line three times: the lowest and highest intersections are stable (slopes below one locally), while the middle intersection is unstable (slope above one). The unstable point $\bar{X}_{\mathrm{m}}$ therefore acts as a threshold: if inherited public precision falls below $\bar{X}_{\mathrm{m}}$, the complementarity



(a) $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$: multiple steady states



(b) $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^c$: complete collapse

Figure 2: Equilibrium dynamics when $\varepsilon > 4$.

between general knowledge and private effort becomes too weak, effort collapses gradually, and the community converges to $\bar{X}_\ell = 0$. If instead X_1 exceeds $\bar{X}_{\mathrm{m}}$, the positive feedback across cohorts dominates and the system converges to the high-knowledge steady state $\bar{X}_{\mathrm{h}}$. This configuration implies a type of *path-dependence*: where the initial knowledge stock determines whether or not society tends towards knowledge collapse.

Panel (b) illustrates the even more adverse regime (from the viewpoint of collective knowledge-building): when τ_{A} exceeds τ_{A}^c, the map F lies everywhere below the 45-degree line for $X > 0$, and there can be no positive steady states. In this case, the dynamic equilibrium converges to the knowledge-collapse steady state regardless of initial conditions. We refer to this phenomenon as complete collapse, and call τ_{A}^c the complete-collapse threshold.

Another interesting feature of Proposition 5 is that the long-run outcome can change discontinuously with τ_{A}. As $\tau_{\mathrm{A}} \uparrow \tau_{\mathrm{A}}^c$ from below, the two positive fixed points $\bar{X}_{\mathrm{m}}$ and $\bar{X}_{\mathrm{h}}$ move toward each other and eventually collide; at τ_{A}^c the positive steady states disappear, after which the only remaining attractor is $\bar{X}_\ell = 0$. Consequently, the long-run stock of general knowledge can exhibit a discontinuous drop as τ_{A} crosses τ_{A}^c.

We next establish comparative statics for the key quantities that emerged endogenously from the characterization: (i) complete-collapse threshold τ_{A}^c, (ii) the threshold for the basin of attraction, $\bar{X}_m$, and (iii) high-knowledge steady state $\bar{X}_h$. We first consider τ_{A}^c.[7]

Proposition 6. *Suppose $\varepsilon > 4$. Then:*

[7]When we increase Δ_{X}, Δ_{G} needs to be adjusted in order to satisfy the normalization $\Delta_{\mathrm{X}} + \Delta_{\mathrm{G}} = 1$. Loosely speaking, this can be interpreted as the source of higher payoffs being reallocated between complementarity and the pure effects of general knowledge. This formulation is adopted for simplicity, and our qualitative results do not depend on this normalization.

(i) τ_{A}^{c} *is increasing in aggregation capacity* I *(strictly so when* $\tau_{\mathrm{A}}^{c} > 0$*).*

(ii) τ_{A}^{c} *is increasing in the degree of complementarity* Δ_{X} *(strictly so when* $\tau_{\mathrm{A}}^{c} > 0$*).*

The remaining proofs in the paper are presented in the Online Appendix B.

The interpretation of Proposition 6 is straightforward: a larger I strengthens the social signal generated by any given level of per-capita effort, shifting the map F upwards, and thus making the knowledge stock more resilient to the effort-substitution effect of agentic AI. Likewise, a higher Δ_{X} raises the complementarity gain from jointly getting both the common state and the idiosyncratic state right, and thus encourages equilibrium effort for a given level of social knowledge stock, shifting the map F upward. In both cases, society can tolerate a higher level of agentic-AI accuracy before the high-knowledge steady state disappears.

When $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$, complete collapse does not occur, and whether the community converges to knowledge collapse or to the high-knowledge steady state depends on the initial stock of general knowledge exceeds the threshold $\bar{X}_{\mathrm{m}}$. The threshold $\bar{X}_{\mathrm{m}}$ can therefore be interpreted as a measure of *fragility* or the likelihood of coordination failure: the larger is $\bar{X}_{\mathrm{m}}$, the less likely a society to end in the high-knowledge steady state. We record the comparative statics for this threshold $\bar{X}_{\mathrm{m}}$ in the next proposition:

Proposition 7. *Suppose* $\varepsilon > 4$ *and* $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$*. The threshold for the basin of attraction,* $\bar{X}_{m}$*, is increasing in* τ_{A} *and decreasing in* I *and* Δ_{X}*.*

In sum, even when a high-knowledge steady state exists, more accurate agentic AI raises the minimum stock of general knowledge required for the community to remain on (or return to) the high-knowledge trajectory. By contrast, stronger aggregation (I) and stronger complementarity (Δ_{X}) reduce the threshold for the basin of attraction, $\bar{X}_{m}$, and make the high-knowledge equilibrium easier to sustain.

Finally, let $(\bar{X}_{\mathrm{h}}, \bar{e}_{\mathrm{h}}, \bar{Y}_{\mathrm{h}})$ denote the high-knowledge steady state when it exists, where

$$\bar{e}_{h} := e(\bar{X}_{h}, \tau_{\mathrm{A}}), \quad \bar{Y}_{h} := \sigma^{-2} + \lambda_{\mathrm{I}} \bar{e}_{h} + \tau_{\mathrm{A}}.$$

The next proposition summarizes how steady-state outcomes respond to τ_{A} and I.

Proposition 8. *Suppose* $\varepsilon > 4$ *and* $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$*. The following monotonicity statements hold strictly:*

(i) As agentic AI precision τ_{A} *increases:*

$$\bar{X}_{h} \downarrow \text{ in } \tau_{\mathrm{A}}, \quad \bar{e}_{h} \downarrow \text{ in } \tau_{\mathrm{A}}.$$

There exists $0 \leq \hat{\tau}_{A} < \tau_{\mathrm{A}}^{c}$ *such that*

$$\bar{Y}_{h} \uparrow \text{ in } \tau_{\mathrm{A}} \text{ when } \tau_{\mathrm{A}} \in [0, \hat{\tau}_{A}], \quad \bar{Y}_{h} \downarrow \text{ in } \tau_{\mathrm{A}} \text{ when } \tau_{\mathrm{A}} \in (\hat{\tau}_{A}, \tau_{\mathrm{A}}^{c}).$$

(ii) The aggregation scale I acts in the opposite direction:

$$\bar{X}_h \uparrow \text{ in } I, \quad \bar{e}_h \uparrow \text{ in } I, \quad \bar{Y}_h \uparrow \text{ in } I.$$

This proposition is similar to Proposition 4, with the only difference being the non-monotonicity of $\bar{Y}_{\mathrm{h}}$ when we increase τ_{A}. This non-monotonicity reflects two opposing channels: a higher τ_{A} directly improves context-specific precision, but indirectly reduces it by crowding out private learning ($\bar{e}_{\mathrm{h}}$). For relatively small τ_{A}, the direct informational benefit dominates and $\bar{Y}_{\mathrm{h}}$ rises. On the other hand, close to the collapse threshold, the system becomes fragile and the crowd-out effect can dominate, so that even context-specific precision falls with a more accurate agentic AI.

4 Welfare

In this section, we evaluate the effects of AI on long-run welfare. We take the relevant welfare object to be the steady-state expected utility of a representative cohort. We drop the constant term $f(0,0)$ throughout, so welfare is measured relative to the no-information baseline.

4.1 Steady-State Expected Utility

For any steady state $(\bar{X}, \bar{e}, \bar{Y})$, define steady-state expected utility as

$$\bar{U} := G(\bar{X}) \cdot \Delta_{\mathrm{G}} + G(\bar{X})G(\bar{Y}) \cdot \Delta_{\mathrm{X}} - \frac{\varepsilon}{\varepsilon+1}\bar{e}^{\frac{\varepsilon+1}{\varepsilon}}.$$

Under Assumption 1 ($\Delta_{\mathrm{I}} = 0$), the knowledge-collapse steady state $(\bar{X}, \bar{e}) = 0$ yields $\bar{U} = 0$.

Because $\bar{e}$ is the privately optimal effort given $(\bar{X}, \tau_A)$, welfare can be written as a function of just $(\bar{X}, \tau_A)$:

$$\bar{U}(\bar{X}, \tau_A) = G(\bar{X}) \cdot \Delta_{\mathrm{G}} + \max_{e \geq 0} \left[G(\sigma^{-2} + \lambda_{\mathrm{I}} e' + \tau_{\mathrm{A}}) \cdot G(\bar{X})\Delta_{\mathrm{X}} - \frac{\varepsilon}{\varepsilon+1}(e'^{\frac{\varepsilon+1}{\varepsilon}} \right]. \tag{10}$$

This representation separates (i) the effect of inherited stock of general knowledge $\bar{X}$, (ii) the effect of accuracy of agentic recommendation τ_{A}, and (iii) the within-cohort private optimization problem over effort, which trades off higher idiosyncratic precision against convex effort costs.

Whenever a positive high-knowledge steady state exists (as defined in Propositions 3 and 5), we let $(\bar{X}_h, \bar{e}_h, \bar{Y}_h)$ denote the associated steady-state triple and define the corresponding steady-state welfare by $\bar{U}^+ := \bar{U}(\bar{X}_h, \tau_{\mathrm{A}})$.

The realized long-run welfare depends on whether the equilibrium dynamics admit a unique (stable) steady state or exhibit multiplicity and path dependence.

(i) Unique steady state regime: $\varepsilon < 4$. From Proposition 3, for any initial condition $X_1 > 0$

the equilibrium converges to the unique high-knowledge steady state $\bar{X}_h$. Hence long-run welfare is uniquely pinned down and equals $\bar{U}^+$. In this regime, there is no equilibrium-selection problem: AI affects welfare only through how it shifts $(\bar{X}_h, \bar{e}_h, \bar{Y}_h)$.

(ii) Multiple steady states regime: $\varepsilon > 4$. In this regime, equilibrium dynamics can display coordination failure. If $\tau_{\mathrm{A}} \geq \tau_{\mathrm{A}}^c$, complete collapse occurs and long-run welfare is 0 for all initial conditions. If instead $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$, the knowledge-collapse steady state and the high-knowledge steady state coexist, separated by the threshold $\bar{X}_m$. Long-run welfare then depends on the initial stock X_1: cohorts converge to $\bar{X}_h$ (and obtain $\bar{U}^+$) if X_1 lies above the threshold for the basin of attraction, but converge to the knowledge-collapse steady state (and obtain 0) if X_1 lies below this threshold. Thus, we need to consider AI's impact on both the high steady state utility $\bar{U}^+$ and the threshold for the basin of attraction $\bar{X}_m$.

Nevertheless, it should be remembered that regardless of whether there is a unique or multiple steady states and whether or not there is knowledge collapse, human effort is always under-supplied because of the general knowledge externality. This under-provision of effort is at the root of the results presented in the rest of the section, including the characterization of optimal policy.

4.2 The Impact of General Knowledge

Proposition 9. *Whenever the high-knowledge steady state exists, $\bar{U}^+$ is strictly increasing in I.*

Loosely speaking, this proposition shows that more general knowledge always improves welfare.

More specifically, higher I enables better aggregation of general knowledge and raises welfare through two distinct channels:

- **Level effect.** A larger I increases the informativeness of the public signal generated by any given per-capita effort, shifting the transition map F upward and raising the steady-state stock of general knowledge $\bar{X}_h$. From equation (10), this higher stock then translates into higher $\bar{U}^+$.
- **Selection effect.** When $\varepsilon > 4$, a larger I not only raises $\bar{U}^+$ but also makes the high-knowledge steady state easier to sustain dynamically by lowering the threshold for the basin of attraction, $\bar{X}_m$. This expands the basin of attraction of the high-knowledge steady state, increasing realized welfare for a larger set of initial conditions.

In practice, a higher I may result from more effective information-sharing communities and from improvements in other information-aggregation technologies that broaden the reach of human-generated general knowledge. Our model thus suggests that this type of information sharing improves both human effort aimed at learning and makes knowledge collapse less likely.

4.3 The Effects of Agentic AI

The welfare effects of a stronger agentic AI (higher τ_{A}) are inherently ambiguous because they operate through two opposing channels in (10): higher τ_{A} directly increases idiosyncratic precision but indirectly depresses the long-run stock of general knowledge by crowding out effort. Differentiating (10) at the high-knowledge steady state and applying the envelope theorem yields

$$\frac{\partial \bar{U}^{+}}{\partial \tau_{\mathrm{A}}} = \underbrace{g(\bar{Y}_h) \cdot G(\bar{X}_h)\Delta_{\mathrm{X}}}_{\text{Direct Effect, } \geq 0} + \underbrace{\frac{\partial G(\bar{X}_h)}{\partial \tau_{\mathrm{A}}} \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}})}_{\text{Indirect Effect, } \leq 0}.$$

The *direct effect* is the mechanical gain from more accurate individualized recommendations. The *indirect effect* reflects dynamic crowd-out: higher τ_{A} lowers equilibrium effort, reducing human-generated general knowledge and thus $\bar{X}_h$.

Let us introduce the following short hands for these two effects:

$$\mathrm{DE} = g(\bar{Y}_h) \cdot G(\bar{X}_h)\Delta_{\mathrm{X}}, \quad \mathrm{IE} = -\frac{\partial G(\bar{X}_h)}{\partial \tau_{\mathrm{A}}} \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}).$$

Then $\frac{\partial \bar{U}^{+}}{\partial \tau_{\mathrm{A}}} > 0$ iff $\mathrm{DE} > \mathrm{IE}$.

Below, we establish our main result that the negative effect is likely to dominate when agentic AI is very accurate. In fact, the ratio IE/DE is always increasing in τ_{A} (under natural regularity constraints on the parameter space), and strictly above one when τ_{A} is sufficiently large. Two economic forces push the ratio IE/DE upward as τ_{A} grows. First, the direct benefit DE exhibits diminishing returns because $g(\cdot)$ declines as idiosyncratic precision rises. This force captures the fact that once individualized recommendations are already fairly accurate, further increases in τ_{A} will not much improve predictions. Second, the indirect loss IE becomes larger when the stock of general knowledge is lower. This is because when $\bar{X}_h$ is low, marginal improvements in general knowledge are particularly valuable (the slope $g(\bar{X}_h)$ is high), so reductions in $\bar{X}_h$ translate into large welfare losses. Together, these forces imply that modest levels of agentic AI can raise welfare, but sufficiently accurate agentic AI reduces welfare.

We first consider the unique steady-state regime $\varepsilon < 4$. For expositional convenience, we also add a simple regularity condition on the prior idiosyncratic precision.[8]

[8] Essentially, Assumption 2 rules out a non-single-peaked welfare profile in which the welfare derivative changes sign twice. It is straightforward to see why this can happen Assumption 2 does not hold. In particular, we can write the indirect effect as $\mathrm{IE} = \left(-\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}}\right) \cdot G'_W(I\bar{e}_h) \cdot I \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}})$, and can conclude that this indirect effect becomes very large when idiosyncratic precision $\bar{Y}_h = \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h + \tau_{\mathrm{A}}$ is close to zero. Intuitively, in this region, equilibrium effort $\bar{e}_h$ becomes very sensitive to τ_A as the substitution effect $|\frac{\partial \bar{e}_h}{\partial \tau_A}|$ becomes very large. To see this, note that $\bar{e}_h$ satisfies $\Delta_{\mathrm{X}} \cdot G_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h + \tau_{\mathrm{A}}) = \bar{e}_h^{1/\varepsilon}$, and when $\bar{Y}_h$ is small, the left-hand side becomes very sensitive to τ_A, because $g(\cdot)$ is very steep near zero; in particular, $g'(\tau) = -\frac{1}{2\sqrt{2\pi}} \frac{e^{-\tau/2}}{\sqrt{\tau}} \left(1 + \frac{1}{\tau}\right) = \Theta(\tau^{-\frac{3}{2}})$ as $\tau \to 0$, and this violates single-peakedness. A simple example violating single-peakedness in this fashion obtains when $\varepsilon = 0.01$, $\Delta_{\mathrm{X}} = 0.005$, $\Delta_{\mathrm{G}} = 0.995$, $\lambda_{\mathrm{I}} = 0.05$, $\lambda_{\mathrm{G}} = 1$, $\sigma^{-2} = 0$, $\Sigma^{-2} = 1$, $I = 1$.

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

The welfare comparative statics for the high-knowledge steady state are summarized in the next proposition:

Proposition 10. *Suppose $\varepsilon < 4$ and Assumption 2 holds. There exists a finite threshold $\tau_{\mathrm{A}}^{\star}$ satisfying $0 \leq \tau_{\mathrm{A}}^{\star} < +\infty$, such that:*

(i) $\bar{U}^{+}$ is strictly increasing in τ_{A} on $(0, \tau_{\mathrm{A}}^{\star})$.

(ii) $\bar{U}^{+}$ is strictly decreasing in τ_{A} on $(\tau_{\mathrm{A}}^{\star}, +\infty)$.

(iii) $\bar{U}^{+} \to 0$ as $\tau_{\mathrm{A}} \to +\infty$.

In the unique steady state regime, there is no selection effect: the economy always converges to $\bar{X}_h$ from any $X_1 > 0$. The welfare trade-off is therefore purely about how $\bar{U}^{+}$ changes with τ_{A}. The proposition therefore implies that $\tau_{\mathrm{A}}^{\star}$ is the welfare-maximizing agentic AI accuracy, holding other parameters fixed. The limit $\bar{U}^{+} \to 0$ as $\tau_{\mathrm{A}} \to \infty$ highlights the limiting case of complete collapse: if agentic AI becomes sufficiently strong, equilibrium effort vanishes and general knowledge collapses, with very adverse welfare effects.

