Cognitive Delegation in Childhood: A Theory of Skill Formation under Generative AI*

Siqi Wei†

David Nazarian College of Business and Economics
California State University Northridge

Naying Zhou‡

Keck School of Medicine
University of Southern California

May, 2026

Abstract

This paper develops a dynamic theory of childhood human-capital formation when a generative technology permits the production of educational output without the underlying cognitive process. We call such uses cognitive delegation and distinguish process-preserving use (active engagement above the capacity-formation threshold) from process-replacing use. The model is a discrete-time skill-formation framework in which a present-biased child accumulates knowledge, discipline, and judgment through effort, delegation, and verification, with age-dependent plasticity and a governance function setting the threshold. Solving the recursive problem under sophisticated quasi-hyperbolic preferences, we establish the Capacity-Wedge Decomposition: the long-run capacity wedge under process-replacing use admits an additive decomposition into a contemporaneous discipline-channel cost and a cascade cost from downstream knowledge complementarities. The marginal welfare cost is strictly decreasing in age, and under sufficient conditions admits a sharp threshold in governance, strictly increasing in AI capability. The framework identifies governance, not access, as the operative variable in the developmental effect of generative AI.

Keywords: Human capital; Skill formation; Cognitive delegation; Active engagement; Generative AI.
JEL codes: I20, I24, I28, J24, D83, D91, O33.

*The usual disclaimers apply; any remaining errors are the authors' responsibility.

†David Nazarian College of Business and Economics at California State University Northridge. 18111 Nordhoff St, Northridge CA 91330. Email: siqi.wei@csun.edu. Corresponding author.

‡Keck School of Medicine at University of Southern California. 1441 Eastlake Ave, Los Angeles, CA 90089, USA. Email: nayingzh@usc.edu.

1. Introduction

For the better part of two centuries, the institution of childhood schooling has operated on a technological assumption about how learning works. The assumption was procedural rather than philosophical. To complete a problem set, a child had to work through the problems. To produce an essay, a child had to read the passage, draft, and revise. To answer a question about a book, a child had to first read the book. The cognitive process and the observable output were joined by a feature of the available technology: the only path from a blank page to a completed assignment ran through the cognitive activity the assignment was designed to elicit. Generations of educators built curricula, assignments, and grading practices around this technological reality. Generations of children developed the capacities those practices were designed to cultivate: sustained attention, the willingness to struggle with difficulty, the discipline to revise, the habit of independent reasoning, the judgment to evaluate one's own work. The dynamic accumulation of such capacities through deliberate investment is the central object of human-capital theory since Becker (1964), Mincer (1974), and Ben-Porath (1967), with the modern formulation of dynamic complementarity and self-productivity due to Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010).

The arrangement worked because the technology of cognitive output enforced the underlying cognitive process. A teacher who assigned twenty math problems was not just measuring whether the child could produce twenty correct answers; she was conscripting the child into twenty acts of active mathematical thinking. A teacher who assigned a five-paragraph essay was not just measuring whether the child could produce five paragraphs of coherent prose; she was requiring the child to engage in five paragraphs of drafting, revising, and reading. The grade attached to the output served as the institution's incentive instrument, but the instrument operated on the process because the underlying technology made the process and the output coextensive. The child who appeared at the end of the year with the capacity to think mathematically, to compose, to read with attention, had become that child precisely because the technology of schooling required these activities of her.

Generative artificial intelligence introduces a new technological possibility. The economic implications of AI as a general-purpose technology have been studied at the production-side level in Brynjolfsson and McAfee (2014), Acemoglu and Restrepo (2018, 2020), Agrawal, Gans and Goldfarb (2018), and Acemoglu (2024); the present paper focuses on the implications of generative AI specifically for childhood human-capital formation, a margin the existing AI-economics literature has not yet engaged. The child can now obtain the same output, the math answers, the essay, the analysis of the book, without performing the underlying cognitive work. At the moment of submission, the output is observationally unchanged. The grading practice continues to operate. The institutional reward is paid. What has changed is the relationship between the output and the cognitive experience that produced it. For one child, the technology operates as a tutor that supports her own active engagement with the task: she attempts the problem before asking for an explanation, she drafts before requesting feedback, she reads before asking for a summary. For another child, the technology operates as a substitute for the active engagement: the AI generates the answers, the essay, the analysis, and the child receives and submits the result. The contemporaneous output looks the same in both cases. The cognitive experience that has occurred in the child does not.

Recent experimental evidence in Bastani et al. (2024) and Kestin et al. (2024) is consistent with this distinction: students who use AI on homework score higher on graded assignments but lower on subsequent unaided exams, and AI tutoring with process-preserving structure can outperform passive instruction while unstructured AI use does not.

This paper develops a dynamic theory of what each cognitive mode does to the formation of childhood capacity. The paper is positive theory. It does not begin from the position that the technology is harmful or beneficial. It begins from the observation that the same technology can be used in two qualitatively different modes, and asks when each mode obtains and what each mode does to the formation of long-run human capital. The substantive concern that has motivated the project, that something has changed in how children learn, receives an analytical answer rather than a verdict.

1.1. Motivation

The stakes of the question extend beyond the surface concerns of the current public debate. Three substantive features of the problem motivate the analytical apparatus we develop. None is sufficient on its own; together they identify a structural shift that the existing economics of education has not yet engaged.

Consider first what childhood schooling does. The framing that dominates current commentary, focused on whether students cheat, whether teachers can detect AI-assisted work, and whether institutions should ban the technology, treats schooling as a procedure for producing output. But the durable economic content of childhood schooling is not the output. It is the repeated cognitive labor through which attention, persistence, the willingness to revise, the habit of independent reasoning, and the judgment to evaluate one's own work are formed. These capacities are inputs into the production function for human capital, not outputs of the schooling procedure. The skill-formation literature of Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010) formalizes this insight as the joint determination of cognitive and non-cognitive capacity through dynamically complementary investments. A technology that allows children to obtain the visible output while bypassing the cognitive process therefore alters the production function itself rather than the procedure of homework completion. The cheating frame examines a symptom; the production-function frame examines the structural change.

The second feature is that capacities formed in childhood are durable life-course inputs, not contemporaneous test inputs. The developmental and life-course literatures, in particular the body of work synthesized by Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010) and the long-horizon follow-up studies in Heckman, Pinto and Savelyev (2013), Kautz et al. (2014), and Heckman and Mosso (2014), have established that non-cognitive capacities measured in childhood predict adult outcomes across a wide range: post-secondary completion, labor-market sorting and earnings, health-relevant decisions, financial decision-making, the capacity to filter and evaluate information, and civic participation. The capacities the present model formalizes as discipline and judgment correspond closely to the capacities the life-course literature has identified as durable inputs into adult welfare. A wedge in capacity formation during childhood is therefore a wedge in life trajectories, not a wedge in school transcripts. The relevant time horizon for evaluating childhood AI policy is decades, and the relevant outcome space includes labor-market earnings, health, and civic competence in addition to measured test scores.

The third feature concerns the distributional implications. The standard equity framing of educational technology is access-based: who has the technology, who can afford its subscription, who attends a school that has integrated it. The framework developed here says that access is the first margin but not the most important. Whether the child's use of the technology preserves the cognitive process through which capacity is formed, or replaces that process while preserving the appearance of completed work, depends on a separate household and institutional variable, which we label governance. Governance varies sharply across households and schools. It is imperfectly correlated with income or parental education and depends on time, attention, and design features that are not transferred through market access. Universal access to generative AI may therefore fail to operate as an equalizing intervention. In the absence of complementary governance investment, it may instead convert unequal governance environments into unequal developmental trajectories. Recasting the equity question from access to governance is, in our view, the most consequential single contribution that the framework can make to the public debate.

These three features motivate the three structural objects the paper develops: the active-passive distinction as a primitive of skill formation; the dynamic cascade structure that generates the lifetime capacity wedge; and the governance threshold that mediates its distributional consequences.

1.2. The research question

The paper organizes itself around one question, expanded by a single related clause to acknowledge that the substantive concern requires more than a qualitative answer:

> *Under what conditions does cognitive delegation in childhood produce the same long-run human-capital outcomes as active cognitive engagement, and when the two modes diverge, what is the structure of the divergence: its magnitude, its age profile, and its dependence on household and institutional environments?*

The paper does not aim to pronounce a verdict on the technology. It aims to specify the conditions under which the two modes diverge, the structure of the divergence when it occurs, and the institutional features that select between them. Neither the position that generative AI is a transformative learning aid nor the position that it is corroding childhood learning is right as a general claim; both are right under conditions the present framework makes explicit.[1]

[1] The cohort that entered kindergarten in the autumn after the public release of large language models will reach the labor market in approximately 2040. The institutional features that determine whether their AI use is process-preserving or process-replacing are being set now.