For the multiple steady states regime, $\varepsilon > 4$, agentic AI has an additional *selection* effect: it can shrink (and eventually eliminate) the basin of attraction of the high-knowledge outcome. We first state a similar welfare comparative static for the high-knowledge steady state.

Proposition 11. *Suppose $\varepsilon > 4$ and Assumption 2 holds. There exists a finite threshold $\tau_{\mathrm{A}}^{\star}$ satisfying $0 \leq \tau_{\mathrm{A}}^{\star} < \tau_{\mathrm{A}}^{c}$, such that:*

(i) $\bar{U}^{+}$ is strictly increasing in τ_{A} on $(0, \tau_{\mathrm{A}}^{\star})$.

(ii) $\bar{U}^{+}$ is strictly decreasing in τ_{A} on $(\tau_{\mathrm{A}}^{\star}, \tau_{\mathrm{A}}^{c})$.

(iii) $\bar{U}^{+} = 0$ as $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^{c}$.

Relative to the unique steady state regime, the welfare consequences of agentic AI are more stark here because a sufficiently accurate AI can induce a discontinuous regime change: once τ_{A} crosses τ_{A}^{c}, the high-knowledge steady state disappears and long-term knowledge collapse becomes inevitable. Moreover, even for $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$, realized welfare need not track $\bar{U}^{+}$ because the economy may fail to coordinate on the high-knowledge basin of attraction, and larger τ_{A} can increase the likelihood of coordination failure by increasing the threshold for the basin of attraction, $\bar{X}_m$.

4.4 Asymptotics for $I \to +\infty$

It is also informative to look at the case where the economy's ability to aggregate information becomes very large, meaning $I \to \infty$.

Proposition 12. *Suppose Assumption 2 holds. Then as the aggregation capacity I goes to infinity (with all other parameters fixed), the following holds:*[9]

$$\lim_{I \to +\infty} \frac{\tau_{\mathrm{A}}^{c}}{\log I} = \frac{2}{\varepsilon},$$

and

$$\lim_{I \to +\infty} \frac{\tau_{\mathrm{A}}^{\star}}{\log I} \;=\; \frac{2}{\varepsilon + 1}.$$

Proposition 12 delivers a sharp "scaling law"-type result. Both the collapse threshold τ_{A}^{c} and the welfare-maximizing precision $\tau_{\mathrm{A}}^{\star}$ grow only *logarithmically* in I, reflecting strong diminishing returns to aggregation. The economic reason is that, as τ_{A} increases, the marginal value of human effort in producing idiosyncratic precision falls very quickly (because $g(\cdot)$ decays exponentially with precision), so equilibrium effort declines sharply. Consequently, sustaining a given public-knowledge level requires the aggregation scale to grow very fast, which translates into τ_{A} scaling with $\log I$.

Finally, this asymptotic result implies that even as aggregation becomes arbitrarily powerful, the welfare-maximizing accuracy remains a fixed fraction away from the collapse boundary:

$$\frac{\tau_{\mathrm{A}}^{\star}}{\tau_{\mathrm{A}}^{c}} \;\longrightarrow\; \frac{\varepsilon}{\varepsilon + 1}.$$

Another important implication of this result is worth noting: because the benefits of larger I only scale with $\log I$, any aggregation benefits of AI will tend to be small relative to AI's agentic capabilities that change τ_{A}^{c}. This theme is explored in greater detail in Section 5.

4.5 Information Design

Motivated by the possible negative effects of agentic AI and the potential for knowledge collapse, we now discuss whether there are simple information design principles or regulation policies that can improve welfare.

Formally, we model regulation as *garbling* of the agentic recommendation by additive independent Gaussian noise (which imposes that the public agency cannot produce the information, but can partially suppress available information). An information design policy can then be represented by the sequence $\kappa = \{\kappa_t\}_{t \geq 1}$, such that $\kappa_t \in [0, +\infty]$, where κ_t is the precision of the noise added in

[9] Here, $\tau_{\mathrm{A}}^{\star}$ denotes the (unique) welfare-maximizing agentic precision characterized in Propositions 10 and 11.

period t.[10] Under policy κ, the signal released to agent i in period t is

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

with $\{\eta_{i,t}\}$ being independent across agents and time. Because signals are Gaussian, garbling is equivalent to replacing the raw agentic precision τ_{A} by an *effective* precision

$$\tau_{\mathrm{A}}^{(t)} = \tau_{\mathrm{A}}(\kappa_t) := \left(\tau_{\mathrm{A}}^{-1} + \kappa_t^{-1}\right)^{-1} \in [0, \tau_{\mathrm{A}}].$$

Therefore, the regulator's choice of κ_t is without loss of generality relative to a time-varying cap $\tau_{\mathrm{A}}^{(t)} \leq \tau_{\mathrm{A}}$ on the informativeness of the agentic signal.

Given a policy κ, the induced dynamic equilibrium is characterized exactly as in Proposition 1, with τ_{A} replaced by $\tau_{\mathrm{A}}^{(t)}$ period by period. Let $U_t(\kappa)$ denote the expected utility of the period-t cohort in the equilibrium induced by κ. We evaluate policies by long-run average welfare:

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

Since finite transitions have vanishing weight under this criterion, it is natural to focus on policies that eventually settle to a constant long-run cap. This policy class captures a permanent regulatory regime, possibly preceded by a temporary "moratorium" phase.

Let us call an information policy κ *eventually constant* if there exist $M < \infty$ and $\bar{\kappa} \in [0, +\infty]$ such that $\kappa_t = \bar{\kappa}$ for all $t \geq M$; let $\mathcal{K}^{EC}$ denote the set of such policies.

We consider the multiple steady-state case $\varepsilon > 4$. The following proposition characterizes the optimal Gaussian garbling policy under this regime.

Proposition 13. *Suppose that Assumption 2 holds and $\varepsilon > 4$, so that (i) the complete-collapse threshold τ_{A}^{c} and the steady-state objects $\{\bar{X}_m(\tau_{\mathrm{A}}), \bar{X}_h(\tau_{\mathrm{A}})\}_{\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}}$ are as in Proposition 5, and (ii) the high-knowledge steady-state welfare $\bar{U}^{+}(\tau_{\mathrm{A}})$ is well-defined and is maximized at a unique $\tau_{\mathrm{A}}^{\star} \in [0, \tau_{\mathrm{A}}^{c})$ as in Proposition 11.*

Fix an initial public precision $X_1 > 0$ satisfying $X_1 > \bar{X}_m(0)$, so that under full suppression ($\kappa_t = 0,\ \forall t$) the equilibrium dynamics converge to the high-knowledge steady state $\bar{X}_h(0)$. Define $\tau^{\star} := \min\{\tau_{\mathrm{A}}, \tau_{\mathrm{A}}^{\star}\}$ and pick $\kappa^{\star}$ such that $\tau_{\mathrm{A}}(\kappa^{\star}) = \tau^{\star}$. Then:

(i) *There exists a finite $M^{\star}$ such that the two-phase policy*

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

[10] $\kappa_t = +\infty$ corresponds to no garbling (full revelation), while $\kappa_t = 0$ corresponds to full suppression.

satisfies

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

(ii) *This two-phase policy is optimal among all eventually constant policies. That is,* $\sup_{\kappa \in \mathcal{K}^{EC}} U(\kappa) = U(\kappa^{2ph}) = \bar{U}^{+}(\tau^{\star})$.

The economic logic of the two-phase policy is informative and can be understood by considering each phase separately:

- **Phase 1 (rebuild the stock).** Setting $\kappa_t = 0$ fully suppresses agentic recommendations, forcing agents to rely on their own learning. Since effort is decreasing in effective agentic precision, this phase maximizes equilibrium effort and accelerates the rebuilding of public general knowledge, pushing the state into the basin of attraction of the high-knowledge steady state under the eventual cap.
- **Phase 2 (cap on the optimal level)** Once the stock of general knowledge has recovered, the regulator relaxes the moratorium but permanently caps the effective precision at $\tau^{\star}$, the welfare-maximizing long-run level (subject to feasibility). If $\tau_{\mathrm{A}} \leq \tau_{\mathrm{A}}^{\star}$, then $\tau^{\star} = \tau_{\mathrm{A}}$ and Phase 2 corresponds to no garbling; if $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^{\star}$, Phase 2 implements a strict precision cap.

Intuitively, Phase 1 is dealing with the problem of *path-dependence*: it ensures that the economy escapes the basin of attraction of the knowledge-collapse steady state. Phase 2 then sets an optimal long-run cap to build a sufficient stock of general knowledge. If the initial stock of general knowledge is sufficiently high then Phase 1 becomes unnecessary.

5 Extensions

In this section we discuss several important extensions. Our main purpose is to show that the simplifying assumptions imposed in our baseline model do not change the general insights rooted in the dynamic effects of agentic AI on collective knowledge.

5.1 Community versus AI Aggregation

In the baseline model, agentic AI only affects context-specific information but does not change the economy's stock of general knowledge directly (or does not improve the aggregation capacity I). This subsection considers an extension in which the same AI technology simultaneously (i) provides agent-specific recommendations and (ii) improves the stock of general knowledge.

Agentic AI provides a recommendation signal to each agent, as in the baseline:

$$s_{i,t}^{A} \sim \mathcal{N}(\theta_{i,t},\ \tau_{\mathrm{A}}^{-1}).$$

We additionally assume that a higher-τ_{A} AI system also increases the effective aggregation capacity. Concretely, we replace the fixed island size I by an AI-dependent *effective* aggregation scale

$$I(\tau_{\mathrm{A}}) := I_0 + \exp(\eta\tau_{\mathrm{A}}), \tag{11}$$

where $I_0 > 0$ is the pre-AI (baseline) aggregation capacity and $\eta \geq 0$ governs how strongly improvements in agentic capability translate into better aggregation.

The public signal technology is unchanged except for this replacement. Under symmetric effort e_t in period t, aggregate effort becomes $E_t = I(\tau_{\mathrm{A}})\, e_t$, so the island-level public signal is

$$s_t^C \sim \mathcal{N}\left(\theta_t, \frac{1}{\lambda_{\mathrm{G}} I(\tau_{\mathrm{A}}) e_t}\right).$$

All remaining parts of the model stay unchanged.

Unlike the baseline model, τ_{A} now bundles together two opposing forces:

- **Substitution (agentic channel):** higher τ_{A} lowers effort $e(X, \tau_{\mathrm{A}})$, reducing the individual-level general-knowledge creation;
- **Complementarity (aggregation channel):** higher τ_{A} increases $I(\tau_{\mathrm{A}})$, making any given amount of human effort more socially informative.

Which force dominates depends on the values of parameters, and specifically on η, which measures the responsiveness of aggregation to agentic AI capability.

Proposition 14. *Suppose that the AI-driven aggregation does not grow too quickly in that* $\eta < \eta^{\mathrm{crit}} := \frac{\varepsilon}{2}$. *Then the high-knowledge steady state (when it exists) satisfies* $\bar{X}_h(\tau_{\mathrm{A}}) \downarrow 0$ *as* $\tau_{\mathrm{A}} \to \infty$. *As a result, the welfare-maximizing level of AI is finite. That is,* $\arg\max_{\tau_{\mathrm{A}} \geq 0} \bar{U}^+(\tau_{\mathrm{A}}) < +\infty$

Proposition 14 highlights that allowing AI to simultaneously improve the stock of general knowledge does not change our main conclusions as long as this effect is not too pronounced. As usual, agentic recommendations are substitutes to human effort, and thus as τ_A rises, there is less effort. The better aggregation provided by AI works in the opposite direction. Provided that $\eta < \eta^{crit} = \varepsilon/2$, this second effect is not too strong and this ensures that the stock of general knowledge still vanishes as $\tau_A \to \infty$.

5.2 Synthetic Data

The baseline model assumes that new *general* knowledge is generated *only* by human effort and then aggregated into the stock of general knowledge. *Synthetic data*, meaning content produced by AI models themselves rather than collected from humans, has been advanced as a way to relax this constraint (Wang et al., 2023; Gunasekar et al., 2023).

We nevertheless view our assumption that human learning is the main source of collective knowledge as a good approximation for many open domains, such as scientific inquiry, free-form essay writing and open-ended medical or financial advice.[11] In such settings, verification against ground truth is slow, costly or impossible, and as a result, recursively training models on their own outputs becomes self-referential rather than genuinely informative.[12] Synthetic data in such settings should therefore be viewed as a *supplement* to human-generated knowledge—valuable primarily for improving training efficiency and/or enhancing instruction-following and alignment—whereas fully model-generated corpora tend to worsen downstream performance (Maini et al., 2024; Kang et al., 2025).

To formalize the possibility of using synthetic data in our model, suppose that in each period t an additional exogenous signal about the common state θ_t is produced with precision $\tau_{\text{syn}} \geq 0$ and pooled into the same public signal observed by future cohorts. The parameter τ_{syn} indexes how informative synthetic data is *independently* of human effort: $\tau_{\text{syn}} = 0$ recovers the baseline, while $\tau_{\text{syn}} = +\infty$ corresponds to the limiting case in which synthetic data fully substitutes for human-generated knowledge. The law of motion of the precision of general knowledge then becomes

$$X_{t+1}^{-1} = (X_t + \lambda_{\text{G}} E_t + \tau_{\text{syn}})^{-1} + \Sigma^2, \qquad E_t = I\, e(X_t, \tau_A). \tag{12}$$

Equivalently, the state transition map becomes

$$F_{\text{syn}}(X) = \left[\Sigma^2 + (X + \lambda_G I e(X, \tau_A) + \tau_{\text{syn}})^{-1}\right]^{-1}. \tag{13}$$

The substitution mechanism of the baseline model survives intact: higher agentic precision τ_A continues to crowd out human effort $e(X, \tau_A)$ and thereby reduce the flow of human-generated general knowledge. What synthetic data adds is a *floor*. Whenever $\tau_{\text{syn}} > 0$, we have $F_{\text{syn}}(0) > 0$, so the degenerate "zero-knowledge" point is no longer a fixed point of the dynamics. In the parameter region $\varepsilon > 4$, where the baseline model exhibits a stable collapse steady state, synthetic data replaces the collapse trap with a *low-knowledge* steady state that remains strictly positive. Because the additive synthetic term breaks the convexity argument used to characterize the full set of steady states in the baseline, we focus here on extremal steady states.[13]

[11] In contrast, in rule-governed or mechanically verifiable environments—such as reinforcement learning applied to games, algorithmic search, or formal proof checking—synthetic data can be paired with a signal based on ground truth, allowing AI to generate genuine novelty. Examples include: *AlphaGo Zero* via self-play Silver et al., 2017; the fast matrix multiplication search of Fawzi et al., 2022; and recent theorem-proving systems built on proof assistants Xin et al., 2024, Lin et al., 2025, Dong and Ma, 2025. These domains correspond to the limiting case in which synthetic data could in principle fully replace human-generated knowledge; in terms of the notation introduced in this subsection, this corresponds to the case where $\tau_{\text{syn}} = +\infty$.

[12] Shumailov et al. (2024) formalize this phenomenon as *model collapse*: when each generation of a generative model is trained, in part, on data sampled from earlier generations of the same model, sampling and approximation errors compound across generations. After several rounds the model effectively forgets the distribution tails, so that the synthetic corpus becomes a degraded image of the original one. In other words, without human input, model-generated content degrades rather than enriching the information environment.

[13] See the proof of Propositions 3 and 5, and in particular Lemma A-2, for details.

Proposition 15. *Suppose $0 < \tau_{syn} < +\infty$. Let $(\bar{X}^\star_\ell, \bar{e}^\star_\ell)$ and $(\bar{X}^\star_h, \bar{e}^\star_h)$ denote, respectively, the least and greatest steady states of the dynamical system $X_{t+1} = F_{\text{syn}}(X_t)$.*

1. *The least steady state is strictly positive: $\bar{X}^\star_\ell > 0$ and $\bar{e}^\star_\ell > 0$.*
2. *All four quantities $\bar{X}^\star_h, \bar{e}^\star_h, \bar{X}^\star_\ell$, and $\bar{e}^\star_\ell$ are strictly increasing in aggregation capacity I, strictly decreasing in agentic precision τ_A, and strictly increasing in synthetic precision τ_{syn}.*

This proposition thus establishes that our main results generalize to this environment with synthetic data, provided that synthetic data cannot fully replace human-generated knowledge.

5.3 (Imperfect) Separability of Effort

In the baseline model, a unit of learning effort jointly produces (i) context-specific knowledge and (ii) a "thin" contribution to general knowledge. Here we allow agents to *partially decouple* these two outputs. We model this partial separation by maintaining the following basic structure

$$s^H_{i,t} \sim \mathcal{N}\left(\theta_{i,t},\ \frac{1}{\lambda_{\text{I}} e_{i,t}}\right), \qquad s^A_{i,t} \sim \mathcal{N}\left(\theta_{i,t},\ \frac{1}{\tau_{\text{A}}}\right), \qquad \text{cost} = \frac{\varepsilon}{\varepsilon+1} e_{i,t}^{\frac{\varepsilon+1}{\varepsilon}},$$

but now assume that agent i's effort contributes a signal about the common state θ_t with precision proportional to

$$\lambda_{\text{G}}\, e^{\beta}_{i,t}, \qquad \beta \in [0, +\infty).$$

As before, the overall effective effort for general knowledge in island m can be computed as $E^{(\beta)}_{m,t} := \int_{I_m} e^{\beta}_{i,t}\, di$, and thus

$$s^C_{m,t} \sim \mathcal{N}\left(\theta_t,\ \frac{1}{\lambda_{\text{G}} E^{(\beta)}_{m,t}}\right). \tag{14}$$

Therefore, this generalization allows us to parameterize with β how tightly individual learning effort is "bundled" with collective knowledge-building. In particular,

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

so that a smaller β means "greater separability," because the two dimension of knowledge do not perfectly co-move anymore. The limiting case $\beta = 0$ corresponds to perfect separability in the sense that changes in $e_{i,t}$ do not alter the public-learning input as $e^{\beta}_{i,t} = 1$ is constant. The following proposition provides our main result in this subsection:

Proposition 16. *Under this extension, if $\beta > 0$ (not perfectly separable), then all results in the baseline model continue to apply, except with the stability condition $\varepsilon < 4$ replaced by $\varepsilon < \frac{4}{\beta}$, and*

with the asymptotic results in Proposition 12 replaced by

$$\lim_{I\to\infty}\frac{\tau_{\mathrm{A}}^{c}}{\log I}=\frac{2}{\varepsilon\beta}\qquad\textit{and}\qquad\lim_{I\to\infty}\frac{\tau_{\mathrm{A}}^{\star}}{\log I}=\frac{2}{\varepsilon\beta+1}.$$

Here we discuss how β shifts stability of the system near knowledge collapse. As clarified in Lemma 2, the local stability of the knowledge-collapse steady state is determined by the strength of public learning near $X = 0$. In that region, the stock of general knowledge scales with e^{β}, so what matters is how quickly e^{β} vanishes as $e \downarrow 0$. If β is large, e^{β} declines very rapidly as $e \downarrow 0$, strengthening the self-reinforcing logic behind collapse. Conversely, if β is small, a small amount of human effort will be sufficient to produce significant improvements in general knowledge, making the knowledge-collapse steady state locally unstable.

6 Conclusion

In this paper, we introduced a simple framework to study the implications of new generative AI technologies that promise to provide context-specific information and recommendations to human decision-makers. Our framework is based on three core ideas:

1. Good human decisions combine general knowledge with context-specific information.
2. Human effort directed at improving cognition generates both types of information, with the primary private return coming from context-specific information.
3. Individual contributions to general knowledge create externalities on others who build on this general knowledge.

These three observations together imply that the main motive for individual effort is often the acquisition of context-specific information, while the general knowledge an individual generates is primarily an externality. Consequently, better general knowledge in society is a *complement* to human learning effort, while better context-specific recommendations are *substitutes*. Because generative AI promises to provide this kind of context-specific information, it can be such a substitute, and reduce human effort. But then as lower human effort reduces general knowledge-building, generative AI (especially agentic AI) can *dynamically* push a society towards lower effective information and even lead to a knowledge-collapse steady state.

Our analysis is purely theoretical and shows that such a framework is tractable and yields a number of new and intuitive comparative statics. It also clarifies the conditions under which a knowledge-collapse steady state emerges and what determines how large its basin of attraction is. The tractability of our model also enables us to consider a number of extensions, aimed at showing that our main qualitative insights are robust.

There are several interesting areas for future research, including on incorporating synthetic data and other new AI capabilities, and exploring whether they can break the strong substitution between context-specific AI recommendations and human effort. The tractability of our model makes a range of other theoretical applications and extensions possible, for example, looking at different types of efforts on the side of humans and the implications of different AI technologies that provide varying mixes of general and context-specific recommendations.

Our framework also provides guidelines on different types of effects that need to be measured empirically to evaluate the overall welfare impacts of new AI advances.

References

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*, 108(6): 1488–1542, 2018.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of economic perspectives*, 33(2):3–30, 2019.

Daron Acemoglu, Munther A Dahleh, Ilan Lobel, and Asuman Ozdaglar. Bayesian learning in social networks. *The Review of Economic Studies*, 78(4):1201–1236, 2011.

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14(4):218–256, 2022.

Daron Acemoglu, Asuman Ozdaglar, and James Siderius. A model of online misinformation. *Review of Economic Studies*, 91(6):3117–3150, 2024.

Nikhil Agarwal, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. Combining human expertise with artificial intelligence: Experimental evidence from radiology. Technical report, National Bureau of Economic Research, 2023.

Nikhil Agarwal, Alex Moehring, and Alexander Wolitzky. Designing human-ai collaboration: A sufficient-statistic approach. Technical report, National Bureau of Economic Research, 2025.

Ajay K Agrawal, Joshua S Gans, and Avi Goldfarb. The turing transformation: Artificial intelligence, intelligence augmentation, and skill premiums. Technical report, National Bureau of Economic Research, 2023.

Kenneth J Arrow. The economic implications of learning by doing. *The review of economic studies*, 29(3):155–173, 1962a.

Kenneth Joseph Arrow. Economic welfare and the allocation of resources for invention. In *Readings in industrial economics: Volume two: Private enterprise and state intervention*, pages 219–236. Springer, 1962b.

Abhijit V Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3): 797–817, 1992.

Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. Generative ai can harm learning. *The Wharton School Research Paper*, 2024.

Dirk Bergemann, Alessandro Bonatti, and Tan Gan. The economics of social data. *The RAND Journal of Economics*, 53(2):263–296, 2022.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

Tilman Börgers, Angel Hernando-Veciana, and Daniel Krähmer. When are signals complements or substitutes? *Journal of Economic Theory*, 148(1):165–195, 2013.

Benjamin Brooks, Alexander Frankel, and Emir Kamenica. Comparisons of signals. *American Economic Review*, 114(9):2981–3006, 2024.

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative ai at work. *The Quarterly Journal of Economics*, 140(2):889–942, 2025.

Herman Budiyono, M Pudjaningsih, B Prastio, and A Maulidina. Exploring the long-term impact of ai writing tools on independent writing skills: a case study of indonesian language education students. *International Journal of Information and Education Technology*, 15(5):1003–1013, 2025.

Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, and Kate Smaje. The economic potential of generative ai. 2023.

Zoë Cullen, Danielle Li, and Shengwu Li. Labor as capital: AI and the ownership of expertise. Preliminary & incomplete draft. First draft August 12, 2025; this draft September 2, 2025. Unpublished manuscript., 2025. URL `https://danielle.li/assets/docs/LaborAsCapital.pdf`.

Krishna Dasaratha and Kevin He. Aggregative efficiency of bayesian learning in networks. *arXiv preprint arXiv:1911.10116*, 2019.

Krishna Dasaratha and Kevin He. Learning from viral content. *arXiv preprint arXiv:2210.01267*, 2022.

R Maria del Rio-Chanona, Nadzeya Laurentsyeva, and Johannes Wachs. Large language models reduce public knowledge sharing on online q&a platforms. *PNAS nexus*, 3(9):pgae400, 2024.

Kefan Dong and Tengyu Ma. Stp: Self-play llm theorem provers with iterative conjecturing and proving. *arXiv preprint arXiv:2502.00212*, 2025.

Maryam Farboodi, Andrew Koh, and Bryant Xia. Data-driven automation, 2025.

Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930):47–53, 2022.

Michael Gerlich. Ai tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1):6, 2025.

Suriya Gunasekar, Yi Zhang, Jyoti Aneja, Caio César Teodoro Mendes, Allie Del Giorno, Sivakanth Gopi, Mojan Javaheripi, Piero Kauffmann, Gustavo de Rosa, Olli Saarikivi, et al. Textbooks are all you need. *arXiv preprint arXiv:2306.11644*, 2023.

Gillian K Hadfield and Andrew Koh. An economy of ai agents. *arXiv preprint arXiv:2509.01063*, 2025.

Wayne Holmes, Fengchun Miao, et al. *Guidance for generative AI in education and research*. Unesco Publishing, 2023.

Enrique Ide. Automation, ai, and the intergenerational transmission of knowledge. *arXiv preprint arXiv:2507.16078*, 2025.

Enrique Ide and Eduard Talamàs. Artificial intelligence in the knowledge economy. *Journal of Political Economy*, 133(12):3762–3800, 2025.

Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman. Co-writing with opinionated language models affects users' views. In *Proceedings of the 2023 CHI conference on human factors in computing systems*, pages 1–15, 2023.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with 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, and Carole-Jean Wu. Demystifying synthetic data in llm pre-training: A systematic study of scaling laws, benefits, and pitfalls. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 10750–10769, 2025.

Godfrey Keller and Sven Rady. Strategic experimentation with poisson bandits. *Theoretical Economics*, 5(2):275–311, 2010.

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. Your brain on chatgpt: Accumulation of cognitive debt when using an ai assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 2025.

Victor Lei, David Strömberg, and Yanhui Wu. The generative ai learning penalty: Evidence from chinese k12 education. Working paper, 2026.

Yong Lin, Shange Tang, Bohan Lyu, Jiayun Wu, Hongzhou Lin, Kaiyu Yang, Jia Li, Mengzhou Xia, Danqi Chen, Sanjeev Arora, et al. Goedel-prover: A frontier model for open-source automated theorem proving. *arXiv preprint arXiv:2502.07640*, 2025.

Liang Lyu, James Siderius, Hannah Li, Daron Acemoglu, Daniel Huttenlocher, and Asuman Ozdaglar. Wikipedia contributions in the wake of chatgpt. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1176–1179, 2025.

Pratyush Maini, Skyler Seto, Richard Bai, David Grangier, Yizhe Zhang, and Navdeep Jaitly. Rephrasing the web: A recipe for compute and data-efficient language modeling. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 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. Artificial intelligence index report 2025. *arXiv preprint arXiv:2504.07139*, 2025.

Richard R Nelson. The simple economics of basic scientific research. *Journal of political economy*, 67(3):297–306, 1959.

OpenAI. Gpt-4, March 2023. URL `https://openai.com/index/gpt-4-research/`. Accessed January 2026.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal. Ai models collapse when trained on recursively generated data. *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. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. Self-instruct: Aligning language models with self-generated instructions. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 13484–13508, 2023.

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. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Huajian Xin, Daya Guo, Zhihong Shao, Zhizhou Ren, Qihao Zhu, Bo Liu, Chong Ruan, Wenda Li, and Xiaodan Liang. Deepseek-prover: Advancing theorem proving in llms through large-scale synthetic data. *arXiv preprint arXiv:2405.14333*, 2024.

Appendix A Main Appendix

A.1 Proof of Lemma 1

Let us fix $\tau_{\mathrm{A}} \geq 0$ and denote $e(X) = e(X, \tau_{\mathrm{A}})$, where $e(X, \tau_{\mathrm{A}})$ is the (unique) best-response effort solving

$$e(X, \tau_{\mathrm{A}}) := \left\{ e \geq 0 \mid \Delta_{\mathrm{X}} G(X) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}}) = e^{1/\varepsilon} \right\}.$$

Define

$$\Psi(e, X) := \Delta_{\mathrm{X}} G(X) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}}) - e^{1/\varepsilon}$$

For any $e, X > 0$, we have $\partial_e \Psi(e, X) < 0$ and $\partial_X \Psi(e, X) > 0$, so by the implicit function theorem $e(X)$ is continuous and strictly increasing on $(0, \infty)$. As $X \downarrow 0$, $G(X) \to 0$, which forces $e(X) \to 0$; thus $e(X)$ is continuous on $[0, \infty)$.

As a result, $F(X) = \left[\Sigma^2 + \left(X + \lambda_{\mathrm{G}} I e(X)\right)^{-1}\right]^{-1}$ is also continuous and strictly increasing in $[0, \infty)$, with $F(0) = 0$ and $\lim_{X \to +\infty} F(X) = \Sigma^{-2}$.

A.2 Proof of Proposition 1

Consider a fixed period t and take (X_t, τ_{A}) as given. We have shown that an agent i's expected utility is strictly concave in effort $e_{i,t}$, so there is a unique best response $e(X_t, \tau_{\mathrm{A}})$; optimal predictions are posterior means. In a symmetric equilibrium $e_{i,t} = e_t = e(X_t, \tau_{\mathrm{A}})$ for all i, so aggregate effort on an island is $E_t = \int_{\mathcal{I}} e_{i,t} di = I e_t$. Given E_t, the public signal has precision $\lambda_{\mathrm{G}} E_t$ and Kalman updating implies $X_{t+1}^{-1} = (X_t + \lambda_{\mathrm{G}} E_t)^{-1} + \Sigma^2$, equivalently, $X_{t+1} = \left[\Sigma^2 + (X_t + \lambda_{\mathrm{G}} I e_t)^{-1}\right]^{-1}$. Since $e(X_t, \tau_{\mathrm{A}})$ is single-valued, the recursion pins down a unique symmetric equilibrium path.

A.3 Proof of Proposition 2

Fix $X > 0$ and define

$$u := X + \lambda_{\mathrm{G}} I e(X, \tau_{\mathrm{A}}),$$

so that F can be written as $F = \left(\Sigma^2 + u^{-1}\right)^{-1}$. Clearly, F is (strictly) increasing in u on $\mathbb{R}_{\geq 0}$.

Since $X > 0$, we have $e(X, \tau_{\mathrm{A}}) > 0$, so that u is strictly increasing in I. By Observation 2, $e(X, \tau_{\mathrm{A}})$ is strictly decreasing in τ_{A}, so u is also strictly decreasing in τ_{A}. Putting together, we conclude that F is strictly increasing in I and strictly decreasing in τ_{A}.

A.4 Proof of Proposition 3 and Proposition 5

In the main text, we define steady-state social precision $\bar{X}$ as a fixed point of the one-step map F, and the corresponding steady-state effort as the best response $\bar{e} = e(\bar{X}, \tau_{\mathrm{A}})$. Here, we first provide

a convenient "effort-first" characterization: we derive an equation that pins down $\bar{e}$ directly.

Rearranging terms in the fixed point condition $F(\bar{X}) = \bar{X}$, we obtain the linking function between aggregate effort and steady-state public precision: $(\bar{X}^{-1} - \Sigma^2)^{-1} - \bar{X} = \lambda_{\mathrm{G}} \cdot (I\bar{e})$.

Solving for $\bar{X}$ yields $\bar{X} = W(I\bar{e})$, where W is given by the (positive) root of the resulting quadratic:

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

Substituting $\bar{X} = W(I\bar{e})$ into the first-order condition for equilibrium effort, we obtain the equation that steady-state effort needs to satisfy: $D(\bar{e}) = 0$, where[14]

$$D(e) := \underbrace{\Delta_{\mathrm{X}} \cdot G_W(Ie) \ \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}})}_{=:\mathrm{MB}(e)} - \underbrace{e^{1/\varepsilon}}_{=:\mathrm{MC}(e)} . \tag{A-1}$$

Here, $G_W(\cdot)$ is the implied link between aggregate effort and the steady-state "hitting" probability for the public task:

$$G_W(E) := G\big(W(E)\big). \tag{A-2}$$

We will study steady states via the sign and roots of $D(e)$ and then map back to steady-state public precision using $\bar{X} = W(I\bar{e})$.

The following sign-linking lemma is useful for studying the global dynamics.

Lemma A-1. *Let W be the effort-precision link, and let $D(e)$ be given by equation* (A-1)*. For every $e \geq 0$, define $X(e) := W(Ie)$. Then*

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

where $\operatorname{sign}(\cdot)$ *denotes the usual sign function.*

Proof. For any public precision $X \geq 0$, define a helper function D_X:

$$D_X(e) := \Delta_{\mathrm{X}} G(X) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}}) - e^{1/\varepsilon}, \qquad e \geq 0.$$

This is exactly the first-order condition for the agent's optimal effort at public precision X. By definition, $e(X)$ is the unique root of D_X: $D_X(e(X)) = 0$. Since $D_X(e)$ is strictly decreasing in e, crosses zero exactly once at $e = e(X)$, therefore

$$\operatorname{sign}(D_X(e)) = \operatorname{sign}(e(X) - e) \quad \text{for all } e \geq 0. \tag{A-3}$$

[14] When we concern comparative statics with respect to τ_{A} and I, we sometimes write $D(e)$ as $D(e; \tau_{\mathrm{A}}, I)$ to highlights its dependency with τ_{A}, I.

In particular, applying (A-3) with $X = X(e)$ yields

$$\operatorname{sign}(D(e)) = \operatorname{sign}(e(X(e)) - e) \quad \text{for all } e \geq 0. \tag{A-4}$$

Define, for arbitrary effort e,

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

The actual dynamic map is $F(X) = T(X, e(X))$. Thus we have

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

Since $T(X(e), \cdot)$ is strictly increasing in effort, the sign of the right-hand side of (A-5) coincides with the sign of $e(X(e)) - e$, that is,

$$\operatorname{sign}(F(X(e)) - X(e)) = \operatorname{sign}(e(X(e)) - e). \tag{A-6}$$

Comparing (A-4) and (A-6), we obtain for all $e > 0$,

$$\operatorname{sign} D(e) = \operatorname{sign}(e(X(e)) - e) = \operatorname{sign}(F(X(e)) - X(e)) = \operatorname{sign}(F(W(Ie)) - W(Ie)).$$

∎

Denote $x := \log e$ so that we can rewrite (A-1) in the log-space:

$$D(x) := \underbrace{\Delta_{\mathrm{X}} \cdot G_W(I \exp(x))\ \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} \exp(x) + \tau_{\mathrm{A}})}_{\mathrm{MB}(x)} - \underbrace{\exp(x/\varepsilon)}_{\mathrm{MC}(x)}.$$

Each real root of $D(x)$ corresponds to a positive steady-state effort. The following lemma is the key to the characterization of the roots of $D(x)$.