1.3. Approach

The paper makes one conceptual move from which all of its analytical content follows. The move is to take seriously, as a primitive of skill-formation theory, the distinction between active cognitive engagement and passive reception of cognitive output. Active engagement is the child constructing, generating, working through, revising, retrieving, checking. Passive reception is the child reading, accepting, submitting cognitive work performed by an external system. The two modes can produce identical observable output. The cognitive-science literature on the generation effect (Slamecka and Graf, 1978), retrieval practice (Roediger and Karpicke, 2006), desirable difficulties (Bjork, 1994), deliberate practice (Ericsson, Krampe and Tesch-Römer, 1993), and productive struggle (Hiebert and Grouws, 2007) has established for decades that the two modes do not produce identical formation of underlying capacity. The paper imports this finding into dynamic skill-formation theory and develops the structural consequences.

The standard dynamic human-capital framework, in the tradition of Becker (1964), Ben-Porath (1967), and the formalization by Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010), treats investment in human capital as a homogeneous flow. A unit of investment is a unit of investment; composition and intensity govern the production of capacity. The paper relaxes this assumption. It treats the cognitive mode of investment as an additional dimension and shows that, when the skill-formation technology exhibits downstream complementarities across the curricular sequence, the cognitive mode determines whether nominal investment accumulates into long-run capacity. The implication is that the technology that has appeared in childhood since the public release of large language models is not a marginal change in the price or availability of educational input. It is a change in the dimension along which inputs vary, and the dynamic theory of skill formation requires extension to accommodate it.

The paper provides that extension. The extension consists of one main theorem, the Capacity-Wedge Decomposition, together with two corollaries that develop its age and governance implications and a small set of supporting results collected in the appendix. All derive from the same primitive: the active-passive distinction interacting with downstream skill complementarities under endogenous capacity formation.

1.4. The main result

The paper's main result is the Capacity-Wedge Decomposition. Under process-replacing delegation, the long-run wedge between active-engagement capacity and realized capacity decomposes into two components: a contemporaneous discipline-channel cost and a cascade cost from downstream knowledge complementarities. Two immediate implications follow. First, the marginal welfare cost is larger earlier in childhood because plasticity is higher and the remaining curricular distance is longer. Second, under sufficient conditions on the household and institutional environment, the welfare effect of AI access varies with governance capacity: access can be welfare-improving in high-governance environments and welfare-reducing in low-governance environments. The substantive content of the decomposition is in Sections 7 and 10; the two corollaries follow there as immediate consequences.

1.5. Contribution and relation to the literature

The paper sits at the intersection of three literatures, each of which contributes machinery and none of which contains the result. From the dynamic skill-formation tradition of Becker (1964), Ben-Porath (1967), Heckman (1976), Cunha and Heckman (2007), Cunha, Heckman and Schennach (2010), and Heckman and Mosso (2014), we take the discipline of treating human capital as a state shaped by deliberate investment and the formal apparatus of dynamic complementarity. From the cognitive-science literature on active engagement, including Slamecka and Graf (1978), Bjork (1994), Roediger and Karpicke (2006), Ericsson, Krampe and Tesch-Römer (1993), and Hiebert and Grouws (2007), we take the substantive distinction between active construction and passive reception as inputs into the formation of durable capacity. From the behavioral-economics literature on present bias and self-control, including Strotz (1956), Laibson (1997), O'Donoghue and Rabin (1999), and Bénabou and Tirole (2004), we take the apparatus of quasi-hyperbolic preferences and sophisticated equilibrium.

The contribution is the integration of these three components through a single technological primitive, the active-passive cognitive-mode distinction, and the demonstration that the integrated framework yields a structural characterization of the developmental consequences of generative AI that none of the three parent literatures contains. The cognitive-mode distinction does not appear in dynamic skill formation; the dynamic complementarity does not appear in cognitive science; the technology that motivates both does not appear in either as a formal primitive. The paper places them in one framework and develops what follows.

We do not claim novelty in the individual machinery. Dynamic complementarity in skill formation is well developed. Quasi-hyperbolic discounting is standard. The cognitive-science findings on active engagement have decades of empirical support. What is new is the unification through a primitive that recent technological change has made empirically central, and the characterization of consequences that follow from this unification.

1.6. Scope of the analysis

It is useful at the outset to be explicit about what the paper does and does not aim to do. The paper develops a theory of conditions under which cognitive delegation in childhood produces long-run capacity equivalent to active engagement, and what determines the structure of the divergence when the two diverge. The paper does not claim that generative AI is, on net, harmful for childhood learning. Under process-preserving use, the technology is unambiguously beneficial in the model. The paper does not claim that all observed AI use by children is process-replacing. The model distinguishes the two modes through the function $\varphi(e; g)$ and treats both as feasible and observable in equilibrium. The paper does not predict universal decline in measured capacity; it predicts heterogeneous outcomes conditional on the household and institutional environment that mediates use.

The paper also abstracts from several features that are likely empirically important. Adult governance is treated as exogenous. The school is treated as a single actor making policy. Output is produced through a single technology aggregating active effort and AI use; we do not distinguish among different forms of AI use (tutoring, generation, summarization, calculation). These abstractions are made for tractability and are common in the formal literature. We return to each in the discussion section.

1.7. Plan of the paper

Section 2 reviews the three literatures the paper draws on and identifies the specific machinery imported from each. Section 3 develops the environment and primitives of the model: time, agents, three capital stocks, three actions, age-dependent agency and present bias, the governance function. Section 4 formalizes the active-passive distinction as a primitive of the skill-formation technology, with explicit connection to the cognitive-science literature. Section 5 introduces downstream complementarity in skill formation and introduces a single baseline cascade specification used in the main analysis, with alternative specifications deferred to the appendix. Section 6 develops the child's dynamic optimization problem under quasi-hyperbolic preferences and proves existence and monotonicity properties of the equilibrium. Section 7 states and proves the Capacity-Wedge Decomposition, together with the age corollary and the governance proposition. Section 8 develops two corollaries and a small set of supporting extensions; further extensions are collected in the appendix. Section 9 analyzes limiting cases and functional-form robustness. Section 10 develops welfare theorems characterizing the planner's problem under explicit sufficient conditions. Section 11 discusses the broader reach of the framework. Section 12 concludes. Proofs deferred from the main text are collected in the Appendix.

2. Related Literature

The paper draws on three literatures in roughly equal measure and contributes to each. We discuss them in turn, emphasizing the specific machinery imported and the substantive distinction between the present treatment and prior work.

2.1. Dynamic skill formation and human-capital accumulation

The discipline of treating learning as the accumulation of a state variable under deliberate investment originates with Becker (1964) and Mincer (1974), with the canonical dynamic formulation due to Ben-Porath (1967). Heckman (1976) extends the framework with risk and learning-by-doing. The modern formulation of *dynamic complementarity* and *self-productivity* is due to Cunha and Heckman (2007) and Cunha, Heckman and Schennach (2010): investments at earlier ages raise the productivity of investments at later ages (dynamic complementarity), and skills accumulated at earlier ages raise the productivity of skill accumulation at later ages (self-productivity). The econometric estimation in Cunha, Heckman and Schennach (2010) documents that early childhood is a sensitive period for non-cognitive skill formation specifically, with downstream productivity that does not diminish over the life cycle.

The paper imports the structural machinery of dynamic complementarity and uses it to motivate the downstream-skill-complementarity assumption that drives the cascade result. The substantive distinction is that the standard skill-formation literature treats investment as a homogeneous flow: a unit of investment at age t is a unit of investment, and the dynamic complementarity operates on the accumulated stock that results from this flow. The present paper distinguishes investment by cognitive mode: the same nominal investment can be undertaken actively or passively, with different effects on the formation of capacity. The cascade theorem then says that the mode distinction interacts with the downstream complementarity to produce structural consequences that the mode-blind framework cannot capture. The present paper is therefore a methodological extension of the skill-formation framework rather than a competitor to it; it preserves all of the framework's existing structure and adds one dimension along which inputs vary.