Lemma A-2. $\log(\mathrm{MB}(x))$ *is a strictly concave function of* x.

Proof. We have the decomposition:

$$\log(\mathrm{MB}(x)) = \log(\lambda_{\mathrm{I}} \Delta_{\mathrm{X}}) + \underbrace{\log G_W(Ie^x)}_{=:t_1(x)} + \underbrace{\log g(\sigma^{-2} + \lambda_{\mathrm{I}} e^x + \tau_{\mathrm{A}})}_{=:t_2(x)}$$

We show that both $t_1(x)$ and $t_2(x)$ are strictly concave. For $t_1(x)$, we have

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

which is strictly decreasing in x due to Lemma A-4.

For $t_2(x)$, we have

$$\begin{aligned} t_2'(x) &= \frac{g'(\sigma^{-2}+\lambda_{\mathrm{I}}e^x+\tau_{\mathrm{A}})}{g(\sigma^{-2}+\lambda_{\mathrm{I}}e^x+\tau_{\mathrm{A}})}\cdot\lambda_{\mathrm{I}}e^x \\ &= -\frac{\lambda_{\mathrm{I}}e^x}{2}-\frac{\lambda_{\mathrm{I}}e^x}{2(\sigma^{-2}+\lambda_{\mathrm{I}}e^x+\tau_{\mathrm{A}})}, \end{aligned}$$

which is also strictly decreasing in x. ∎

Proof of Proposition 3. Recall that

$$D(e) := \underbrace{\Delta_{\mathrm{X}}\cdot G(W(Ie))\ \cdot\lambda_{\mathrm{I}}g(\sigma^{-2}+\lambda_{\mathrm{I}}e+\tau_{\mathrm{A}})}_{\mathrm{MB}(e)} - \underbrace{e^{1/\varepsilon}}_{\mathrm{MC}(e)}.$$

By Lemma A-3, as $E\to 0^+$ we have $G(W(E))=\kappa_1 E^{1/4}+o(E^{1/4})$ for some $\kappa_1>0$. Setting $E=Ie$ and using the continuity of $g(\cdot)$, we obtain, as $e\to 0^+$,

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

for some $C>0$. Thus, $D(e)$ is strictly positive near 0. It is also easy to verify that $D(e)$ is strictly negative for large e.

Next, Lemma A-2 shows that $\log(\mathrm{MB}(x))$ is strictly concave in $x=\log e$, while $\log(\mathrm{MC}(x))=x/\varepsilon$ is linear. Moreover, the previous limiting behavior of $D(e)$ implies that the MB curve is larger at $x=-\infty$, while the MC curve larger at $x=+\infty$. Hence these two curves must cross *exactly once*. Therefore, $D(e)=0$ has at exactly one positive root; denoted as $\bar{e}_h>0$. From the concavity structure we obtain that $D(e)>0$ for $0<e<\bar{e}_h$ and $D(e)<0$ for $e>\bar{e}_h$.

Let $\bar{X}_h := W(I\bar{e}_h)>0$. Together with the trivial steady state $\bar{X}_\ell=0$, we obtain exactly two steady-state public precisions $0=\bar{X}_l<\bar{X}_h<\infty$. By Lemma A-1, for every $e\geq 0$,

$$\operatorname{sign} D(e) = \operatorname{sign}(F(X(e))-X(e)), \qquad X(e):=W(Ie).$$

Thus

$$F(X)-X\begin{cases} >0, & 0<X<\bar{X}_h, \\ <0, & X>\bar{X}_h. \end{cases}$$

Consider the dynamical system $X_{t+1}=F(X_t)$ on $[0,\infty)$. Lemma 1 gives that F is continuous, strictly increasing, $F(0)=0$ and $F(+\infty)=\Sigma^{-2}$. If $X_1\in(0,\bar{X}_h)$, then $\{X_t\}$ is increasing and bounded above by $\bar{X}_h$, hence converging to $\bar{X}_h$. Same argument applies to $X_1\in(\bar{X}_h,+\infty)$. Thus every trajectory with $X_1>0$ converges to $\bar{X}_h$, while $X_1=0$ stays at 0. ∎

Proof of Proposition 5. For each $\tau\geq 0$, let $F\tau(\cdot)$ denote the state transition map $F(\cdot)$ when $\tau_{\mathrm{A}}=\tau$,

and let

$$D_\tau(e) := \underbrace{\Delta_{\mathrm{X}} \cdot G_W(Ie) \ \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau)}_{=:\mathrm{MB}\tau(e)} - \underbrace{e^{1/\varepsilon}}_{=:\mathrm{MC}(e)}. \tag{A-8}$$

be the steady-state effort equation (A-1) when $\tau_{\mathrm{A}} = \tau$.

By Lemma A-3, as $E \to 0^+$ we have $G_W(E) = \kappa_1 E^{1/4} + o(E^{1/4})$ for some $\kappa_1 > 0$. Setting $E = Ie$ and using the continuity of $g(\cdot)$, we obtain, as $e \to 0^+$,

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

for some $C(\tau) > 0$ with $C'(\tau) < 0$. Since $1/\varepsilon < \frac{1}{4}$, we have $D\tau(e)/e^{1/4} \to -\infty$ as $e \to 0^+$, uniformly in τ on $\tau \in [0, \infty)$. Hence there exists $\delta > 0$ (independent of τ) such that

$$D\tau(e) < 0 \quad \text{for all } 0 < e \leq \delta. \tag{A-9}$$

For large effort, note that $0 \leq G_W(Ie) \leq G(\Sigma^{-2})$ and $g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau) \to 0$ as $e \to \infty$. Thus the marginal benefit term in $D\tau(e)$ is bounded, whereas $e^{1/\varepsilon} \to \infty$. As a result, there exists $R > 0$ (again uniform on τ) such that

$$D\tau(e) < 0 \quad \text{for all } e \geq R. \tag{A-10}$$

Therefore, for each fixed τ, $D\tau(e)$ is continuous, strictly negative near 0 and for large e, and any positive root must lie in (δ, R).

Next, Lemma A-2 shows that $\log(\mathrm{MB}\tau(x))$ is strictly concave in $x = \log e$, while $\log(\mathrm{MC}(x)) = x/\varepsilon$ is linear. Hence these two curves cross at most twice, so $D\tau(e)$ has at most two positive roots. Combining this with (A-9)–(A-10), we conclude:

(i) for each τ, $D\tau$ has 0, 1, or 2 positive roots;

(ii) if $0 < e_m(\tau) < e_h(\tau)$ are the two roots, then

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

Since $g(\cdot)$ is strictly decreasing, $D\tau(e)$ is strictly decreasing in τ for every $e > 0$. Let

$$\mathcal{S} := \{\tau \geq 0 : \exists\, e > 0 \text{ s.t. } D\tau(e) = 0\}$$

be the set of τ for which there exists at least one positive steady–state effort. If $\mathcal{S} = \emptyset$, then there is never a positive steady state and Proposition 5(ii) holds with $\tau_{\mathrm{A}}^c = 0$; in this case part (i) is vacuous. From now on, we assume $\mathcal{S} \neq \emptyset$.

Because $D\tau(e)$ is (pointwise) strictly decreasing in τ for $e > 0$, the set $\mathcal{S}$ is downward closed: if $\tau_2 \in \mathcal{S}$ and $\tau_1 < \tau_2$, then $\tau_1 \in \mathcal{S}$. Since for sufficiently large τ we have $D\tau(e) < 0$ for all $e > \delta$ (by

the uniform bound $g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau) \le g(\sigma^{-2} + \tau) \to 0$), the set $\mathcal{S}$ is bounded above.

Then the complete-collapse threshold corresponds to $\tau_{\mathrm{A}}^c := \sup \mathcal{S} < \infty$.

We now claim:

(i) For every $\tau < \tau_{\mathrm{A}}^c$, $D\tau$ has exactly two positive roots $0 < e_m(\tau) < e_h(\tau)$.

(ii) For every $\tau > \tau_{\mathrm{A}}^c$, $D\tau(e) < 0$ for all $e > 0$ (i.e, no positive root).

(iii) if $\tau_{\mathrm{A}}^c > 0$, then at $\tau = \tau_{\mathrm{A}}^c$ there is exactly one positive root $\bar{e} > 0$, and it is a tangency ($D\tau(\bar{e}) = D\tau'(\bar{e}) = 0$).

Point (ii) is immediate from the definition of τ_{A}^c as a supremum. For (i), fix $\tau < \tau_{\mathrm{A}}^c$ and choose $\tilde{\tau} \in (\tau, \tau_{\mathrm{A}}^c)$ so that $\tilde{\tau} \in \mathcal{S}$. Then $D_{\tilde{\tau}}$ has at least one positive root $e^* > 0$, so by monotonicity $D_\tau(e^*) > 0$. Together with (A-9)–(A-10), continuity implies that D_τ must cross zero twice on $(0, \infty)$, hence has exactly two positive roots $e_m(\tau)$ and $e_h(\tau)$ by the at–most–two–roots property above.

For (iii), take a sequence $\tau_n \uparrow \tau_{\mathrm{A}}^c$ with $\tau_n \in \mathcal{S}$, and let $e_m(\tau_n), e_h(\tau_n)$ be the two roots of D_{τ_n}. We have uniform bounds $\delta \le e_m(\tau_n) \le e_h(\tau_n) \le R$, so after passing to a subsequence we may assume $e_m(\tau_n) \to \bar{e}_1$ and $e_h(\tau_n) \to \bar{e}_2$ with $\delta < \bar{e}_1 \le \bar{e}_2 \le R$. By continuity, $\bar{e}_1$ and $\bar{e}_2$ are both roots of $D\tau$ when $\tau = \tau_{\mathrm{A}}^c$. If $\bar{e}_1 < \bar{e}_2$, the Implicit Function Theorem implies that for all τ close to τ_{A}^c there are two distinct nearby roots, contradicting the definition of τ_{A}^c as a supremum of $\mathcal{S}$. Hence $\bar{e}_1 = \bar{e}_2 =: \bar{e} > 0$, and D_τ has a *double* root at $\bar{e}$ when $\tau = \tau_{\mathrm{A}}^c$.

Positive steady–state efforts correspond one-to-one to positive steady–state social precisions via $X = W(Ie)$. Thus for each $\tau < \tau_{\mathrm{A}}^c$ there are exactly three steady states in X–space:

$$\bar{X}_l(\tau) = 0, \quad \bar{X}_m(\tau) = W(Ie_m(\tau)), \quad \bar{X}_h(\tau) = W(Ie_h(\tau)),$$

with $0 < \bar{X}_m(\tau) < \bar{X}_h(\tau) < \infty$. For $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^c$ we have only one steady state $\bar{X}_l(\tau) = 0$. This yields the existence and number of steady states stated in parts (i) and (ii).

Fix $\tau < \tau_{\mathrm{A}}^c$ and consider the one–dimensional dynamical system $X_{t+1} = F\tau(X_t)$ on $[0, \infty)$. Lemma 1 implies $F\tau$ is continuous and strictly increasing with $F\tau(0) = 0$ and $F\tau(+\infty) = \Sigma^{-2} < \infty$. Moreover, F has exactly the three fixed points $0 < \bar{X}_m(\tau) < \bar{X}_h(\tau)$, and it follows by Lemma A-1 that

$$F\tau(X) - X \begin{cases} < 0, & X \in (0, \bar{X}_m(\tau)), \\ > 0, & X \in (\bar{X}_m(\tau), \bar{X}_h(\tau)), \\ < 0, & X \in (\bar{X}_h(\tau), \infty). \end{cases}$$

Now let $X_1 \ge 0$ be arbitrary. If $X_1 \in (0, \bar{X}_m(\tau))$, then $\{X_t\}$ will remain in $(0, \bar{X}_m(\tau))$ and will be strictly decreasing, hence converges to a fixed point in $[0, \bar{X}_m(\tau)]$, necessarily $\bar{X}_l(\tau) = 0$. If

$X_1 \in (\bar{X}_m(\tau), \bar{X}_h(\tau))$, then the sequence increases and remains in $(\bar{X}_m(\tau), \bar{X}_h(\tau))$; thus $X_t \to \bar{X}_h(\tau)$. Similar arguments apply to the case where $X_1 > \bar{X}_h(\tau)$.

This proves the basin partition in part (i):

$$X_1 < \bar{X}_m(\tau) \;\Rightarrow\; X_t \to 0, \qquad X_1 > \bar{X}_m(\tau) \;\Rightarrow\; X_t \to \bar{X}_h(\tau).$$

When $\tau_{\mathrm{A}} > \tau_{\mathrm{A}}^c$, we have shown that 0 is the unique steady state. Moreover, for all $X > 0$, $F\tau(X) < X$ due to Lemma A-1. Hence for any $X_1 > 0$, the sequence $X_{t+1} = F\tau(X_t)$ is strictly decreasing and bounded below by 0, thus must converge to 0. This yields the basin partition in part (ii). ∎

A.5 Proof of Lemma 2

By Lemma 1, the state transition map F is continuous, strictly increasing on $\mathbb{R}_{\geq 0}$, and satisfies $F(0) = 0$. Recall that from (A-7), we have for some $C > 0$, $D(e) = Ce^{1/4} - e^{1/\varepsilon} + o(e^{1/4})$.

Therefore, when $\varepsilon < 4$, there exists $\delta_e > 0$, such that $D(e) > 0$ for $e \in (0, \delta_e)$. By Lemma A-1, there exists $\delta_X > 0$, such that $F(X) > X$ for $X \in (0, \delta_X)$, i.e., $F(X) > X$ in a neighborhood of 0. Combining with the above property of F, we conclude that the zero steady state is locally unstable.

Similarly, when $\varepsilon > 4$, there exists $\delta_e' > 0$, such that $D(e) < 0$ for $e \in (0, \delta_e')$. By Lemma A-1, there exists $\delta_X' > 0$, such that $F(X) < X$ for $X \in (0, \delta_X')$. In this case, the zero steady state is locally stable.

A.6 Additional Technical Lemmas

Lemma A-3. *For the function $G_W(\cdot)$ defined in* (A-2)*, the following statements hold:*

- *$G_W(\cdot)$ is strictly increasing, with $\lim_{E\to 0} G_W(E) = 0$, and $\lim_{E\to+\infty} G_W(E) = G(\Sigma^{-2})$.*
- *Under the regime $E \to 0^+$, we have*
  - $G_W(E) = \kappa_1 E^{\frac{1}{4}} + o(E^{\frac{1}{4}})$
  - $G_W'(E) = \frac{1}{4}\kappa_1 E^{-\frac{3}{4}} + o(E^{-\frac{3}{4}})$

  where $\kappa_1 = 2^{\frac{1}{2}}\lambda_{\mathrm{G}}^{\frac{1}{2}} a^{\frac{1}{4}}\phi(0) > 0$.
- *Under the regime $E \to +\infty$,*
  - $G_W(E) = G(\Sigma^{-2}) - \frac{\kappa_2}{E} + o(\frac{1}{E})$
  - $G_W'(E) = \frac{\kappa_2}{E^2} + o(\frac{1}{E^2})$

  where $\kappa_2 = \phi(\Sigma^{-1})\lambda_{\mathrm{G}}^{-1}\Sigma^{-3} > 0$.

Proof. Since both $G(\cdot)$ and $W(\cdot)$ are strictly increasing, $G_W(\cdot)$ is also strictly increasing.

Denote

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

Then we have that $\frac{2}{\lambda_{\mathrm{G}}}W(E) = E(S(E)-1)$. Note that

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

Thus

$$\frac{2}{\lambda_{\mathrm{G}}}W'(E) = S(E) - 1 - \frac{a}{2ES(E)} = \frac{(S(E)-1)^2}{2S(E)}. \tag{A-11}$$

Regime $E \to 0^+$: We have the following asymptotics:

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

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

and

$$W'(E) = \frac{\lambda_{\mathrm{G}}}{2} \cdot \frac{(S(E)-1)^2}{2S(E)} \sim \frac{\lambda_{\mathrm{G}} a^{\frac{1}{2}}}{4} \cdot E^{-\frac{1}{2}}.$$

Therefore,

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

Regime $E \to +\infty$: In this case, we have asymptotics:

$$S(E) - 1 \sim \frac{a}{2E},$$

$$\Sigma^{-2} - W(E) = \Sigma^{-2} \cdot \left(1 - \frac{2}{S(E)+1}\right) = \Sigma^{-2} \cdot \left(\frac{S(E)-1}{S(E)+1}\right) \sim \Sigma^{-2} \cdot \frac{a}{4E} = \lambda_{\mathrm{G}}^{-1}\Sigma^{-4} \cdot E^{-1},$$

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

and

$$\begin{aligned} W'(E) &= \frac{\lambda_{\mathrm{G}}}{2}\frac{(S(E)-1)^2}{2S(E)} \\ &= \frac{\lambda_{\mathrm{G}}}{4}\frac{(S(E)^2-1)^2}{S(E)(S(E)+1)^2} \\ &\sim \frac{\lambda_{\mathrm{G}}}{16}\left(\frac{a}{E}\right)^2 \\ &= \lambda_{\mathrm{G}}^{-1}\Sigma^{-4}E^{-2}. \end{aligned}$$

As a result,

$$\begin{aligned} G'_W(E) &= \frac{\phi(\sqrt{W(E)})}{\sqrt{W(E)}}W'(E) \\ &\sim \frac{\phi(\sqrt{\Sigma^{-2}})}{\sqrt{\Sigma^{-2}}}\lambda_{\mathrm{G}}^{-1}\Sigma^{-4}E^{-2} \\ &= \phi(\Sigma^{-1})\lambda_{\mathrm{G}}^{-1}\Sigma^{-3}E^{-2}. \end{aligned}$$

∎

Lemma A-4. *Let*

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

We have that $R(E)$ is strictly decreasing in $(0,+\infty)$, with $\lim_{E\to 0^+} R(E) = \frac{1}{4}$.

Proof. Recall that $G(\tau) = 2\Phi(\sqrt{\tau}) - 1$, and $W(E) = \frac{\lambda_{\mathrm{G}}E}{2}\left(\sqrt{1+\frac{4}{\lambda_{\mathrm{G}}\Sigma^2\cdot E}}-1\right)$.

We have the following factorization of $R(E)$:

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

so that $R(E) = L(E)\cdot Q(W(E))$. To prove that $R(E)$ is strictly decreasing, it suffices to prove both $L(\cdot)$ and $Q(\cdot)$ are strictly decreasing.

Following the same notation as in the proof of Lemma A-3, we have

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

where $a := \frac{4}{\lambda_{\mathrm{G}}\Sigma^2}$ and $S(E) := \sqrt{1+\frac{a}{E}}$. Since $S(E)$ is strictly decreasing in E, so is $L(E)$.

We then prove that $Q(\tau) = \frac{\tau G'(\tau)}{G(\tau)}$ is strictly decreasing. Note that

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

where $z = \sqrt{\tau}$. We only need to prove that P is strictly decreasing, which follows because

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

Moreover, for $z > 0$, we have

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

which implies that for $z \in (0, +\infty)$,

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

Thus we have that for $z \in (0, +\infty)$, $P'(z) < 0$. ∎

A.7 Proof of Proposition 4

We divide the proof in several parts:

1. $\bar{e}_h$ in I and τ_{A}.

By Proposition 3, $D(e; \tau_{\mathrm{A}}, I)$ has a unique positive root $\bar{e}_h > 0$ satisfying $\partial_e D(\bar{e}_h; \tau_{\mathrm{A}}, I) < 0$. Moreover, for every $e > 0$, we have $\partial\tau D(e; \tau_{\mathrm{A}}, I) < 0$ and $\partial_I D(e; \tau_{\mathrm{A}}, I) > 0$. Thus the implicit function theorem implies that $\bar{e}_h$ is C^1 in (τ_{A}, I), and

$$\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} = -\frac{\partial \tau D(\bar{e}_h; \tau_{\mathrm{A}}, I)}{\partial_e D(\bar{e}_h; \tau_{\mathrm{A}}, I)}, \qquad \frac{\partial \bar{e}_h}{\partial I} = -\frac{\partial_I D(\bar{e}_h; \tau_{\mathrm{A}}, I)}{\partial_e D(\bar{e}_h; \tau_{\mathrm{A}}, I)}.$$

Therefore, we have $\frac{\partial \bar{e}_h}{\partial I} > 0$ and $\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} < 0$.

2. $\bar{Y}_h$ in I and τ_{A}.

For $\bar{Y}_h = \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h + \tau_{\mathrm{A}}$, we have $\frac{\partial \bar{Y}_h}{\partial I} = \lambda_{\mathrm{I}} \cdot \frac{\partial \bar{e}_h}{\partial I} > 0$.

The comparative static of $\bar{Y}_h$ with respect to τ_{A} requires more refined analysis. We first show a lemma.

Lemma A-5. *The following holds:*

$$-\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} = \frac{(1+1/\bar{Y}_h)\,\bar{e}_h}{\lambda_{\mathrm{I}}(1+1/\bar{Y}_h)\,\bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))}. \tag{A-12}$$

where $R(E) := \frac{EG''_W(E)}{G_W(E)}$.

Proof. Direct calculations show that

$$\frac{\partial D(\bar{e}_h;\tau_{\mathrm{A}},I)}{\partial e} = A - B - C, \quad \frac{\partial D(\bar{e}_h;\tau_{\mathrm{A}},I)}{\partial \tau_{\mathrm{A}}} = -\frac{1}{\lambda_{\mathrm{I}}}B$$

where $A := \Delta_{\mathrm{X}} \cdot IG'_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}} g\left(\sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h + \tau_{\mathrm{A}}\right)$; $B := -\Delta_{\mathrm{X}} G_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}}^2 g'\left(\sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h + \tau_{\mathrm{A}}\right)$; and $C := \frac{1}{\varepsilon}(\bar{e}_h)^{1/\varepsilon - 1}$.

Therefore, we have

$$\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} = -\frac{\partial \tau D(\bar{e}_h;\tau_{\mathrm{A}},I)}{\partial_e D(\bar{e}_h;\tau_{\mathrm{A}},I)} = -\frac{B}{\lambda_{\mathrm{I}}(B+C-A)}$$

Moreover, the quantities can be simplified as follows:

$$\begin{aligned} A &= \Delta_{\mathrm{X}} \cdot IG'_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}} g\left(\bar{Y}_h\right) \\ &= \Delta_{\mathrm{X}} \cdot G_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}} g\left(\bar{Y}_h\right) \cdot R(I\bar{e}_h) \cdot (\bar{e}_h)^{-1} \\ &= (\bar{e}_h)^{1/\varepsilon} \cdot R(I\bar{e}_h) \cdot (\bar{e}_h)^{-1} \\ &= R(I\bar{e}_h) \cdot (\bar{e}_h)^{1/\varepsilon - 1}. \end{aligned} \qquad \begin{aligned} B &= -\Delta_{\mathrm{X}} G_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}}^2 g'\left(\bar{Y}_h\right) \\ &= \Delta_{\mathrm{X}} G_W(I\bar{e}_h) \cdot \lambda_{\mathrm{I}} g\left(\bar{Y}_h\right) \cdot \frac{\lambda_{\mathrm{I}}}{2}(1 + \frac{1}{\bar{Y}_h}) \\ &= \frac{\lambda_{\mathrm{I}}}{2}(1 + \frac{1}{\bar{Y}_h}) \cdot (\bar{e}_h)^{1/\varepsilon} \end{aligned}$$

The following string of equalities establishes the desired result:

$$\begin{aligned} \frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} &= -\frac{B}{\lambda_{\mathrm{I}}(B+C-A)} \\ &= -\frac{\frac{\lambda_{\mathrm{I}}}{2}(1+\frac{1}{\bar{Y}_h}) \cdot (\bar{e}_h)^{1/\varepsilon}}{\lambda_{\mathrm{I}}\left(\frac{\lambda_{\mathrm{I}}}{2}(1+\frac{1}{\bar{Y}_h}) \cdot (\bar{e}_h)^{1/\varepsilon} + \frac{1}{\varepsilon}(\bar{e}_h)^{1/\varepsilon - 1} - R(I\bar{e}_h) \cdot (\bar{e}_h)^{1/\varepsilon - 1}\right)} \\ &= -\frac{(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h}{\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))}. \end{aligned}$$

∎

Lemma A-4 shows $R(E)$ is strictly decreasing with $\lim_{E \downarrow 0} R(E) = \frac{1}{4}$, so for all $E > 0$, $R(E) \leq \frac{1}{4} < 1/\varepsilon$. Hence, in (A-12) the denominator is strictly larger than $\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_h)\bar{e}_h$, implying $0 < -\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} < \frac{1}{\lambda_{\mathrm{I}}}$. Therefore, $\frac{\partial \bar{Y}_h}{\partial \tau_{\mathrm{A}}} = 1 + \lambda_{\mathrm{I}} \frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} > 0$, so that $\bar{Y}_h$ also increases in τ_{A}.

3. $\bar{X}_h$ in I and τ_{A}.

Since $\bar{X}_h = W'(I\bar{e}_h)$, we have $\frac{\partial \bar{X}_h}{\partial \tau_{\mathrm{A}}} = W'(I\bar{e}_h)\, I\, \frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} < 0, \quad \frac{\partial \bar{X}_h}{\partial I} = W'(I\bar{e}_h)\left(\bar{e}_h + I\, \frac{\partial \bar{e}_h}{\partial I}\right) > 0$, as desired.

Appendix B Online Appendix for AI, Human Cognition and Knowledge Collapse

B.1 Proof of Proposition 6

We first prove Part (i). Recall that in the proof of Proposition 5, we derive the formula for the complete-collapse threshold as (here we define $\sup \emptyset = 0$)

$$\tau_{\mathrm{A}}^{c}(I) = \sup S(I), \qquad S(I) := \{\tau \geq 0 : \exists e > 0 \text{ s.t. } D(e; \tau_{\mathrm{A}}, I) = 0\}.$$

Here, we write τ_{A}^{c} as $\tau_{\mathrm{A}}^{c}(I)$ to highlight its dependency on I.

If $\tau_{\mathrm{A}}^{c}(I) > 0$, there is a tangency at the threshold:

$$\exists\, \bar{e} > 0 : \quad D(\bar{e}; \tau_{\mathrm{A}}^{c}(I), I) = 0, \qquad \partial_e D(\bar{e}; \tau_{\mathrm{A}}^{c}(I), I) = 0. \tag{B-1}$$

Note that $D(e; \tau_{\mathrm{A}}, I)$ is strictly increasing in I pointwise in (τ_{A}, e). Let $I_2 > I_1$. If $\tau \in S(I_1)$ there exists $\bar{e} > 0$ with $D(\bar{e}; \tau, I_1) = 0$, hence $D(\bar{e}; \tau, I_2) > 0$. Recall that we have shown that $D(\cdot; \tau, I_2)$ is continuous, strictly negative near 0 and $+\infty$. So $D(\cdot; \tau, I_2)$ must have two positive roots. Therefore, $\tau \in S(I_2)$, which implies that $S(I_1) \subseteq S(I_2)$ and

$$\tau_{\mathrm{A}}^{c}(I_2) = \sup S(I_2) \;\geq\; \sup S(I_1) = \tau_{\mathrm{A}}^{c}(I_1),$$

so $\tau_{\mathrm{A}}^{c}(I)$ is weakly increasing in I.

If $\tau_{\mathrm{A}}^{c}(I_1) > 0$, then by (B-1) there exists $\bar{e} > 0$ such that $D(\bar{e}; \tau_{\mathrm{A}}^{c}(I_1), I_1) = 0$. Strict monotonicity of D in I again implies that $D(\bar{e}; \tau_{\mathrm{A}}^{c}(I_1), I_2) > 0$. Since $D(\cdot; \tau_{\mathrm{A}}^{c}(I_1), I_2)$ is continuous, negative near 0 and $+\infty$, a small increase of τ above $\tau_{\mathrm{A}}^{c}(I_1)$ and a small perturbation of $\bar{e}$ restore a root; so there exists $\tau' > \tau_{\mathrm{A}}^{c}(I_1)$ with $D(\bar{e}'; \tau', I_2) = 0$ for some $\bar{e}' > 0$. Thus $\tau_{\mathrm{A}}^{c}(I_2) \geq \tau' > \tau_{\mathrm{A}}^{c}(I_1)$, proving strict monotonicity whenever $\tau_{\mathrm{A}}^{c}(I_1) > 0$.

The proof for part (ii) (the Δ_{X} part) is identical; it suffices to note that $D(e; \tau_{\mathrm{A}}, I)$ is also strictly increasing in Δ_{X} pointwise in (τ_{A}, e).

B.2 Proof of Proposition 7 and Proposition 8

We divide the proof in several parts:

1. $\bar{e}_m,\ \bar{e}_h$ in I and τ_{A}.

For $\varepsilon > 4$ and each $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^{c}$, Proposition 5 implies that $D(e; \tau_{\mathrm{A}}, I)$ has exactly two positive roots $0 < \bar{e}_m < \bar{e}_h$ with

$$D_e(\bar{e}_m; \tau_{\mathrm{A}}, I) > 0, \qquad D_e(\bar{e}_h; \tau_{\mathrm{A}}, I) < 0.$$

The implicit function theorem implies that $\bar{e}_k$ ($k \in \{m, h\}$) is C^1 in (τ_{A}, I) and

$$\frac{\partial \bar{e}_k}{\partial \tau} = -\frac{\partial \tau D(\bar{e}_k; \tau_{\mathrm{A}}, I)}{\partial_e D(\bar{e}_k; \tau_{\mathrm{A}}, I)}, \qquad \frac{\partial \bar{e}_k}{\partial I} = -\frac{\partial_I D(\bar{e}_k; \tau_{\mathrm{A}}, I)}{\partial_e D(\bar{e}_k; \tau_{\mathrm{A}}, I)},$$

which implies

$$\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} < 0, \quad \frac{\partial \bar{e}_m}{\partial \tau_{\mathrm{A}}} > 0, \qquad \frac{\partial \bar{e}_h}{\partial I} > 0, \quad \frac{\partial \bar{e}_m}{\partial I} < 0.$$

Thus the high steady–state effort decreases in τ and increases in I, while the middle root moves in the opposite direction.

2. $\bar{X}_m$, $\bar{X}_h$ in I and τ_{A}.

By definition,

$$\bar{X}_k = W(I\bar{e}_k), \qquad k \in \{m, h\},$$

and W is strictly increasing. Since $\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} < 0$ and $\frac{\partial \bar{e}_m}{\partial \tau_{\mathrm{A}}} > 0$, we have

$$\frac{\partial \bar{X}_h}{\partial \tau_{\mathrm{A}}} < 0, \qquad \frac{\partial \bar{X}_m}{\partial \tau_{\mathrm{A}}} > 0.$$

This proves the τ_{A}–comparative statics of $\bar{X}_h$ and $\bar{X}_m$.

For the I–comparative statics of $\bar{X}_h$ and $\bar{X}_m$ it is convenient to work with the law of motion $X_{t+1} = F(X_t; I, \tau_{\mathrm{A}})$ directly. For $\tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$ there are three steady states

$$0 = \bar{X}_l < \bar{X}_m < \bar{X}_h,$$

and the sign pattern of $F(X; I, \tau_{\mathrm{A}}) - X$ is

$$F(X; I, \tau_{\mathrm{A}}) - X \begin{cases} < 0, & X \in (0, \bar{X}_m), \\ > 0, & X \in (\bar{X}_m, \bar{X}_h), \\ < 0, & X \in (\bar{X}_h, \infty). \end{cases}$$

Moreover, Proposition 2 states that $F(X; I, \tau_{\mathrm{A}})$ is strictly increasing in I for every $X > 0$.

Fix τ_{A} and take $I_2 > I_1$. Write $F_j(X) = F(X; I_j, \tau_{\mathrm{A}})$ and $\bar{X}_{m,j} = \bar{X}_m(I_j, \tau_{\mathrm{A}})$, $j = 1, 2$. At $X = \bar{X}_{m,1}$ we have $F_1(\bar{X}_{m,1}) = \bar{X}_{m,1}$, so $F_2(\bar{X}_{m,1}) - \bar{X}_{m,1} > 0$. On the other hand, for $X > 0$ sufficiently small we have $F_2(X) - X < 0$. By continuity, $F_2(X) - X$ crosses zero at some smallest positive X strictly less than $\bar{X}_{m,1}$; this root is exactly $\bar{X}_{m,2}$. Therefore, we have $\bar{X}_m(I_2, \tau) < \bar{X}_m(I_1, \tau)$, which shows that $\bar{X}_m$ is strictly decreasing in I. Identically, $\bar{X}_h$ is strictly increasing in I

3. $\bar{Y}_h$ in I.

At the high steady state $\bar{Y}_h = \sigma^{-2} + \lambda_{\mathrm{I}} \bar{e}_h + \tau_{\mathrm{A}}$, so

$$\frac{\partial \bar{Y}_h}{\partial I} = \lambda_{\mathrm{I}} \frac{\partial \bar{e}_h}{\partial I} > 0.$$

4. $\bar{Y}_h$ in τ_{A}.

Lemma A-5 applies to the high equilibrium here: we have

$$\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} = -\frac{(1 + 1/\bar{Y}_h)\, \bar{e}_h}{\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_h)\, \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))}.$$

where both the numerator and the denominator are strictly positive.

Hence

$$\frac{\partial \bar{Y}_h}{\partial \tau_{\mathrm{A}}} = 1 + \lambda_{\mathrm{I}} \frac{\partial \bar{e}_h}{\partial \tau} = \frac{2(1/\varepsilon - R(I\bar{e}_h)}{\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_h)\, \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))}.$$

The denominator is positive, so $\operatorname{sign}\left(\frac{\partial \bar{Y}_h}{\partial \tau_{\mathrm{A}}}\right) = \operatorname{sign}(1/\varepsilon - R(I\bar{e}_h))$.

Lemma A-4 shows that $R(E)$ is strictly decreasing on $(0, \infty)$ with $\lim_{E \downarrow 0} R(E) = \frac{1}{4}$ and $\lim_{E \to \infty} R(E) = 0$. Since $1/\varepsilon \in (0, \frac{1}{4})$, there is a unique $E_* > 0$ such that $R(E_*) = 1/\varepsilon$.

We have shown that $\bar{e}_h$ is continuous and strictly decreasing in τ_{A} on $[0, \tau_{\mathrm{A}}^c)$. At the complete-collapse threshold $\tau_{\mathrm{A}} = \tau_{\mathrm{A}}^c$ the two positive roots coalesce at a tangency $\bar{e}_\star > 0$ with $D(\bar{e}_\star; \tau_{\mathrm{A}}^c, I) = 0$ and $D_e(\bar{e}_\star; \tau_{\mathrm{A}}^c, I) = 0$. The tangency condition implies

$$\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_\star)\bar{e}_\star + 2(1/\varepsilon - R(I\bar{e}_\star)) = 0,$$

hence $R(I\bar{e}_\star) > 1/\varepsilon$ and $I\bar{e}_\star < E_*$. By continuity of $\bar{e}_h$ in τ_{A}, for τ_{A} sufficiently close to τ_{A}^c we therefore have $I\bar{e}_h < E_*$.