Heckman, Pinto and Savelyev (2013), Heckman and Mosso (2014), and Kautz et al. (2014) develop the empirical implications of dynamic complementarity for non-cognitive skill formation.[2]

2.2. Cognitive science of active engagement and passive reception

The substantive distinction between active cognitive construction and passive reception of cognitive output has decades of empirical support in the cognitive-science literature and is treated as established in cognitive psychology textbooks. We summarize the five literatures most directly relevant to the paper's primitive.

The *generation effect* (Slamecka and Graf, 1978; Bertsch et al., 2007) establishes that material generated by the learner is recalled and transferred substantially better than the same material received passively. The effect has been replicated across hundreds of studies, holds across paradigms (word recall, problem solving, conceptual learning), and is robust to controls for time-on-task and exposure intensity.

The *testing effect*, or retrieval practice, (Roediger and Karpicke, 2006; Karpicke and Roediger, 2008), establishes that active retrieval of information from memory produces stronger long-run retention than passive re-reading, even when the re-reading is matched in clock-time and apparent effort. The mechanism is that retrieval itself is a learning event, not merely a measurement of prior learning.

Desirable difficulties (Bjork, 1994; Bjork and Bjork, 2011) establishes that learning conditions that feel harder in the moment (spaced practice, interleaved practice, productive struggle, retrieval before instruction) produce better long-run outcomes than smoother, easier learning experiences. The substantive content of the term is that the cognitive difficulty of an experience and its long-run productivity are positively related, in contrast to the lay intuition that easier learning is better learning.

Deliberate practice (Ericsson, Krampe and Tesch-Römer, 1993; Ericsson, 2008) establishes that expert performance is built through specific, effortful engagement with one's current limits, and is not produced by passive exposure or by undirected practice. The framework

[2]The Heckman program documents that early-childhood interventions targeting non-cognitive capacities produce returns that compound across the life cycle. The structural mechanism in this literature, that early investments enable later investments through their effect on persistence, attention, and discipline, is precisely the mechanism the present paper formalizes as the cascade in the process-replacing regime. The empirical findings of the Heckman program provide the substantive grounding for the cascade specification developed in Section 5.

has been applied to domains from music to chess to medicine and consistently identifies the active-engagement requirement as central.

Productive struggle in mathematics education (Hiebert and Grouws, 2007; Kapur, 2014) establishes that students who work through difficult problems with appropriate support before receiving instruction develop deeper conceptual understanding than students who receive instruction first and then practice. The effect is consistent across studies and across mathematical domains.

The five literatures converge on a single empirical claim: active cognitive engagement and passive reception of cognitive output are not interchangeable inputs into the formation of durable learning. The active mode produces durable capacity; the passive mode produces transient familiarity. The paper takes this empirical convergence as the substantive basis for the formal primitive developed in Section 4.

A subsidiary observation from the cognitive-science literature is that the active-passive distinction matters more for some cognitive outputs than others. Foundational skill mastery, mathematical fluency, reading decoding, retention of factual material, requires active engagement at the time of acquisition; partial substitutes are largely ineffective. Conceptual understanding and creative work admit broader supportive assistance; some passive reception of expert examples appears to be complementary to subsequent active practice rather than process-replacing. The paper's formalization treats the strict-substitution case as the baseline, with the partial-complementarity case appearing as a parametric extension.

2.3. Behavioral economics of present bias and self-control

The recognition that agents may not act in their own long-run interest, with the corresponding distinction between short-run and long-run preferences, originates with Strotz (1956). The formal apparatus of quasi-hyperbolic discounting is due to Laibson (1997); O'Donoghue and Rabin (1999, 2001) develop the sophisticated-versus-naive distinction; Bénabou and Tirole (2002, 2004) model self-control as a problem of personal rules and identity. The application of the framework to children specifically requires modification because children's intertemporal capacity is itself developing; we adopt an age-dependent present-bias parameter following developmental work on the maturation of intertemporal choice.

The paper imports the quasi-hyperbolic apparatus to capture two structural features of the childhood-learning environment. First, the wedge between the child's short-run evaluation of effort and the long-run welfare implications of effort is the formal reason schooling exists as an institution: a child who fully internalized the long-run welfare consequences of her cognitive choices would face a substantially weaker version of the problem the paper studies. Second, the wedge is what makes the process-replacing mode privately attractive even when its long-run consequences are negative: the contemporaneous gain from delegation is salient to the child, while the cascade cost is distant and discounted.

The paper does not claim that present bias is necessary for the main results. With $\beta = 1$, the Capacity-Wedge Decomposition retains the decomposition under the substantive interpretation that the planner's evaluation differs from the child's evaluation through the long-run-welfare channel, but the policy implications of Proposition 3 require additional structure to justify paternalistic intervention. With $\beta < 1$, the policy content of the theorem is sharper and the welfare analysis of Section 10 admits a clean Pareto-improvement characterization.

2.4. Economics of AI in education

The economics of artificial intelligence has developed rapidly. Brynjolfsson and McAfee (2014) and Acemoglu and Restrepo (2018, 2020) provide influential analyses of automation and labor markets. Agrawal, Gans and Goldfarb (2018) formalize AI as a prediction technology that reduces the marginal cost of forecasts. Acemoglu (2024) considers the macroeconomic implications of large language models. The empirical literature on AI in workplaces, Brynjolfsson, Li and Raymond (2023), Noy and Zhang (2023), Peng et al. (2023), establishes that capable AI raises measured productivity in many settings while leaving open the question of how the productivity gains interact with skill formation.

The empirical literature on AI in education is younger and less settled. Recent experimental evidence is consistent with the wedge our framework formalizes: Bastani et al. (2024) find that students who use AI on homework score higher on graded assignments but lower on subsequent unaided exams, and Kestin et al. (2024) report that process-preserving AI tutoring can outperform passive instruction while unstructured AI use does not.[3] Across the empirical literature in education and learning sciences, the emerging finding is that AI's effect on learning depends on the institutional and pedagogical structure surrounding its use, a finding for which the present paper supplies the theoretical structure.

The paper's contribution to the AI economics literature is to identify cognitive delegation as a distinct class of technological change with structural implications for human-capital formation that automation, prediction, or generation alone do not capture. Cognitive delegation operates on the formation of capacity, not only on the production of output. The framework developed here is intended to make the developmental dimension of generative AI legible to economists working on technology and skill formation.

3. Environment and Primitives

This section develops the formal environment of the model. The substantive content is the specification of (i) the time and agent structure, (ii) the three capital stocks the child accumulates, (iii) the three choice variables the child controls, (iv) the age-dependent primitives of plasticity, present bias, and strategic agency, and (v) the output and governance technologies. The active-passive distinction that animates the paper enters formally in Section 4, building on the primitives developed here.

[3] The empirical pattern is suggestive rather than dispositive. Identification of the long-run capacity wedge requires longitudinal outcome measurement that the post-2022 data infrastructure has only begun to make possible; we develop the testable predictions of the framework in Section 11.

3.1. Time, agents, and capital stocks

Time is discrete, $t = 0, 1, 2, \ldots, T$, with $T \in \mathbb{N}$. We interpret t as the age of the child measured in periods (a period may be a quarter, a semester, or a year, depending on application). The schooling life cycle runs from $t = 0$ (entry into formal schooling) to $t = T$ (exit from secondary education). Throughout, we hold the labor-market post-schooling life cycle implicit: terminal capital stocks s_{T+1} enter the planner's evaluation through a continuation value function $\Omega(s_{T+1})$ that we treat as a primitive of the welfare analysis.

A representative child enters period t with a state vector

$$s_t \;=\; (h_t, D_t, J_t) \;\in\; \mathcal{S} \subseteq \mathbb{R}^3_+, \tag{1}$$

where h_t is *knowledge*, D_t is *learning discipline*, and J_t is *judgment*. The economic interpretation of each is given below. We do not impose a state-space boundary except where required by specific functional forms; we assume $\mathcal{S}$ is a compact convex set with non-empty interior and contains a non-empty subset on which all three stocks are strictly positive.

The household's adult governance is a parameter $g \in [0, \bar{g}]$ that does not vary over time within a given child. The paper treats g as exogenous at the household level; the distribution $\Gamma(g)$ across households is also treated as exogenous. Endogenizing g is the subject of Extension 1 in Section 8.