Define

$$\hat{\tau}_A := \inf\{\tau_{\mathrm{A}} \in [0, \tau_{\mathrm{A}}^c) : I\bar{e}_h < E_*\}.$$

When $\tau_{\mathrm{A}}^c > 0$, the set in braces is an non-empty interval, and $\hat{\tau}_A \in [0, \tau_{\mathrm{A}}^c)$ is well defined. By construction,

$$1/\varepsilon - R(E(\tau_{\mathrm{A}})) \begin{cases} > 0, & \text{if } \tau_{\mathrm{A}} \in [0, \hat{\tau}_{\mathrm{A}}), \\ < 0, & \text{if } \tau_{\mathrm{A}} \in (\hat{\tau}_{\mathrm{A}}, \tau_{\mathrm{A}}^c). \end{cases}$$

Using the sign identity above, we conclude that

$$\frac{\partial \bar{Y}_h}{\partial \tau_{\mathrm{A}}} > 0 \quad \text{for } \tau_{\mathrm{A}} \in [0, \hat{\tau}_A), \qquad \frac{\partial \bar{Y}_h}{\partial \tau_{\mathrm{A}}} < 0 \quad \text{for } \tau_{\mathrm{A}} \in (\hat{\tau}_A, \tau_{\mathrm{A}}^c),$$

as desired.

B.3 Proof of Proposition 9

Fix τ_{A} and other parameters, and suppose the high-knowledge steady state exists. Let $\bar{X}_h(I) > 0$ denote the corresponding steady-state public precision. Recall the (within-cohort) value function

$$\bar{U}(\bar{X}, \tau_{\mathrm{A}}) = G(\bar{X}) \cdot \Delta_{\mathrm{G}} + \max_{e' \geq 0} \left[G(\sigma^{-2} + \lambda_{\mathrm{I}} e' + \tau_{\mathrm{A}}) \cdot G(\bar{X}) \Delta_{\mathrm{X}} - \frac{\varepsilon}{\varepsilon + 1} (e')^{\frac{\varepsilon+1}{\varepsilon}} \right],$$

so that $\bar{U}^+(I) = \bar{U}(\bar{X}_h(I), \tau_{\mathrm{A}})$.

First, note that $\bar{U}(X, \tau_{\mathrm{A}})$ is strictly increasing in X for $X > 0$. Second, comparative static results in Proposition 4 (ii) and 8 (ii) imply that $\bar{X}_h(I)$ is strictly increasing in I. Therefore, $\bar{U}^+(I)$ is strictly increasing in I.

B.4 Proof of Proposition 10

By Propositions 3, when $\varepsilon < 4$, the high-knowledge steady state is always unique for all τ_{A} and changes smoothly on τ_{A}. So by the Envelope theorem,

$$\begin{aligned}
\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} &= g(\bar{Y}_h) \cdot G_W(I\bar{e}_h) \Delta_{\mathrm{X}} + \frac{\partial G_W(I\bar{e}_h)}{\partial \tau_{\mathrm{A}}} \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}) \\
&= \underbrace{g(\bar{Y}_h) \cdot G_W(I\bar{e}_h) \Delta_{\mathrm{X}}}_{\mathrm{DE}} - \underbrace{G'_W(I\bar{e}_h) \cdot I \cdot \left(-\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} \right) \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}})}_{\mathrm{IE}}.
\end{aligned}$$

To study the sign of $\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}}$ it is therefore convenient to work with the ratio $\frac{\mathrm{IE}}{\mathrm{DE}}$. We have the expression:

$$\begin{aligned}
\frac{\mathrm{IE}}{\mathrm{DE}} &= \frac{G'_W(I\bar{e}_h) \cdot I}{G_W(I\bar{e}_h)} \cdot \frac{1}{g(\bar{Y}_h)} \cdot \left(-\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} \right) \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{\Delta_{\mathrm{X}}} \\
&= \frac{R(I\bar{e}_h)}{\bar{e}_h} \cdot \frac{1}{g(\bar{Y}_h)} \cdot \frac{(1 + \frac{1}{\bar{Y}_h}) \cdot \bar{e}_h}{\lambda_{\mathrm{I}}(1 + \frac{1}{\bar{Y}_h}) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{\Delta_{\mathrm{X}}}.
\end{aligned}$$

Rearranging terms, we have expressed $\frac{\mathrm{IE}}{\mathrm{DE}}$ as a function of $\bar{e}_h$ and $\bar{Y}_h$:

$$\frac{\mathrm{IE}}{\mathrm{DE}} = \frac{(1 + 1/\bar{Y}_h) \cdot R(I\bar{e}_h)}{\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_h) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{g(\bar{Y}_h)\Delta_{\mathrm{X}}}. \tag{B-2}$$

Lemma B-1. *When $\bar{Y}_h \geq \sqrt{2} - 1$, the mapping*

$$(\bar{e}_h, \bar{Y}_h) \longmapsto \frac{\mathrm{IE}}{\mathrm{DE}}$$

given by (B-2) *is strictly* decreasing *in $\bar{e}_h$ and strictly* increasing *in $\bar{Y}_h$.*

Proof. Since $R(\cdot)$ is a strictly decreasing function, the monotonicity with respect to $\bar{e}_h$ is easy to check. For the $\bar{Y}_h$ part, it suffices to prove that $\frac{1+\frac{1}{\bar{Y}_h}}{g(\bar{Y}_h)}$ is increasing in $\bar{Y}_h$. In fact,

$$\frac{1+\frac{1}{\bar{Y}_h}}{g(\bar{Y}_h)} = \sqrt{2\pi} e^{\frac{\bar{Y}_h}{2}} \left(\sqrt{\bar{Y}_h} + \frac{1}{\sqrt{\bar{Y}_h}} \right),$$

which is indeed increasing in $\bar{Y}_h$ due to the assumption that $\bar{Y}_h \geq \sqrt{2} - 1$. ∎

Proposition 4 shows that as τ_{A} increases, $\bar{e}_h$ decreases while $\bar{Y}_h$ increases, making the ratio $\frac{\mathrm{IE}}{\mathrm{DE}}$ strictly increasing in τ_{A} on $(0, \infty)$. We separate two cases:

Case I: $\frac{\mathrm{IE}}{\mathrm{DE}}\Big|_{\tau_{\mathrm{A}}=0} \geq 1$**.** In this case, the above arguments guarantees that for all $\tau_{\mathrm{A}} > 0$, we have $\frac{\mathrm{IE}}{\mathrm{DE}} > 1$, so that $\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} < 0$. Therefore choosing $\tau_{\mathrm{A}}^\star = 0$ and statements (i) and (ii) of the proposition follow.

Case II: $\frac{\mathrm{IE}}{\mathrm{DE}}\Big|_{\tau_{\mathrm{A}}=0} < 1$**.** From (B-2), we have that for every $\tau_{\mathrm{A}} \in (0, +\infty)$, $\frac{\mathrm{IE}}{\mathrm{DE}}$ is strictly positive, continuous and strictly increasing in τ_{A}. Moreover, it holds that $\lim_{\tau_{\mathrm{A}} \uparrow \infty} \frac{\mathrm{IE}}{\mathrm{DE}} = +\infty$. Combining with our case assumption, there exists a finite, unique threshold $\tau_{\mathrm{A}}^\star$ at which $\frac{\mathrm{IE}}{\mathrm{DE}} = 1$.

By strict monotonicity of $\frac{\mathrm{IE}}{\mathrm{DE}}$ with respect to τ_{A}, we conclude that:

$$\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} > 0 \quad \text{for } 0 < \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^\star, \qquad \frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} < 0 \quad \text{for } \tau_{\mathrm{A}} > \tau_{\mathrm{A}}^\star.$$

Let's now prove part (iii). From the first-order condition it is easy to show that $\bar{e}_h, \bar{X}_h \to 0$ as $\tau_{\mathrm{A}} \to +\infty$, which gives $\bar{U}^+(\tau_{\mathrm{A}}) \to 0$ as well.

B.5 Proof of Proposition 11

We restrict τ_{A} to the range $(0, \tau_{\mathrm{A}}^c)$ where the high-knowledge steady state exists: $\bar{X}_h > 0$. By Proposition 5 the high steady state is unique in this region and depends smoothly on τ_{A}, so we may differentiate $\bar{U}^+$ with respect to τ_{A}. By the Envelope theorem,

$$\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} = \underbrace{g(\bar{Y}_h) \cdot G_W(I\bar{e}_h)\Delta_{\mathrm{X}}}_{\mathrm{DE}} - \underbrace{G_W'(I\bar{e}_h) \cdot I \cdot \left(-\frac{\partial \bar{e}_h}{\partial \tau_{\mathrm{A}}} \right) \cdot (\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}})}_{\mathrm{IE}}.$$

Similar to the proof of Proposition 10, we can express $\frac{\mathrm{IE}}{\mathrm{DE}}$ as a function of $\bar{e}_h$ and $\bar{Y}_h$:

$$\frac{\mathrm{IE}}{\mathrm{DE}} = \frac{(1 + 1/\bar{Y}_h) \cdot R(I\bar{e}_h)}{\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_h) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{g(\bar{Y}_h)\Delta_{\mathrm{X}}}, \tag{B-3}$$

and Lemma B-1 continues to hold here.

By Proposition 8(i), comparative statics of the high steady states with respect to τ_{A} is as follows:

$$\bar{X}_h \downarrow, \qquad \bar{e}_h \downarrow \quad \text{for all } \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c,$$

while $\bar{Y}_h$ is increasing for small $\tau_{\mathrm{A}} < \hat{\tau}_A$ and eventually decreasing when $\hat{\tau}_A < \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$.

Here, we first show the following two arguments:

(a) Within the region where $0 \le \tau_{\mathrm{A}} < \hat{\tau}_A$ (so that $R(I\bar{e}_h) < 1/\varepsilon$), $\bar{e}_h$ is decreasing in τ_{A}, and $\bar{Y}_h$ is increasing in τ_{A}, so that $\frac{\mathrm{IE}}{\mathrm{DE}}$ is increasing in τ_{A}.

(b) Within the region where $\hat{\tau}_A \le \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$ (so that $R(I\bar{e}_h) \ge 1/\varepsilon$), even if $\bar{Y}_h$ is decreasing in τ_{A}, $\frac{\mathrm{IE}}{\mathrm{DE}}$ is still increasing in τ_{A}

Let's prove (b). When $R(I\bar{e}_h) > 1/\varepsilon$, recall

$$\frac{\mathrm{IE}}{\mathrm{DE}} = \frac{1}{g(\bar{Y}_h)} \cdot \frac{(1+\frac{1}{\bar{Y}_h}) \cdot R(I\bar{e}_h)}{\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{\Delta_{\mathrm{X}}}.$$

Lemma B-2. *When $\hat{\tau}_A \le \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$, $\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h))$ is decreasing in τ_{A}, moreover,*

$$\frac{\mathrm{d}\log(\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h + 2(1/\varepsilon - R(I\bar{e}_h)))}{\mathrm{d}\,\tau_{\mathrm{A}}} \le \frac{\mathrm{d}\log(\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h)}{\mathrm{d}\,\tau_{\mathrm{A}}}$$

Proof. Let

$$T_1 := \lambda_{\mathrm{I}}\left(1+\frac{1}{\bar{Y}_h}\right) \cdot \bar{e}_h, \quad T_2 := 2(1/\varepsilon - R(I\bar{e}_h)),$$

and let $T = T_1 + T_2$.

We first show that T_1 and T_2 are both strictly decreasing in τ_{A}. Note that

$$T_1 = \lambda_{\mathrm{I}}\left(\bar{e}_h + \frac{\bar{e}_h}{\sigma^{-2} + \lambda_I \bar{e}_h + \tau_{\mathrm{A}}}\right).$$

Since $\bar{e}_h$ is decreasing, both T_1 and T_2 are decreasing. So T is also decreasing.

Therefore, we have

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

On $[\hat{\tau}_{\mathrm{A}}, \tau_{\mathrm{A}}^c)$ we also have $T_2 \le 0$. Because $T_1' < 0$ and $T_1 + T_2 > 0$, dividing by the smaller positive denominator yields

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

Combining the last two inequalities gives $\frac{T'}{T} \leq \frac{T'_1}{T_1}$, i.e.

$$\frac{d}{d\tau_{\mathrm{A}}} \log T \leq \frac{d}{d\tau_{\mathrm{A}}} \log T_1,$$

as claimed. ∎

By this lemma, to show that $\frac{\mathrm{IE}}{\mathrm{DE}}$ is increasing in τ_{A}, it suffices to prove that

$$\frac{1}{g(\bar{Y}_h)} \cdot \frac{(1+\frac{1}{\bar{Y}_h}) \cdot R(I\bar{e}_h)}{\lambda_{\mathrm{I}}(1+\frac{1}{\bar{Y}_h}) \cdot \bar{e}_h} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{\Delta_{\mathrm{X}}} = \frac{1}{g(\bar{Y}_h)} \cdot \frac{R(I\bar{e}_h)}{\lambda_{\mathrm{I}} \cdot \bar{e}_h} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{\Delta_{\mathrm{X}}}$$

is increasing in τ_{A}.

Simplifying and omitting factors that does not depend on τ_{A}, we only need to show that

$$\frac{G'_W(I\bar{e}_h)}{G_W(I\bar{e}_h)} \cdot \frac{\Delta_{\mathrm{G}} + G(\bar{Y}_h)\Delta_{\mathrm{X}}}{g(\bar{Y}_h)}$$

is increasing in τ_{A}.

Here, the denominator $G_W(I\bar{e}_h)g(\bar{Y}_h) = \frac{\bar{e}_h^{1/\varepsilon}}{\lambda_{\mathrm{I}}\Delta_{\mathrm{X}}}$ is decreasing in τ_{A}. To finish the proof of (b), it suffices to show the following lemma.

Lemma B-3. *When $\hat{\tau}_A \leq \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$, $G'_W(I\bar{e}_h)G(\bar{Y}_h)$ is increasing in τ_{A}.*

Proof. Along the high steady-state path, define

$$E(\tau_{\mathrm{A}}) := I\bar{e}_h(\tau_{\mathrm{A}}), \qquad Y(\tau_{\mathrm{A}}) := \bar{Y}_h(\tau_{\mathrm{A}}) = \sigma^{-2} + \lambda_I \bar{e}_h(\tau_{\mathrm{A}}) + \tau_{\mathrm{A}}.$$

Differentiating the objective yields

$$\frac{d}{d\tau_{\mathrm{A}}} \left(G'_W(E(\tau_{\mathrm{A}}))\, G(Y(\tau_{\mathrm{A}})) \right) = G''_W(E)G(Y)E'(\tau_{\mathrm{A}}) + G'_W(E)g(Y)Y'(\tau_{\mathrm{A}}).$$

Thus it suffices to show that

$$G''_W(E)G(Y)E'(\tau_{\mathrm{A}}) + G'_W(E)g(Y)Y'(\tau_{\mathrm{A}}) \geq 0. \tag{B-4}$$

At the high steady state, $\bar{e}_h$ satisfies the steady-state effort equation (A-1):

$$G_W(I\bar{e}_h)g(\bar{Y}_h) = \frac{\bar{e}_h^{1/\varepsilon}}{\lambda_{\mathrm{I}}\Delta_{\mathrm{X}}} \tag{B-5}$$

Since the right-hand side of (B-5) is strictly decreasing in τ_{A}, so is the left-hand side. Differenti-

ating, we have

$$G'_W(E)g(Y)E'(\tau_{\mathrm{A}}) + G_W(E)g'(Y)Y'(\tau_{\mathrm{A}}) \leq 0. \tag{B-6}$$

Dividing (B-6) by $E'(\tau_{\mathrm{A}}) < 0$ and then by $G_W(E)g'(Y) < 0$ yields the upper bound

$$\frac{Y'(\tau_{\mathrm{A}})}{E'(\tau_{\mathrm{A}})} \leq -\frac{G'_W(E)g(Y)}{G_W(E)g'(Y)}. \tag{B-7}$$

Because $E'(\tau_{\mathrm{A}}) < 0$ and $G'_W(E)g(Y) > 0$, inequality (B-4) is equivalent to

$$\frac{Y'(\tau_{\mathrm{A}})}{E'(\tau_{\mathrm{A}})} \leq -\frac{G''_W(E)G(Y)}{G'_W(E)g(Y)}. \tag{B-8}$$

Hence, by (B-7), it suffices to prove that

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

which is equivalent to

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

In what follows, we prove that the right-hand side of (B-9) is weakly smaller than 1, while the left-hand side weakly bigger than 1.

Recall that $g'(Y) = -\frac{Y+1}{2Y}g(Y)$, therefore

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

Let $z = \sqrt{Y}$. Using $g(Y) = \phi(z)/z$ and $G(Y) = 2\Phi(z) - 1$, we obtain

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

For the left-hand side of (B-9), note that

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

Substituting into the left-hand side gives

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

We claim $W''(E) < 0$ for all $E > 0$. Indeed, Lemma A-3 implies

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

Since $S'(E) < 0$ and that $W'(E)$ is strictly increasing in $S(E)$, we have $W''(E) < 0$.