The parameter g is intended to be *independently interpretable*, not merely the structural object that converts AI use from process-replacing to process-preserving. We interpret g as a composite index of four observable features of the child's institutional and household environment: (i) parental supervision time directed to schoolwork, (ii) teacher process-monitoring intensity and willingness to require visible intermediate work, (iii) the prevalence of process-based assignments in the child's school (oral defense, supervised drafting, in-class problem-solving), and (iv) household digital literacy and explicit rules governing AI use. Each of these features is measurable in principle, varies non-trivially across households and schools in observed data, and is correlated with but distinct from household income or parental education. The composite index g is the projection of these features onto the single dimension that affects the child's effective access to process-preserving versus process-replacing modes. This grounding matters because the threshold result in Section 7 depends on g being substantively meaningful: without it the threshold characterization would be a tautological restatement of the assumption that some households shift use toward process-preserving modes. With it, the threshold result becomes an empirically testable prediction about heterogeneity in observable features of the household and school environment.

Knowledge h_t. The knowledge state is the conventional human-capital object: the stock of facts, procedures, conceptual understanding, and integrated subject matter the child can deploy in cognitive tasks. It depreciates through forgetting and accumulates through investment, with the specific accumulation technology developed in Section 4. Knowledge is the state most directly observable through assessments and the state most directly affected by the technology of output.

Learning discipline D_t. The discipline state is the capacity to engage in costly cognitive work: attention, patience, willingness to struggle with difficulty, persistence in revision, tolerance for being corrected. It is not a preference parameter and it is not knowledge in the conventional sense. It is a productive capacity whose primary effect is to lower the future cost of cognitive effort. Discipline is the state most directly built by active engagement and most directly degraded by sustained substitution. We discuss the interpretation of discipline more fully in Subsection 3.7 and connect it to the cognitive-science literature on persistence and self-regulation.

Judgment J_t. The judgment state is the capacity to evaluate claims, detect errors, compare sources, and recognize uncertainty. It is built by verification practice (checking AI output, comparing one's own work to instructed standards, explaining one's reasoning) and, to a lesser extent, by the integrative experience of active cognition itself. Judgment is the state most directly relevant to the long-run social externalities of cognitive delegation, since a population's collective judgment determines its collective ability to evaluate the outputs of AI systems on which it relies.

3.2. The child's choice variables

In each period t the child chooses an action vector

$$x_t \;=\; (e_t, a_t, v_t) \;\in\; \mathcal{X} \;\equiv\; [0, \bar{e}] \times [0, \bar{a}] \times [0, \bar{v}], \tag{2}$$

where $e_t \geq 0$ is *active effort*, $a_t \geq 0$ is *cognitive delegation*, and $v_t \geq 0$ is *verification*. The action set $\mathcal{X}$ is compact and convex with $\bar{e}, \bar{a}, \bar{v}$ strictly positive. The upper bounds reflect time and attention constraints within a single period and play no substantive role in the analysis. Below, we describe each action.

Active effort e_t. Active effort is the child's own cognitive engagement with the task: reading the passage, working through the problem, drafting the essay, retrieving facts from memory, revising her own work. The economic content of effort is not time-on-task but cognitive engagement intensity; we treat the two as interchangeable in the formal model. The substantive content of e_t is that it is the mode in which the child constructs cognitive output through her own work. This will be made formal in the active-passive distinction of Section 4.

Cognitive delegation a_t. Cognitive delegation is the use of an external technology, generative AI, to produce or assist in producing cognitive output that the child submits. Delegation does not occur in a single mode: the same nominal level of a_t can be undertaken in a process-preserving mode (using AI to support the child's own active engagement) or in a process-replacing mode (using AI to replace the child's active engagement).[4] The distinction

[4] Process-preserving delegation corresponds to what education researchers often call *scaffolding*. We do not use that term as our central economics terminology because its educational valence is somewhat narrative,

between these modes is one of the central modeling moves of the paper and is formalized in Section 4.

Verification v_t. Verification is the child's effortful checking of AI output: explaining the AI's reasoning back, comparing the AI's output to sources, replicating computations, identifying errors. Verification differs from active effort in that it is performed on AI-produced content rather than on the original task, but it shares with effort the property of being an active engagement of the child's cognitive system. For tractability, we focus the main analysis on the effort-delegation pair (e, a) and treat verification as a separate channel that activates only when the model is extended to address judgment formation in Section 8.

3.3. Age-dependent primitives

The model has three age-dependent primitives that capture the developmental features of childhood. We introduce them in this subsection and use them throughout the analysis.

Plasticity π_t. Plasticity governs the rate at which active engagement and substitution build or degrade the discipline state. It is age-declining:

$$\pi_t \;=\; \pi_0 \,\exp(-\kappa_\pi\, t), \qquad \pi_0 \in (0, 1],\; \kappa_\pi > 0. \tag{3}$$

The substantive content is that the same investment in active engagement at an earlier age has a larger effect on the formation of discipline than the same investment at a later age, and symmetrically that the same process-replacing use at an earlier age has a larger negative effect.[5]

Present bias β_t. The child evaluates the discounted sum of period payoffs with quasi-hyperbolic preferences. The present-bias parameter is itself age-dependent:

$$\beta_t \;=\; \beta_\infty - (\beta_\infty - \beta_0)\,\exp(-\xi\, t), \qquad 0 < \beta_0 < \beta_\infty \le 1,\; \xi > 0. \tag{4}$$

The youngest children face the most severe present bias (β_0) and asymptote to a less-biased value (β_∞) over the schooling life cycle. The interpretation is that intertemporal capacity matures over childhood, with the rate of maturation governed by ξ.

Strategic agency ω_t. The child's realized action is not in general the action a sophisticated quasi-hyperbolic optimizer would choose. The child's strategic capacity to optimize over multi-period horizons itself develops over childhood. We model this through an agency

but the substantive content is the same.

[5]The exponential functional form is conventional; any strictly decreasing $\pi : [0, T] \to (0, 1]$ delivers the qualitative content of the age-gradient corollary. The age-decline of plasticity is documented empirically in the developmental literature (Cunha, Heckman and Schennach, 2010; Heckman and Mosso, 2014) as a feature of non-cognitive skill formation specifically.

parameter $\omega_t \in [0, 1]$ that interpolates between the sophisticated-optimizer choice and the adult-default action:

$$x_t^{\text{realized}} \; = \; \omega_t \, x_t^* + (1 - \omega_t) \, x_t^{\text{default}}(g), \tag{5}$$

where x_t^* is the solution to the child's sophisticated optimization problem and $x_t^{\text{default}}(g)$ is the action determined by adult defaults under governance g. The agency parameter is itself age-dependent and weakly increasing:

$$\omega_t \; = \; \omega_\infty \Big[1 - \exp(-\lambda_a \, t) \Big], \qquad \omega_\infty \in (0, 1], \;\; \lambda_a > 0. \tag{6}$$

At $t = 0$ the realized action equals the adult default. As $t \to \infty$ the realized action approaches the strategic-optimizer choice. The parameter λ_a governs the rate at which strategic agency develops; the asymptote ω_∞ captures the upper bound on autonomous decision-making (which need not equal one if adult oversight persists throughout schooling).

For analytical clarity, the main results in Sections 7 and 10 are stated for the limiting case $\omega_t = 1$ (full agency), with the age-dependent agency case treated as an extension in Section 8. The qualitative conclusions are robust to interior ω_t, with the welfare implications strengthening for ω_t small (younger children).

3.4. The output technology

Schoolwork output in period t is produced by a constant-elasticity-of-substitution (CES) aggregator combining the child's own cognitive contribution and the AI's contribution. The own contribution is the interaction of effort and knowledge, $h_t e_t$, reflecting the fact that effort is more productive when the child knows more. The AI's contribution is θa_t, where $\theta \geq 0$ is AI capability. Output is

$$y_t \; = \; F(e_t, a_t; \theta, h_t) \; = \; \Big[\alpha \, (h_t e_t)^\rho + (1 - \alpha) \, (\theta a_t)^\rho \Big]^{1/\rho}, \tag{7}$$

with $\alpha \in (0, 1)$ and $\rho \in (0, 1]$.

The parameter α governs the weight of own cognition in output; ρ governs the elasticity of substitution between own cognition and AI. As $\rho \to 1$, the inputs become perfect substitutes; as $\rho \to 0$, they become Cobb-Douglas; intermediate $\rho \in (0, 1)$ gives finite, strictly positive substitutability.[6]

The substantive content of equation (7) is that the technology of output does not, by itself, distinguish active from passive cognition. The output y_t is observationally equivalent whether produced by high e and low a or by low e and high a (with appropriate substitution). The mode-distinction enters the model not through the output technology but through the skill-formation technology developed in Sections 4 and 5.

[6]The CES form is chosen for three reasons: (i) it admits the relevant limiting cases that bracket the substantive economics; (ii) it allows separate comparative statics in the substitution elasticity and the input share; (iii) it delivers tractable closed-form input-share expressions. The qualitative results extend to any continuously differentiable, concave, increasing production function with appropriately signed cross-partial.

3.5. The governance function

The governance function $\varphi(e; g)$ identifies the process-preservation threshold: the level of delegation that constitutes use of AI as an aid to the child's active engagement rather than as a substitute for it. The function is continuous and weakly increasing in both arguments:

$$\varphi : [0, \bar{e}] \times [0, \bar{g}] \to [0, \bar{a}], \quad \varphi(0; g) = 0, \quad \partial\varphi/\partial e > 0, \quad \partial\varphi/\partial g > 0. \tag{8}$$

The interpretation is that $\varphi(e; g)$ is the maximum amount of AI use that, conditional on the child's effort level e and the household's governance level g, qualifies as process-preserving. AI use up to this level supports active engagement; AI use above this level substitutes for it. The asymmetric treatment of delegation below versus above the process-preservation threshold is the formal expression of the active-passive distinction at the level of the skill-formation technology.

Three representative specifications illustrate the range of admissible φ.[7]

The substantive content of φ is what adults do when they "structure" a child's AI use. The function may capture explicit household rules, school-level policy, parental supervision intensity, or some combination. The paper treats φ as a primitive of the environment rather than as an outcome of a separate optimization problem; the endogenization of φ as a household choice subject to parental time and competence constraints is Extension 1 in Section 8. Figure 1 displays the geometry of the boundary in (e, a) space and the comparative-static effect of higher governance. Because $\varphi(0; g) = 0$, the boundary passes through the origin for every g: with no prior effort, no delegation counts as process-preserving. Higher governance rotates the boundary upward, so the two boundaries fan apart from the origin as effort rises.

[7] The *first-attempt rule* is a widely-advocated household rule, "the child must produce a draft or attempt before AI assistance"; this is captured by $\varphi(e; g) = \min\{ge, \bar{a}\}$, where $g > 0$ permits AI use proportional to the child's prior effort. The *ratio rule* requires that delegation never exceed a fixed fraction of total work, captured by $\varphi(e; g) = e/g$ for $g \geq 1$ and $\varphi = 0$ otherwise. The *threshold rule* forbids delegation below a particular effort floor $\underline{e}(g)$ and permits unlimited process-preserving use above it: $\varphi(e; g) = 0$ for $e < \underline{e}(g)$ and $\varphi(e; g) = \bar{a}$ otherwise, with $\underline{e}(\cdot)$ decreasing in g.

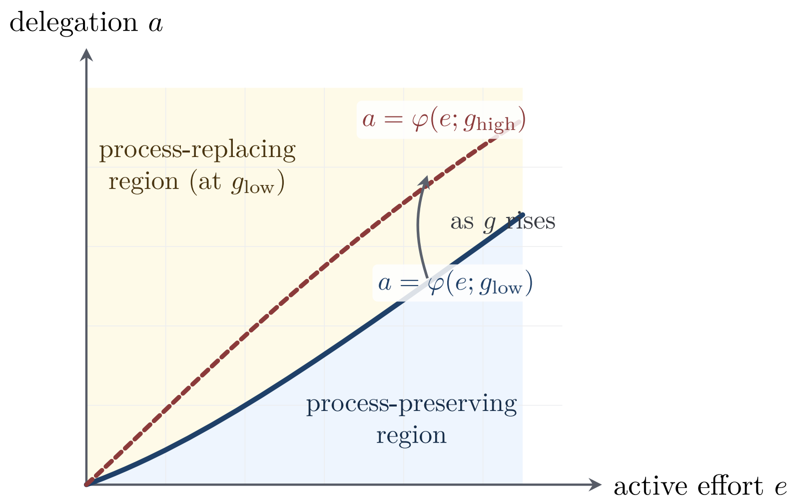

Figure 1: The process-preservation boundary in effort–delegation space. The horizontal axis is the child's active effort e; the vertical axis is delegation a. The blue curve plots $a = \varphi(e; g_{\text{low}})$ for a low-governance household: combinations below the curve (paleblue region) are process-preserving; combinations above (paleamber region) are process-replacing. The dashed red curve plots $a = \varphi(e; g_{\text{high}})$ for a high-governance household. Both boundaries pass through the origin, reflecting the assumption $\varphi(0; g) = 0$ that no delegation counts as process-preserving without prior effort; they fan apart as effort rises, with the high-governance boundary always lying above the low-governance boundary. The upward shift enlarges the process-preserving region. The threshold characterization of Proposition 3 operates through this shift: at the same observed level of AI use, low-governance households are more likely to be in the process-replacing region, and high-governance households are more likely to be in the process-preserving region.

3.6. Per-period preferences and the welfare wedge

The per-period payoff for the child is

$$u_t(x_t; s_t) \;=\; w(y_t) \;-\; c_e(e_t; D_t) \;-\; c_a(a_t) \;-\; c_v(v_t; J_t), \tag{9}$$

where $w(\cdot)$ is the reward attached to observable output (the grade, the institutional approval), c_e is the cost of effort, c_a is the cost of delegation, and c_v is the cost of verification. We assume:

- w is twice continuously differentiable, strictly increasing, weakly concave, with $w(0) = 0$.
- $c_e(e; D) = e^{1+\psi}\,[(1+\psi)(D + \underline{D})^{\sigma}]^{-1}$ for $\psi > 0$, $\sigma > 0$, $\underline{D} > 0$: cost is convex in effort and strictly decreasing in discipline. The constant $\underline{D}$ ensures finite cost at $D = 0$.
- $c_a(a) = qa + \frac{1}{2}q_2 a^2$ for $q \geq 0$, $q_2 > 0$: the marginal cost of delegation is increasing.
- $c_v(v; J) = v^{1+\psi_v}\,[(1+\psi_v)(J + \underline{J})^{\sigma_v}]^{-1}$ for $\psi_v > 0$, $\sigma_v > 0$, $\underline{J} > 0$: cost is convex in verification and strictly decreasing in judgment.

The child evaluates discounted lifetime utility under quasi-hyperbolic preferences:

$$U_t^{\text{child}} \;=\; u_t \;+\; \beta_t \sum_{s=1}^{T-t} \delta^s u_{t+s}, \tag{10}$$

where $\delta \in (0,1)$ is the long-run discount factor and $\beta_t \in (0,1]$ is the age-dependent present-bias parameter from equation (4). The social planner evaluates outcomes with exponential discounting:

$$W_t^{\text{plan}} \;=\; \sum_{s=0}^{T-t} \delta^s u_{t+s} \;+\; \delta^{T-t+1}\,\Omega(s_{T+1}). \tag{11}$$

The wedge between $\beta_t\delta$ and δ is the welfare basis for paternalistic intervention.[8]

3.7. Interpretation of the capital stocks

The three capital stocks (h, D, J) map onto three distinct empirical objects that the cognitive and developmental literatures have studied. We pause to articulate the mapping, since the analytical content of the paper depends on the substantive distinction among the three.

Knowledge h corresponds to cognitive content. Vocabulary, mathematical facts, reading decoding, scientific concepts, historical narratives, procedural fluency. This is the state measured by content assessments: state standardized tests, end-of-course examinations, vocabulary tests, reading-level assessments. The active-engagement requirement for knowledge formation is partial: a child can acquire knowledge through reading, instruction, and passive review under some conditions. The generation effect establishes that active engagement is more productive than passive reception even for content learning, but the magnitude of the active-engagement advantage is smaller for h than for D or J.

Discipline D corresponds to non-cognitive capacity. Persistence on hard tasks, willingness to revise, attention, working-memory engagement, self-regulation, tolerance for difficulty. This is the state measured by behavioral assessments and persistence tasks: time-on-task on difficult problems, willingness to attempt revision, performance on tasks requiring sustained attention. The active-engagement requirement for discipline formation is strict: discipline is built only by active engagement with difficulty. Passive reception does not build discipline, and substitution that bypasses the engagement with difficulty actively degrades the formation of discipline through the missed practice. This is the strongest empirical claim of the cognitive-science literature on active engagement.

Judgment J corresponds to epistemic capacity. Ability to detect errors, evaluate claims, compare sources, recognize what one does not know. This is the state measured by

[8] We do not assume $\beta_t < 1$ is necessary for the main results; with $\beta_t = 1$ the planner-child welfare alignment is restored at the individual level, and the policy content of the framework operates through the cross-household inequality channel rather than through the individual-level commitment channel.

error-detection tasks, source-evaluation exercises, and metacognitive assessments. The active-engagement requirement for judgment is moderate: judgment is built primarily by verification practice (which is itself active), with a secondary contribution from active engagement on original tasks (which provides experience with what correct output looks like).