Consequently, the second term in (B-10) is nonnegative:

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

Moreover, analogous to the analysis of the right-hand side of (B-9), we have $-\frac{g(W)^2}{G(W)\, g'(W)} \leq 1$, so the first term in (B-10) is at least 1. Hence

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

Therefore (B-9) holds, so that

$$\frac{d}{d\tau_{\mathrm{A}}} \left(G'_W(E(\tau_{\mathrm{A}})) G(Y(\tau_{\mathrm{A}})) \right) \geq 0,$$

as desired. ∎

Combining (a) and (b), we conclude: the ratio $\frac{\mathrm{IE}}{\mathrm{DE}}$ is strictly increasing in τ_{A} as long as $0 \leq \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$.

We separate two cases:

Case I: $\frac{\mathrm{IE}}{\mathrm{DE}}\Big|_{\tau_{\mathrm{A}}=0} \geq 1$**.** In this case, the above arguments guarantees that for all $0 < \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c$, we have $\frac{\mathrm{IE}}{\mathrm{DE}} > 1$, so that $\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} < 0$. Therefore choosing $\tau_{\mathrm{A}}^\star = 0$ and Statements (i) and (ii) of the proposition follow.

Case II: $\frac{\mathrm{IE}}{\mathrm{DE}}\Big|_{\tau_{\mathrm{A}}=0} < 1$**.** From (B-3), we have that for every $\tau_{\mathrm{A}} \in (0, \hat{\tau}_A)$, $\frac{\mathrm{IE}}{\mathrm{DE}}$ is strictly positive, continuous and strictly increasing in τ_{A}. Moreover, it is easy to verify that $\lim_{\tau_{\mathrm{A}} \uparrow \tau_{\mathrm{A}}^c} \frac{\mathrm{IE}}{\mathrm{DE}} = +\infty$. Combining with our case assumption, there exists a finite, unique threshold $\tau_{\mathrm{A}}^\star < \hat{\tau}_A$ at which $\frac{\mathrm{IE}}{\mathrm{DE}} = 1$. We conclude that:

$$\frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} > 0 \quad \text{for } 0 < \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^\star, \qquad \frac{\partial \bar{U}^+}{\partial \tau_{\mathrm{A}}} < 0 \quad \text{for } \tau_{\mathrm{A}}^\star < \tau_{\mathrm{A}} < \tau_{\mathrm{A}}^c.$$

Part (iii) follows directly from the definition of τ_{A}^c.

B.6 Proof of Proposition 12

We first consider the asymptotic for $\tau_{\mathrm{A}}^c(I)$. As $I \to +\infty$, we must have $\tau_{\mathrm{A}}^c(I) > 0$, so that the tangency condition (B-1) applies. Denote by $\bar{e}_I$ and $\bar{Y}_I$ the effort and idiosyncratic precision at the tangency point corresponding to I and $\tau_{\mathrm{A}} = \tau_{\mathrm{A}}^c(I)$. Using the formulas derived from the proof of Proposition 4, (B-1) implies that

$$0 = \Delta_{\mathrm{X}} \cdot G_W(\bar{E}_I) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} \bar{e}_I + \tau_{\mathrm{A}}) - \bar{e}_I^{1/\varepsilon} \tag{B-11}$$

$$0 = \lambda_{\mathrm{I}}(1 + 1/\bar{Y}_I)\bar{e}_I + 2(1/\varepsilon - R(\bar{E}_I)) \tag{B-12}$$

where

$$\bar{E}_I := I\bar{e}_I, \qquad \bar{Y}_I := \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_I + \tau_{\mathrm{A}}^c(I).$$

Lemma A-4 shows that $R(E)$ is strictly decreasing on $(0, \infty)$ with $\lim_{E \downarrow 0} R(E) = \frac{1}{4}$ and $\lim_{E \to \infty} R(E) = 0$. Hence, since $1/\varepsilon < \frac{1}{4}$, there is a unique $E_* > 0$ such that $R(E_*) = 1/\varepsilon$.

First, we show that $\bar{E}_I$ converges to E^* as $I \to \infty$. From (B-12) we have $R(\bar{E}_I) > 1/\varepsilon$, so $\bar{E}_I \leq E_*$ for all I. As a result, we have

$$\bar{e}_I = \bar{E}_I / I \leq E^*/I \to 0.$$

As a result, $\lambda_{\mathrm{I}}(1 + 1/\bar{Y}_I)\bar{e}_I \to 0$ and (B-12) yields $R(\bar{E}_I) \to 1/\varepsilon$. Therefore, $\bar{E}_I \to E_* \in (0, \infty)$.

Next, from (B-11) we have

$$g(\bar{Y}_I) = \frac{\bar{e}_I^{1/\varepsilon}}{\Delta_{\mathrm{X}} \lambda_{\mathrm{I}} G_W(\bar{E}_I)} = \frac{\bar{E}_I^{1/\varepsilon}}{\Delta_{\mathrm{X}} \lambda_{\mathrm{I}} G_W(\bar{E}_I)} I^{-(1/\varepsilon)}.$$

Since $E_I \to E_*$ and $G_W(E_*) > 0$, the prefactor converges to some constant $C_0 > 0$, so

$$g(\bar{Y}_I) \sim C_0 \, I^{-(1/\varepsilon)} \qquad (I \to \infty).$$

Because g is strictly decreasing with $g(y) \downarrow 0$ as $y \to \infty$, this implies $\bar{Y}_I \to \infty$. Using the expression $g(y) = (2\pi)^{-1/2} y^{-1/2} e^{-y/2}$, we obtain:

$$\bar{Y}_I = (2/\varepsilon) \log I + O(\log \log I) \qquad (I \to \infty).$$

Finally, from the expression of $\bar{Y}_I$ and the fact that $\bar{e}_I \to 0$, we have

$$\tau_{\mathrm{A}}^c(I) = (2/\varepsilon) \log I + O(\log \log I),$$

and therefore

$$\lim_{I \to \infty} \frac{\tau_{\mathrm{A}}^c(I)}{\log I} = 2/\varepsilon.$$

This completes the proof.

Second, we show the asymptotic of $\tau_{\mathrm{A}}^\star(I)$. Denote by $e_I^\star$ and $Y_I^\star$ the effort and idiosyncratic precision at the high steady state corresponding to I and $\tau_{\mathrm{A}} = \tau_{\mathrm{A}}^\star(I)$:

$$e_I^\star := \bar{e}_h(\tau_{\mathrm{A}}^\star(I), I), \quad Y_I^\star := \sigma^{-2} + \lambda_{\mathrm{I}} e_I^\star + \tau_{\mathrm{A}}^\star(I),$$

and also set $E_I^\star := I e_I^\star$ for convenience.

At any positive steady state we have the effort condition

$$(e_I^\star)^{1/\varepsilon} = \Delta_{\mathrm{X}} \lambda_{\mathrm{I}} G_W(E_I^\star) g(Y_I^\star). \tag{B-13}$$

If the welfare optimum $\tau_{\mathrm{A}}^\star(I) > 0$ (which is the case when $I \to +\infty$), our previous analysis implies $\mathrm{IE} \,/\, \mathrm{DE} = 1$, which is

$$1 = \frac{\mathrm{IE}}{\mathrm{DE}} = \frac{(1 + 1/Y_I^\star) \cdot R(E_I^\star)}{\lambda_{\mathrm{I}}(1 + 1/Y_I^\star) \cdot e_I^\star + 2(1/\varepsilon - R(E_I^\star))} \cdot \frac{\Delta_{\mathrm{G}} + G(Y_I^\star)\Delta_{\mathrm{X}}}{g(Y_I^\star)\Delta_{\mathrm{X}}}.$$

Solving for $g(Y_I^\star)$ from (B-13) and substituting it yields

$$(e_I^\star)^{1/\varepsilon}[\lambda_{\mathrm{I}}(1+1/Y_I^\star)\cdot e_I^\star + 2(1/\varepsilon - R(E_I^\star))] = \lambda_{\mathrm{I}} G_W(E_I^\star)(1+1/Y_I^\star) R(E_I^\star)(\Delta_{\mathrm{G}} + G(Y_I^\star)\Delta_{\mathrm{X}}). \tag{B-14}$$

We first show that along the optimal sequence $I \to \infty$,

$$E_I^\star \to \infty, \qquad e_I^\star \to 0.$$

If $E_{I_n}^\star \to \tilde{E} < \infty$ along some $I_n \to \infty$, then $e_{I_n}^\star = E_{I_n}^\star / I_n \to 0$, while $G_W(E_{I_n}^\star) \to G_W(\tilde{E}) > 0$ and $R(E_{I_n}^\star) \to R(\tilde{E}) > 0$. Letting $n \to \infty$ in (B-14), the left-hand side tends to 0 and the right-hand side to at least $\lambda_{\mathrm{I}} G_W(\tilde{E}) R(\tilde{E})(\Delta_{\mathrm{G}} + \Delta_{\mathrm{X}} G(\sigma^{-2})) > 0$, a contradiction. Thus $E_I^\star \to \infty$.

Suppose instead $e_{I_n}^\star \to \tilde{e} < \infty$ along some $I_n \to \infty$. For large E, we have

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

with constants $G_\infty > 0$, $c_R > 0$. Substituting into (B-14) and letting $n \to \infty$, the right-hand side converges to zero, while the left-hand side converges to at least $\tilde{e}^{1/\varepsilon}(\lambda_{\mathrm{I}} \tilde{e} + 2(1/\varepsilon)) > 0$; again a contradiction. Hence $e_I^\star \to 0$, which also implies $E_I^\star = o(I)$. From (B-13) we also get $Y_I^\star \to \infty$.

Plugging all the asymptotics into (B-14), we have

$$\lim_{I\to\infty} \frac{(e_I^\star)^{1/\varepsilon}}{R(E_I^\star)} = \lim_{I\to\infty} \frac{\lambda_{\mathrm{I}} G_W(E_I^\star)(1 + 1/Y_I^\star)(\Delta_{\mathrm{G}} + G(Y_I^\star)\Delta_{\mathrm{X}})}{\lambda_{\mathrm{I}}(1 + 1/Y_I^\star) \cdot e_I^\star + 2(1/\varepsilon - R(E_I^\star))} = \frac{\lambda_{\mathrm{I}} G_\infty (\Delta_{\mathrm{G}} + \Delta_{\mathrm{X}})}{2/\varepsilon}. \tag{B-15}$$

Using $R(E) = c_R/E + O(E^{-2})$ and $E_I^\star = Ie_I^\star$, (B-15) becomes

$$(e_I^\star)^{\frac{\varepsilon+1}{\varepsilon}} \sim \frac{\lambda_{\mathrm{I}} G_\infty(\Delta_{\mathrm{G}} + \Delta_{\mathrm{X}})c_R}{2/\varepsilon} \cdot I^{-1},$$

hence

$$e_I^\star \sim C_E \cdot I^{-\frac{\varepsilon}{\varepsilon+1}}, \tag{B-16}$$

for some constant $C_E > 0$.

From the steady-state condition (B-13) and the large-E limit $G_W(E) \to G_\infty$ we obtain

$$g(Y_I^\star) = \frac{(e_I^\star)^{1/\varepsilon}}{\Delta_{\mathrm{X}}\lambda_{\mathrm{I}} G_W(E_I^\star)} \sim C_g\, I^{-1/(1+\varepsilon)},$$

for some constant $C_g > 0$, by (B-16).

Using the expression $g(y) = (2\pi)^{-1/2} y^{-1/2} e^{-y/2}$, we conclude:

$$Y_I^\star = \frac{2}{\varepsilon+1}\log I + O(\log\log I).$$

Using the definition of $Y_I^\star$ and that $e_I^\star \to 0$, we obtain

$$\tau_{\mathrm{A}}^\star(I) = \frac{2}{\varepsilon+1}\log I + O(\log\log I).$$

Thus we have $\lim_{I\to\infty} \frac{\tau_{\mathrm{A}}^\star(I)}{\log I} = \frac{2}{\varepsilon+1}$, which is the desired asymptotic bound.

B.7 Proof of Proposition 13

Consider an eventually-constant information policy $\kappa \in \mathcal{K}^{EC}$, so there exist $M < \infty$ and $\bar\kappa$ such that $\kappa_t = \bar\kappa$ for all $t \geq M$. Let $\bar\tau := \tau_{\mathrm{A}}(\bar\kappa)$ be the corresponding effective precision. From period M onward, the induced equilibrium recursion coincides with the recursion in an economy with constant effective precision $\bar\tau$, starting from initial condition X_M:

$$X_{t+1} = F_{\bar\tau}(X_t), \qquad t \geq M,$$

where $F_{\bar\tau}$ is the constant-$\bar\tau$ transition map. Proposition 5 then implies the convergence of X_t:

- If $\bar\tau \geq \tau_{\mathrm{A}}^c$, then 0 is the unique globally stable steady state, hence $X_t \to 0$ as $t \to \infty$.
- If $\bar\tau < \tau_{\mathrm{A}}^c$, then there exist three steady states $0 = \bar X_\ell(\bar\tau) < \bar X_m(\bar\tau) < \bar X_h(\bar\tau)$ and the basin partition holds: $X_M > \bar X_m(\bar\tau) \Rightarrow X_t \to \bar X_h(\bar\tau)$, while $X_M < \bar X_m(\bar\tau) \Rightarrow X_t \to 0$.

Therefore we have

$$U_t(\kappa) \to \begin{cases} \bar{U}^+(\bar{\tau}), & \text{if } \bar{\tau} < \tau_{\mathrm{A}}^c \text{ and } X_M > \bar{X}_m(\bar{\tau}), \\ 0, & \text{otherwise.} \end{cases}$$

Since a convergent bounded sequence has the same Cesàro limit as its pointwise limit,

$$U(\kappa) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} U_t(\kappa) = \begin{cases} \bar{U}^+(\bar{\tau}), & \text{if } \bar{\tau} < \tau_{\mathrm{A}}^c \text{ and } X_M > \bar{X}_m(\bar{\tau}), \\ 0, & \text{otherwise.} \end{cases} \tag{B-17}$$

We then construct the optimal two-phase policy. Consider the full-suppression regime $\kappa_t = 0\ \forall t$, equivalently $\tau_t = 0\ \forall t$. Because $X_1 > \bar{X}_m(0)$ by assumption, Proposition 5 implies that under $\tau_t = 0\ \forall t$,

$$X_t^{supp} \to \bar{X}_h(0) \quad \text{as } t \to \infty.$$

Next, $\tau^\star \le \tau_{\mathrm{A}}^\star < \tau_{\mathrm{A}}^c$, so the high-knowledge steady state under $\tau^\star$ exists and satisfies

$$0 < \bar{X}_m(\tau^\star) < \bar{X}_h(\tau^\star).$$

Moreover, since the high steady state $\bar{X}_h(\tau)$ is (strictly) decreasing in τ on $[0, \tau_{\mathrm{A}}^c)$, we have

$$\bar{X}_h(0) \ge \bar{X}_h(\tau^\star) > \bar{X}_m(\tau^\star).$$

Hence $\bar{X}_h(0) - \bar{X}_m(\tau^\star) > 0$, and by convergence of X_t there exists a finite $M^\star$ such that

$$X_{M^\star} > \bar{X}_m(\tau^\star).$$

Define the two-phase policy κ^{2ph} by $\kappa_t^{2ph} = 0$ for $t < M^\star$ and $\kappa_t^{2ph} = \kappa^\star$ for $t \ge M^\star$. By construction, the realized public precision at the switching date satisfies $X_{M^\star}(\kappa^{2ph}) = X_{M^\star}^{supp} > \bar{X}_m(\tau^\star)$. Applying (B-17) with $\bar{\tau} = \tau^\star$ yields $U(\kappa^{2ph}) = \bar{U}^+(\tau^\star)$, which proves part (i).

Finally, we prove the optimality of κ^{2ph}. Let $\kappa \in \mathcal{K}^{EC}$ be arbitrary and let $\bar{\tau}$ be its tail effective precision. (B-17) implies that either $U(\kappa) = 0$ or $U(\kappa) = \bar{U}^+(\bar{\tau})$. In either case,

$$U(\kappa) \le \max_{\tau \in [0, \tau_{\mathrm{A}}]} \bar{U}^+(\tau).$$

By Proposition 11, $\bar{U}^+(\tau)$ is increasing on $(0, \tau_{\mathrm{A}}^\star)$ and decreasing on $(\tau_{\mathrm{A}}^\star, \tau_{\mathrm{A}}^c)$, with $\bar{U}^+(\tau) = 0$ for $\tau \ge \tau_{\mathrm{A}}^c$. Therefore the maximizer of $\bar{U}^+(\tau)$ over the feasible set $[0, \tau_{\mathrm{A}}]$ is $\tau^\star$.

B.8 Proof of Proposition 14

Here, agentic AI with precision τ_{A} both (i) provides the idiosyncratic signal $s^A_{i,t} \sim \mathcal{N}(\theta_{i,t}, \tau_{\mathrm{A}}^{-1})$ and (ii) raises the effective aggregation scale from a baseline I_0 to

$$I(\tau_{\mathrm{A}}) = I_0 + \exp(\eta\tau_{\mathrm{A}}).$$

Fix $\tau_{\mathrm{A}} \geq 0$ and suppose a positive (high-knowledge) steady state exists, with per-agent effort $\bar{e}_h(\tau_{\mathrm{A}}) > 0$ and aggregate (island-level) effort

$$\bar{E}_h(\tau_{\mathrm{A}}) := I(\tau_{\mathrm{A}})\, \bar{e}_h(\tau_{\mathrm{A}}).$$

The steady-state public precision is $\bar{X}_h(\tau_{\mathrm{A}}) = W(\bar{E}_h(\tau_{\mathrm{A}}))$, where $W(\cdot)$ is continuous on $\mathbb{R}_+$ and satisfies $W(0) = 0$, and $W(E) > 0$ for $E > 0$.