The asymmetric effects of active versus passive engagement across the three stocks are central to Corollary 2 in Section 8, where we show that the three states degrade at different rates under substitution and have different recovery dynamics.

4. The Active-Passive Distinction

This section formalizes the central modeling primitive of the paper: the distinction between active cognitive engagement and passive reception of cognitive output as inputs into the formation of capacity. The distinction is not a behavioral assumption about preferences and is not a normative judgment about modes of learning. It is a property of the skill-formation technology, supported by an empirical literature whose core findings have decades of replication.

4.1. Cognitive mode as a property of the skill-formation technology

We have seen in Subsection 3.4 that the output technology $F(e, a; \theta, h)$ does not distinguish active from passive cognition. The output y_t depends on the input combination but is observationally identical across different combinations that yield the same y_t. A child who produces a five-paragraph essay through her own drafting and revision and a child who receives a five-paragraph essay from an AI system have produced identical y_t from the institution's standpoint.

The skill-formation technology, by contrast, distinguishes the two modes sharply. We now specify the laws of motion for the three capital stocks. The specification embeds the active-passive distinction as a primitive of the dynamics:

$$h_{t+1} = (1 - \delta_h)h_t + \mu_h(D_t)\, e_t + \nu_h\, a_t \cdot \mathbf{1}[a_t \leq \varphi(e_t; g)] - \lambda_h\, (a_t - \varphi(e_t; g))^+, \tag{12}$$

$$D_{t+1} = (1 - \delta_D)D_t + \pi_t\gamma_e\, e_t - \pi_t\gamma_a\, (a_t - \varphi(e_t; g))^+, \tag{13}$$

$$J_{t+1} = (1 - \delta_J)J_t + \mu_J(J_t)\, [\, v_t + \eta\, e_t\,] - \lambda_J\, (a_t - \varphi(e_t; g))^+, \tag{14}$$

where $(z)^+ \equiv \max\{z, 0\}$, all parameters $\delta_h, \delta_D, \delta_J, \mu_h, \mu_J, \nu_h, \lambda_h, \lambda_J, \gamma_e, \gamma_a, \eta$ are non-negative with $\delta_h, \delta_D, \delta_J \in (0, 1)$, and the functions $\mu_h(\cdot)$ and $\mu_J(\cdot)$ are weakly increasing.

The structure of the three laws of motion encodes the active-passive distinction in three formal observations.

Observation 1: Active effort is the unambiguous input. The effort variable e_t enters all three accumulation equations with strictly positive sign. Effort builds knowledge (through $\mu_h(D_t)e_t$), discipline (through $\pi_t\gamma_e e_t$), and judgment (through $\mu_J(J_t)\eta e_t$, with $\eta < 1$ reflecting the partial contribution of original engagement to evaluative capacity). Effort is the active-engagement input.

Observation 2: Process-preserving delegation contributes to knowledge only. When $a_t \leq \varphi(e_t; g)$, delegation enters the knowledge equation with positive sign (through $\nu_h a_t$), reflecting the empirically supported claim that process-preserving use of AI as a tutor can complement active engagement in content learning. Process-preserving delegation contributes nothing to the discipline or judgment states: in our specification, the indicator $\mathbf{1}[a \leq \varphi]$ activates only for h. The economic interpretation is that AI used as a tutor can deliver content knowledge but cannot substitute for the active engagement that builds discipline or for the verification practice that builds judgment.

Observation 3: Process-replacing delegation actively degrades all three stocks. When $a_t > \varphi(e_t; g)$, the excess $(a_t - \varphi(e_t; g))^+$ enters all three accumulation equations with strictly negative sign. The economic interpretation is that delegation in excess of the process-preservation threshold represents replacement of the child's active engagement: the cognitive task that should have been the child's becomes the AI's, and the formation of capacity that should have occurred in the child does not. The asymmetry between process-preserving and process-replacing delegation is the technological counterpart of the cognitive-science distinction between active engagement and passive reception.

4.2. Functional-form choices and their substantive content

We have made several specific functional-form choices in equations (12)–(14). Each choice has substantive content that we discuss.

The hinge at φ. The indicator $\mathbf{1}[a \leq \varphi]$ produces a sharp transition between the process-preserving regime and the process-replacing regime. This is a stylized assumption: in practice the transition may be smoother, with intermediate ranges of a producing partial-substitution effects. The sharp specification is chosen for analytical clarity; the qualitative results are robust to smooth specifications, as we show in the robustness section.

The asymmetric coefficient structure. Effort enters with coefficient γ_e in the discipline equation; process-replacing delegation enters with coefficient $-\gamma_a$. The two coefficients are independent primitives. We do not impose $\gamma_e = \gamma_a$: the empirical evidence from the cognitive-science literature does not pin down the relative magnitudes. The qualitative results require both to be strictly positive but do not require any particular ordering.

The plasticity factor in the discipline equation. The factor π_t multiplies both the gain from effort and the loss from process-replacing delegation in the discipline equation but not in the knowledge or judgment equations. This reflects the cognitive-science finding that discipline (non-cognitive capacity) has an age-dependent plasticity that is sharper than that of knowledge or judgment. Adjusting this choice (multiplying π_t into h or J as well) does not change the qualitative structure of the Capacity-Wedge Decomposition.

The dependence of μ_h on D_t. The productivity of effort in producing knowledge is increasing in discipline. This is the dynamic-complementarity link between non-cognitive and cognitive skill formation that Cunha, Heckman and Schennach (2010) document empirically. It implies that a child with low discipline cannot easily produce knowledge through effort, even at the same nominal effort level. The dependence of μ_J on J_t is analogous: judgment is itself a productive input into the formation of further judgment, through the role of existing evaluative capacity in detecting errors during verification.

4.3. The active-passive distinction at the level of cognitive content

We have presented the active-passive distinction at the level of the skill-formation technology. It is useful to articulate what the distinction means in terms of the cognitive experience of the child, with concrete examples drawn from the empirical literature.

Mathematics. Active engagement in mathematics consists of attempting problems, generating solution paths, performing the steps of a calculation, retrieving facts from memory, checking one's own work. Process-replacing delegation consists of asking an AI system for the answer, the worked solution, or the explanation, and submitting or accepting the result without performing the underlying activity. The cognitive-science literature on mathematics learning (Hiebert and Grouws, 2007; Kapur, 2014) establishes that the active-engagement mode produces both content knowledge (mathematical facts, procedural fluency) and process capacity (the ability to attempt new problems, to recognize when an answer is wrong, to persist through difficulty). The process-replacing mode produces the appearance of content knowledge at the moment of submission but neither the durable content nor the process capacity.

Reading and writing. Active engagement in reading consists of reading the passage, working through unfamiliar vocabulary, monitoring one's own comprehension, re-reading when comprehension fails. Active engagement in writing consists of drafting, reviewing one's draft, revising for clarity and structure, reading the revised version. Process-replacing delegation in reading consists of asking an AI for a summary; in writing, of asking an AI for a draft. The cognitive-science literature on reading comprehension (Pressley, 2002) and writing development (Bereiter and Scardamalia, 1987) establishes that the active mode builds reading stamina and compositional skill, while the process-replacing mode produces neither.

Foreign language and vocabulary. Active engagement consists of retrieval practice, generation of sentences in the target language, attempting comprehension with limited support. Process-replacing delegation consists of using AI translation, accepting generated text, or receiving paraphrased explanations of foreign-language material. The active mode builds productive language capacity; the process-replacing mode produces transient familiarity that does not transfer.

The examples illustrate that the active-passive distinction is not a single binary but a continuum that takes a particular form in each cognitive domain. The model's stylized e-vs-a

distinction is an abstraction over this continuum. The empirical claim that underwrites the abstraction is that, within each domain, the active mode produces formation of capacity and the passive mode does not, by margins that are well-documented and quantitatively material.

4.4. Connection to cognitive-science measurement

The empirical content of the active-passive distinction has been measured in many ways across the cognitive-science literature. We summarize the most direct measurements to clarify the empirical grounding of the modeling primitive.

Standard measurements report quantitatively material effect sizes: Slamecka and Graf (1978) found 60–80% more material recalled under generation than under passive reading in paired-associate paradigms.[9]

Roediger and Karpicke (2006) measured the testing effect using free-recall paradigms with matched study time, finding that students who studied through retrieval practice recalled approximately 50% more material at one-week retention intervals than students who studied through repeated re-reading. The effect strengthens with longer retention intervals, indicating that the active-engagement mode produces specifically durable rather than transient learning.

Karpicke (2012) measured the practical difference between retrieval-based study and elaborative study (concept mapping with the text available) in undergraduate biology courses, finding that retrieval-based study produced approximately 50% better performance on transfer tests one week later. The substantive implication is that the active-engagement advantage extends beyond surface retention to inferential and conceptual transfer.

Kapur (2014) measured the productive-struggle effect in mathematics, finding that students who attempted complex problems before instruction performed better on subsequent transfer problems than students who received instruction first and then practiced. The effect is again moderate-to-large in standard variants.

The convergence of these measurements supports the treatment of the active-passive distinction as an empirically established primitive of the skill-formation technology, with effect sizes that are not marginal. The asymmetric production-function structure in equations (12)–(14) is a formal reflection of these measurements.

5. Downstream Complementarity

The skill-formation technology developed in Section 4 already exhibits a form of complementarity: the productivity of effort in producing knowledge depends on the discipline state, through $\mu_h(D_t)$. This is the canonical dynamic complementarity of the Cunha-Heckman framework. The cascade result of the paper requires an additional structural feature: *downstream complementarity*, the property that knowledge acquired at age t is an input into the productivity of further investment at all subsequent ages. This section formalizes downstream

[9] The generation effect has been replicated across decades and across paradigms; meta-analyses (Bertsch et al., 2007) report effect sizes consistently in the moderate-to-large range, and the corresponding effects for the testing effect, desirable difficulties, and productive struggle are similarly robust.

complementarity, introduces three specifications under which it operates, and discusses the empirical interpretation across curricular domains.

5.1. The downstream-complementarity assumption

We extend the knowledge production function in equation (12) to admit dependence on the knowledge stock at all prior ages, not only on the contemporaneous discipline state. Specifically, we replace $\mu_h(D_t)$ with a function $\mu_h(D_t, \mathbf{h}_{<t})$ where $\mathbf{h}_{<t} = (h_0, h_1, \dots, h_{t-1})$ is the history of knowledge stocks prior to period t:

$$h_{t+1} \;=\; (1-\delta_h)h_t + \mu_h(D_t, \mathbf{h}_{<t})\, e_t + \nu_h a_t \cdot \mathbf{1}[a_t \le \varphi(e_t; g)] - \lambda_h(a_t - \varphi(e_t; g))^+, \tag{15}$$

with $\partial\mu_h/\partial h_s > 0$ for $s < t$. The dependence captures the empirically supported claim that the productivity of effort at any age depends on the foundations already acquired. A child without arithmetic fluency cannot easily make use of effort spent on algebra; a child without phonics cannot easily make use of effort spent on reading comprehension; a child without basic vocabulary cannot easily make use of effort spent on advanced composition.

The substantive content of equation (15) is the explicit assertion that knowledge formation depends not only on the contemporaneous flow of effort and the contemporaneous state of discipline, but on the entire history of prior knowledge stocks. The implication, as we shall see in Section 7, is that substitution at any age has consequences that extend to all subsequent ages: a foundation skipped at age six is an input missing at ages seven, eight, nine, and so on through the schooling sequence.

5.2. The baseline cascade specification

For the main analysis we work with a single baseline specification of μ_h. We take the polynomial-in-distance form: knowledge productivity at age t depends on the entire history of prior knowledge stocks with weight that decreases polynomially in the curricular distance from the present:

$$\mu_h(D_t, \mathbf{h}_{<t}) \;=\; \mu_h^0(D_t) \cdot \left[1 + \sum_{s=1}^{t} \frac{\kappa}{s^p}\, h_{t-s}\right], \tag{16}$$

where $\kappa > 0$ governs the strength of the cascade and $p > 1$ governs its decay rate. Under this specification, the cascade operates through the entire prior history with diminishing weight at greater distance. The present-value cost of substitution at age t depends on the foundations the child has yet to use, which is roughly the remaining curricular sequence from t to T. The choice of polynomial decay reflects the empirical observation that foundational dependencies in core subjects (mathematics, foreign language, reading) extend across many curricular periods rather than vanishing after a single period.

Alternative specifications (uniform single-period dependence, exponential decay) yield the same qualitative results with different quantitative magnitudes. We develop these in Appendix C.8 as robustness checks and confine the main analysis to the baseline above.

5.3. Empirical interpretation across curricular domains

The strength and structure of downstream complementarity vary across curricular domains. We discuss four representative cases.

Mathematics. Mathematics has the strongest and most sequential downstream complementarity in the standard curriculum. Arithmetic fluency is a prerequisite for fraction operations, which are prerequisites for algebraic manipulation, which are prerequisites for calculus. Each foundation is an input into the productivity of effort on the next layer. A child who has not acquired arithmetic fluency cannot productively engage with fraction operations: the cognitive load of computing each arithmetic step consumes the working-memory capacity that would otherwise be available for the fraction-operation logic. Specification P with high κ_2 and low p is the natural match for mathematics, with the cascade extending across the entire schooling life cycle.

Reading and language. Reading exhibits strong downstream complementarity through the dependence of comprehension on decoding fluency. A child who has not automatized decoding cannot productively engage with comprehension instruction: the cognitive load of decoding consumes the working-memory capacity that would otherwise be available for comprehension. Foreign-language learning exhibits similar structure: vocabulary and basic grammar are inputs into all subsequent productive language use. Specification P or E with moderate parameters matches reading and language learning.

Writing. Writing exhibits more complex complementarity. Mechanics (grammar, spelling, sentence structure) function as foundational inputs; compositional skill (argument structure, voice, rhetorical strategy) builds on these. But the cascade in writing is less strict than in mathematics: a child can produce competent writing in some genres without prior mastery of all mechanical elements, and the relationship between practice on simple and on complex writing is bidirectional. Specification U (with single-period horizon) or specification E (with rapid decay) matches writing learning.

Civic education and historical reasoning. These domains exhibit the weakest downstream complementarity. Knowledge of specific historical events is not in general a foundation for the analysis of other historical events; civic understanding builds through accumulation rather than through strict sequential dependence. Specification U with low κ_1, or no cascade at all, may best match these domains.

The substantive implication of the cross-domain variation is that the policy implications of the Capacity-Wedge Decomposition are subject-specific. Effort-contingent restriction on AI use should be tightest in domains with the deepest cascades (mathematics, foreign language, foundational reading) and may be relaxed in domains with weaker cascades (civic education, advanced literature). We develop this implication formally as Corollary 3 in Section 8.

5.4. Standing assumptions for the analysis

We collect the maintained assumptions for the analysis that follows. Assumptions A1 and A2 are regularity and functional-form conditions; Assumptions A3 through A7 are the economically substantive primitives, each of which we interpret in turn.

Assumption A1 (Smoothness)**.** The functions $w(\cdot)$, $c_e(\cdot;\cdot)$, $c_a(\cdot)$, $c_v(\cdot;\cdot)$, $\mu_h(\cdot,\cdot)$, $\mu_J(\cdot)$, $\varphi(\cdot;\cdot)$ are twice continuously differentiable on the interior of their respective domains.

This is a standard regularity condition ensuring that the first-order conditions of the child's problem are well-defined.

Assumption A2 (CES output technology)**.** The output technology F takes the CES form of equation (7) with $\alpha \in (0,1)$ and $\rho \in (0,1]$.

The CES form fixes the output technology for tractability; Section 9 shows that the qualitative results extend to any smooth, concave, increasing technology in which own cognition and AI use are imperfectly substitutable.

Assumption A3 (Active engagement is the unambiguous input)**.** The coefficients $\gamma_e, \mu_h^0, \mu_J, \eta$ in equations (12)–(14) are strictly positive.

Active effort builds all three capital stocks. This is the formal counterpart of the cognitive-science finding that active construction is productive across knowledge, discipline, and judgment alike.

Assumption A4 (Process-replacing delegation degrades capacity)**.** The coefficients $\gamma_a, \lambda_h, \lambda_J$ in equations (12)–(14) are strictly positive.

This is the central asymmetric primitive of the model: process-replacing delegation degrades the formation of capacity, whereas process-preserving delegation does not. It operationalizes the active-passive distinction documented in the cognitive-science literature reviewed in Section 2, under which active construction and passive reception are not interchangeable inputs into durable learning. We take this qualitative content as a primitive and characterize its dynamic and welfare implications; the specific functional form, including the hinge at φ and the linear-in-excess degradation rate, is a modeling choice made for tractability rather than a property derived from the empirical literature.