Thus, to show $\bar{X}_h(\tau_{\mathrm{A}}) \to 0$ as $\tau_{\mathrm{A}} \to \infty$, it suffices to prove that

$$\bar{E}_h(\tau_{\mathrm{A}}) = I(\tau_{\mathrm{A}})\bar{e}_h(\tau_{\mathrm{A}}) \longrightarrow 0. \tag{B-18}$$

Here, we show that effort must decay at least exponentially in τ_{A}. At any positive steady state, per-agent effort $\bar{e}_h(\tau_{\mathrm{A}})$ satisfies the first-order condition:

$$\bar{e}_h(\tau_{\mathrm{A}})^{1/\varepsilon} = \Delta_X G(\bar{X}_h(\tau_{\mathrm{A}}))\lambda_{\mathrm{I}} g(\bar{Y}_h(\tau_{\mathrm{A}})), \qquad \bar{Y}_h(\tau_{\mathrm{A}}) = \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_h(\tau_{\mathrm{A}}) + \tau_{\mathrm{A}}.$$

Since $g(\cdot)$ is strictly decreasing, we obtain the crude upper bound

$$\bar{e}_h(\tau_{\mathrm{A}})^{1/\varepsilon} \leq \Delta_X \lambda_I\, g(\sigma^{-2} + \tau_{\mathrm{A}}) \leq \Delta_X \lambda_I\, g(\tau_{\mathrm{A}}). \tag{B-19}$$

Recalling the Gaussian formula $g(\tau) = \varphi(\sqrt{\tau})/\sqrt{\tau} = g(\tau) = \frac{1}{\sqrt{2\pi}}\frac{e^{-\tau/2}}{\sqrt{\tau}}$, we can rewrite (B-19) as

$$\bar{e}_h(\tau_{\mathrm{A}}) \leq (\Delta_X \lambda_I (2\pi)^{-1/2})^{\varepsilon} \cdot \tau_{\mathrm{A}}^{-\frac{\varepsilon}{2}} \cdot \exp\left(-\frac{\tau_{\mathrm{A}}\varepsilon}{2}\right). \tag{B-20}$$

Multiply the bound (B-20) by $I(\tau_{\mathrm{A}}) = I_0 + \exp(\eta\tau_{\mathrm{A}})$:

$$\begin{aligned}
\bar{E}_h(\tau_{\mathrm{A}}) &= I(\tau_{\mathrm{A}})\bar{e}_h(\tau_{\mathrm{A}}) \\
&\leq \left(\Delta_X \lambda_I (2\pi)^{-1/2}\right)^{\varepsilon} \cdot (I_0 + e^{\eta\tau_{\mathrm{A}}}) \cdot \tau_{\mathrm{A}}^{-\frac{\varepsilon}{2}} \cdot \exp\left(-\frac{\tau_{\mathrm{A}}\varepsilon}{2}\right) \\
&= C \cdot \tau_{\mathrm{A}}^{-\frac{\varepsilon}{2}} \left(I_0\, e^{-\frac{\tau_{\mathrm{A}}\varepsilon}{2}} + \exp\left((\eta - \tfrac{\varepsilon}{2})\tau_{\mathrm{A}}\right)\right),
\end{aligned} \tag{B-21}$$

for a constant $C > 0$ independent of τ_{A}.

Recall that

$$\eta_{\mathrm{crit}} := \frac{\varepsilon}{2}.$$

Thus if $\eta < \eta_{\text{crit}}$, then the exponent $(\eta - \eta_{\text{crit}})$ in (B-21) is negative, so the bracketed term converges to 0 as $\tau_{\text{A}} \to \infty$; the polynomial factor $\tau_{\text{A}}^{-\varepsilon/2}$ also converges to 0. Hence the upper bound (B-21) implies $\bar{E}_h(\tau_{\text{A}}) \to 0$, establishing (B-18).

Thus we have shown that $\bar{X}_h(\tau_{\text{A}}) \to 0$ as $\tau_{\text{A}} \to \infty$, which implies $\lim_{\tau_{\text{A}} \to \infty} \bar{U}^+(\tau_{\text{A}}) = 0$. Consequently, the supremum of $\bar{U}^+(\tau_{\text{A}})$ over $\tau_{\text{A}} \geq 0$ cannot be attained "at infinity": $\arg\max_{\tau_{\text{A}} \geq 0} \bar{U}^+(\tau_{\text{A}}) < +\infty$.

B.9 Proof of Proposition 15

Let $\mathcal{X} := [0, \Sigma^{-2}]$. For any $X \geq 0$, $0 < F_{\text{syn}}(X) < \Sigma^{-2}$. Thus $F_{\text{syn}} : \mathcal{X} \to \mathcal{X}$. Moreover, F_{syn} is continuous and strictly increasing in X. Therefore, by Tarski's fixed point theorem, the set of fixed points of F_{syn} is nonempty and admits a least and a greatest element; denote them by $\bar{X}_\ell^\star$ and $\bar{X}_h^\star$. Denote the corresponding steady-state efforts by

$$\bar{e}_k^\star := e(\bar{X}_k^\star, \tau_{\text{A}}), \qquad k \in \{\ell, h\}.$$

Because $\tau_{\text{syn}} > 0$ and $e(0, \tau_{\text{A}}) = 0$,

$$F_{\text{syn}}(0) = \left[\Sigma^2 + \tau_{\text{syn}}^{-1}\right]^{-1} > 0.$$

Since $\bar{X}_\ell^\star$ is a fixed point and is the *least* fixed point in $\mathcal{X}$, we must have $\bar{X}_\ell^\star \geq F_{\text{syn}}(0) > 0$, hence $\bar{X}_\ell^\star > 0$, which also implies $\bar{e}_\ell^\star = e(\bar{X}_\ell^\star, \tau_{\text{A}}) > 0$. This proves part (1).

We first note pointwise monotonicity of F_{syn} in parameters. In particular,

- $F_{\text{syn}}(X)$ is strictly increasing in I for all $X > 0$.
- $F_{\text{syn}}(X)$ is strictly increasing in τ_{syn} for all $X \geq 0$.
- $F_{\text{syn}}(X)$ is strictly decreasing in τ_{A} for all $X > 0$.

We prove the statements for the least fixed point $\bar{X}_\ell$; the argument for the greatest fixed point is analogous.

Aggregation capacity I. Let $I_2 > I_1$, and write $F_j(\cdot) := F_{\text{syn}}(\cdot; I_j, \tau_{\text{A}}, \tau_{\text{syn}})$ and $\bar{X}_{\ell,j}^\star$ for the least fixed point of F_j. By monotonicity in I, $F_2(X) \geq F_1(X)$ for all $X \in \mathcal{X}$, and $F_2(X) > F_1(X)$ for all $X > 0$. Standard monotone-comparative-statics for extremal fixed points implies $\bar{X}_{\ell,2}^\star \geq \bar{X}_{\ell,1}^\star$. To see the inequality is strict, suppose toward a contradiction that $\bar{X}_{\ell,2}^\star = \bar{X}_{\ell,1}^\star =: x$. By part (1), $x > 0$, and thus Step 3 gives $F_2(x) > F_1(x) = x$, contradicting that x is a fixed point of F_2. Hence $\bar{X}_\ell^\star$ is strictly increasing in I.

The monotone comparative statics for synthetic precision τ_{syn} and agentic precision τ_{A} is analogous.

For $k \in \{\ell, h\}$, $\bar{e}_k^\star = e(\bar{X}_k^\star, \tau_{\mathrm{A}})$. Since $e(\cdot, \tau_{\mathrm{A}})$ is strictly increasing in X, the strict monotonicity of $\bar{X}_k^\star$ in I and τ_{syn} implies $\bar{e}_k^\star$ is strictly increasing in I and τ_{syn}.

For τ_{A}, let $\tau_{\mathrm{A}}^2 > \tau_{\mathrm{A}}^1$. Then $\bar{X}_k^\star(\tau_{\mathrm{A}}^2) < \bar{X}_k^\star(\tau_{\mathrm{A}}^1)$ and $e(X, \tau_{\mathrm{A}})$ is strictly decreasing in τ_{A}, so

$$\bar{e}_k^\star(\tau_{\mathrm{A}}^2) = e(\bar{X}_k^\star(\tau_{\mathrm{A}}^2), \tau_{\mathrm{A}}^2) < e(\bar{X}_k^\star(\tau_{\mathrm{A}}^1), \tau_{\mathrm{A}}^2) < e(\bar{X}_k^\star(\tau_{\mathrm{A}}^1), \tau_{\mathrm{A}}^1) = \bar{e}_k^\star(\tau_{\mathrm{A}}^1),$$

establishing that $\bar{e}_k^\star$ is strictly decreasing in τ_{A}.

B.10 Proof of Proposition 16

Under the modified public-learning technology (14), a symmetric equilibrium still has $e_{i,t} = e_t \ \forall i$ and the individual best response $e_t = e(X_t, \tau_{\mathrm{A}})$ is unchanged relative to the baseline model because β only affects how effort maps into the next cohort's public signal.

With $\beta > 0$, the island-level public signal precision generated by cohort t equals $\lambda_G E_t^{(\beta)} = \lambda_G I e_t \beta$. Therefore, the public precision evolves according to

$$X_{t+1}^{-1} = (X_t + \lambda_G I e_t \beta)^{-1} + \Sigma^2, \qquad e_t = e(X_t, \tau_{\mathrm{A}}). \tag{B-22}$$

Define the induced one-step map

$$F\beta(X) := \left[\Sigma^2 + (X + \lambda_G I e(X, \tau_{\mathrm{A}})\beta)^{-1}\right]^{-1}.$$

Then the knowledge-collapse steady state $X = 0$ remains a fixed point of $F\beta$ when $\beta > 0$ because $e(0, \tau_{\mathrm{A}}) = 0$ implies $F\beta(0) = 0$.

The condition steady-state effort needs to satisfy is changed to

$$D\beta(e; \tau_{\mathrm{A}}, I) := \Delta_{\mathrm{X}} \cdot G_W(Ie\beta) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} e + \tau_{\mathrm{A}}) - e^{1/\varepsilon}.$$

Using Lemma A-3, $G_W(E) = \kappa_1 E^{1/4} + o(E^{1/4})$ as $E \downarrow 0$, hence $G_W(Ie\beta) = \tilde{\kappa} e^{\beta/4} + o(e^{\beta/4})$ as $e \downarrow 0$ (for $\tilde{\kappa} > 0$). Therefore, as $e \downarrow 0$,

$$D\beta(e; \tau_{\mathrm{A}}, I) = C(\tau_{\mathrm{A}}, I) e^{\beta/4} - e^{1/\varepsilon} + o(e^{\beta/4}),$$

and the same local-stability argument as in Lemma 2 goes through with $1/4$ replaced by $\beta/4$. The core arguments used to characterizes steady states remain unchanged with the stability condition replaced. Therefore, Proposition 3 and 5 hold with condition $\varepsilon < 4$ (or $\varepsilon > 4$) replaced by $\varepsilon < \frac{4}{\beta}$ (or $\varepsilon > \frac{4}{\beta}$). The comparative statics hold similarly.

We next consider the asymptotic for $\tau_{\mathrm{A}}^c(I)$. For this threshold to exist, we assume $\varepsilon > \frac{4}{\beta}$. Analogous to the proof of Proposition 12, we also have the the tangency conditions hold at

$\tau_{\mathrm{A}} = \tau_{\mathrm{A}}^c(I)$, which are now in the form:

$$0 = \Delta_{\mathrm{X}} \cdot G_W(\bar{E}_I) \cdot \lambda_{\mathrm{I}} g(\sigma^{-2} + \lambda_{\mathrm{I}} \bar{e}_I + \tau_{\mathrm{A}}^c(I)) - \bar{e}_I^{1/\varepsilon}, \tag{B-23}$$

$$0 = \lambda_{\mathrm{I}}(1 + 1/\bar{Y}_I)\bar{e}_I + 2\big(1/\varepsilon - \beta R(\bar{E}_I)\big), \tag{B-24}$$

where

$$\bar{E}_I := I\bar{e}_I^{\beta}, \qquad \bar{Y}_I := \sigma^{-2} + \lambda_{\mathrm{I}}\bar{e}_I + \tau_{\mathrm{A}}^c(I), \qquad R(E) := \frac{EG_W'(E)}{G_W(E)}.$$

Since $\varepsilon > \frac{4}{\beta}$, due to Lemma A-3, there is a unique $E_* > 0$ such that $R(E_*) = \frac{1}{\varepsilon\beta}$.

Similar to the original proof, we can show that $\bar{e}_I \to 0$ and $\bar{E}_I \to E_*$ as $I \to \infty$. Next, from (B-23) we have

$$g(\bar{Y}_I) = \frac{\bar{e}_I^{1/\varepsilon}}{\Delta_{\mathrm{X}}\lambda_{\mathrm{I}} G_W(\bar{E}_I)} = \frac{\bar{E}_I^{1/\varepsilon\beta}}{\Delta_{\mathrm{X}}\lambda_{\mathrm{I}} G_W(\bar{E}_I)}\, I^{-1/\varepsilon\beta}.$$

Since $\bar{E}_I \to E_*$ and $G_W(E_*) > 0$, the prefactor converges to a constant $C_0 > 0$, so

$$g(\bar{Y}_I) \sim C_0\, I^{-1/\varepsilon\beta} \qquad (I \to \infty).$$

This then implies $\lim_{I\to\infty} \frac{\tau_{\mathrm{A}}^c(I)}{\log I} = \frac{2}{\varepsilon\beta}$ following the same procedure in the proof of Proposition 12.

Next, we show the asymptotic of $\tau_{\mathrm{A}}^\star(I)$. Define $e_I^\star$ and $Y_I^\star$ similarly as the effort and idiosyncratic precision at the high steady state corresponding to I and $\tau_{\mathrm{A}} = \tau_{\mathrm{A}}^\star(I)$

$$e_I^\star := \bar{e}_h(\tau_{\mathrm{A}}^\star(I), I), \qquad Y_I^\star := \sigma^{-2} + \lambda_{\mathrm{I}} e_I^\star + \tau_{\mathrm{A}}^\star(I),$$

and set $E_I^\star := I(e_I^\star)^\beta$.

At any positive steady state we have the effort condition

$$(e_I^\star)^{1/\varepsilon} = \Delta_{\mathrm{X}}\lambda_{\mathrm{I}} G_W(E_I^\star)\, g(Y_I^\star). \tag{B-25}$$

In the β-economy, the ratio $\mathrm{IE} \,/\, \mathrm{DE} = 1$ becomes

$$1 = \frac{\mathrm{IE}}{\mathrm{DE}} = \frac{(1 + 1/Y_I^\star)\cdot \beta R(E_I^\star)}{\lambda_{\mathrm{I}}(1 + 1/Y_I^\star)\cdot e_I^\star + 2(1/\varepsilon - \beta R(E_I^\star))} \cdot \frac{\Delta_{\mathrm{G}} + G(Y_I^\star)\Delta_{\mathrm{X}}}{g(Y_I^\star)\Delta_{\mathrm{X}}}.$$

Solving for $g(Y_I^\star)$ from (B-25) and substituting it yields

$$(e_I^\star)^{1/\varepsilon}\Big[\lambda_{\mathrm{I}}(1 + 1/Y_I^\star)\cdot e_I^\star + 2(1/\varepsilon - \beta R(E_I^\star))\Big] = \lambda_{\mathrm{I}} G_W(E_I^\star)(1 + 1/Y_I^\star)\cdot \beta R(E_I^\star)\cdot \big(\Delta_{\mathrm{G}} + G(Y_I^\star)\Delta_{\mathrm{X}}\big). \tag{B-26}$$

Similar to the proof of Proposition 12, we can show that as $I \to \infty$,

$$E_I^\star \to \infty, \qquad e_I^\star \to 0.$$

Plugging these asymptotics into (B-26) yields

$$\lim_{I\to\infty} \frac{(e_I^\star)^{1/\varepsilon}}{R(E_I^\star)} = \frac{\lambda_{\mathrm{I}} G_\infty \cdot \varepsilon\beta(\Delta_{\mathrm{G}} + \Delta_{\mathrm{X}})}{2}. \tag{B-27}$$

Using $R(E) = c_R/E + O(E^{-2})$ and $E_I^\star = I(e_I^\star)\beta$, (B-27) becomes

$$I\,(e_I^\star)^{1/\varepsilon+\beta} \sim \frac{\lambda_{\mathrm{I}} G_\infty \cdot \varepsilon\beta(\Delta_{\mathrm{G}} + \Delta_{\mathrm{X}})c_R}{2},$$

therefore $e_I^\star \sim C_E \cdot I^{-1/(1/\varepsilon+\beta)}$ for some constant $C_E > 0$.

From the steady-state condition (B-25) and $G_W(E_I^\star) \to G_\infty$ we obtain

$$g(Y_I^\star) = \frac{(e_I^\star)^{1/\varepsilon}}{\Delta_{\mathrm{X}}\lambda_{\mathrm{I}} G_W(E_I^\star)} \sim C_g\, I^{-1/(\varepsilon\beta+1)},$$

which implies

$$\lim_{I\to\infty} \frac{\tau_{\mathrm{A}}^\star(I)}{\log I} = \frac{2}{\varepsilon\beta + 1}.$$