Assumption A5 (Discipline complementarity)**.** The cost-of-effort function $c_e(e;D)$ is strictly decreasing in D on the relevant range: $\partial c_e/\partial D < 0$.

This embeds self-productivity: a higher discipline stock lowers the marginal cost of effort, so that capacity accumulated earlier raises the return to investment later.

Assumption A6 (Downstream complementarity)**.** The function $\mu_h(D, \mathbf{h}_{<t})$ is strictly increasing in h_s for at least one $s < t$ with $t - s \leq T$.

This is the source of the cascade: knowledge acquired earlier raises the productivity of effort later, so a foundation missed at one age depresses learning at all subsequent ages.

Assumption A7 (Plasticity decline and agency development)**.** The functions π_t, β_t, ω_t from equations (3), (4), (6) are well-defined with the stated monotonicity properties.

This captures the developmental structure of childhood: plasticity is highest, and strategic agency lowest, in early childhood, with both attenuating as the child ages.

The seven assumptions are the minimal structure required for the Capacity-Wedge Decomposition. We do not impose any particular shape on μ_h beyond Assumption A6; the baseline polynomial-in-distance specification (16) and the alternative specifications in Appendix C.8 are illustrative special cases. Section 9 discusses the role of each assumption.

6. The Child's Dynamic Problem and Equilibrium

This section develops the child's intertemporal optimization problem under the primitives introduced in Sections 3 through 5. The development is standard in form. The substantive content lies in the characterization of the equilibrium policy and the monotonicity properties that will be used in the proof of the main theorem.

6.1. The recursive formulation

For the main analysis we work with the case $\omega_t = 1$ (full agency), with the agency-weighted case treated as Extension 1 in Section 8. The state at the start of period t is $s_t = (h_t, D_t, J_t)$ supplemented by the history $\mathbf{h}_{<t}$ that enters μ_h under the cascade specification. We define the augmented state $\tilde{s}_t = (s_t, \mathbf{h}_{<t})$ and the augmented state space $\tilde{\mathcal{S}}$. The action set $\mathcal{X}$ is as in equation (2).

The child's recursive value function $V_t(\tilde{s}_t)$ satisfies the Bellman equation

$$V_t(\tilde{s}_t) \;=\; \max_{x_t \in \mathcal{X}} \left\{ u_t(x_t; s_t) + \beta_t \delta \, W_{t+1}\Big(\tilde{s}_{t+1}(x_t, \tilde{s}_t)\Big) \right\}, \tag{17}$$

where $u_t(x_t; s_t)$ is the per-period payoff of equation (9), the state transition $\tilde{s}_{t+1}(x_t, \tilde{s}_t)$ is governed by equations (15), (13), and (14), and the continuation value W_{t+1} is the value function evaluated by future selves with their own (potentially different) present-bias parameters:

$$W_{t+1}(\tilde{s}_{t+1}) \;=\; u_{t+1}(x^*_{t+1}; s_{t+1}) + \delta \, W_{t+2}\Big(\tilde{s}_{t+2}(x^*_{t+1}, \tilde{s}_{t+1})\Big), \tag{18}$$

with x^*_{t+1} the equilibrium policy at time $t+1$. The distinction between V_t and W_t is the standard quasi-hyperbolic device: V_t is the current self's perceived value (with present-bias weight β_t), while W_t is the value used by current and future selves to evaluate continuation from period $t+1$ onward (with exponential weight δ). The two coincide when $\beta_t = 1$.

6.2. Sophisticated equilibrium

We adopt the sophisticated-equilibrium concept following Harris and Laibson (2001) and Krusell and Smith (2003).

Definition 1 (Sophisticated equilibrium)**.** A *sophisticated equilibrium* of the child's problem is a sequence of policy functions $\{x_t^*(\tilde{s})\}_{t=0}^{T}$ and continuation values $\{W_t(\tilde{s})\}_{t=0}^{T+1}$ such that for each $t \in \{0, \dots, T\}$ and each $\tilde{s} \in \tilde{\mathcal{S}}$:

(i) $x_t^*(\tilde{s}) \in \arg\max_{x \in \mathcal{X}} \Big\{ u_t(x; s) + \beta_t \delta \, W_{t+1}(\tilde{s}_{t+1}(x, \tilde{s})) \Big\}$;

(ii) $W_t(\tilde{s}) = u_t(x_t^*(\tilde{s}); s) + \delta \, W_{t+1}(\tilde{s}_{t+1}(x_t^*(\tilde{s}), \tilde{s}))$;

(iii) $W_{T+1}(\tilde{s}) = \Omega(s)$, the terminal continuation value.

The sophisticated equilibrium is the appropriate solution concept because the child's future selves take their own actions given their own present-bias parameters. The current self anticipates the future selves' play (sophistication) but cannot directly choose it. Existence and uniqueness of the sophisticated equilibrium follows from standard arguments.

Proposition 1 (Existence)**.** *Under Assumptions A1–A7, a sophisticated equilibrium exists. If, in addition, w is strictly concave or the cost functions are strictly convex in their first arguments, the equilibrium is unique up to a set of measure zero in $\tilde{\mathcal{S}}$.*

Proof. The action set $\mathcal{X}$ is compact and convex. The per-period payoff u_t is continuous in x for each $\tilde{s}$ and continuous in $\tilde{s}$ for each x. The state transition is continuous. The terminal value Ω is bounded. The Bellman operator

$$(T_t W)(\tilde{s}) \;=\; \max_{x \in \mathcal{X}} \Big\{ u_t(x; s) + \delta \, W(\tilde{s}_{t+1}(x, \tilde{s})) \Big\}$$

maps bounded continuous functions on $\tilde{\mathcal{S}}$ to bounded continuous functions and is a contraction with modulus $\delta < 1$. The fixed point is the continuation value function W. The policy follows by Berge's theorem of the maximum. The strict-concavity conditions deliver uniqueness via standard arguments. For the quasi-hyperbolic case ($\beta_t < 1$), the analogous fixed-point argument applies to the system of equations (17)–(18); see Krusell and Smith (2003) for the foundational treatment. □

6.3. First-order conditions

We derive the first-order conditions of the child's problem to characterize the equilibrium policy. The derivation is straightforward but produces the bounds that are used in the proof of the main theorem.

The interior FOC for effort, in the process-preserving regime $a \le \varphi(e; g)$, is

$$w'(y_t) \, \frac{\partial F}{\partial e} \;-\; \frac{\partial c_e}{\partial e} \;+\; \beta_t \delta \, \Lambda_h^{t+1} \, \mu_h \, (\partial \mu_h / \partial e)^{-1} \, [\text{cascade chain}] \;+\; \beta_t \delta \, \Lambda_D^{t+1} \, \pi_t \gamma_e \;=\; 0, \quad (19)$$

where $\Lambda_h^{t+1} = \partial W_{t+1} / \partial h$ and $\Lambda_D^{t+1} = \partial W_{t+1} / \partial D$ are the shadow values of the next-period knowledge and discipline states, and the "cascade chain" term represents the cascade contribution from the dependence of μ_h at future ages on the current h_t. We will derive this term explicitly in Section 7 when we prove the cascade theorem; here we note only that it enters the FOC with strictly positive sign when downstream complementarity is present.

The interior FOC for delegation, in the process-replacing regime $a > \varphi(e; g)$, is

$$w'(y_t)\,\frac{\partial F}{\partial a} \;-\; \frac{\partial c_a}{\partial a} \;-\; \beta_t\delta\,\Lambda_h^{t+1}\,\lambda_h \;-\; \beta_t\delta\,\Lambda_D^{t+1}\,\pi_t\gamma_a \;-\; \beta_t\delta\,\Lambda_J^{t+1}\,\lambda_J \;=\; 0. \tag{20}$$

The marginal value of process-replacing delegation comprises the current output reward minus the direct delegation cost minus the future costs through the three capital channels. In the process-preserving regime $a < \varphi(e; g)$, the equation simplifies to remove the three discount-weighted state-channel terms.

The first-order conditions exhibit a kink at $a = \varphi(e; g)$. The marginal value of a is strictly higher on the process-preserving side of the boundary than on the process-replacing side, because the state-channel costs $\lambda_h, \gamma_a, \lambda_J$ activate above the boundary. The kink determines the qualitative structure of the equilibrium policy, with three potential regimes: (i) interior in $[0, \varphi]$ (governed delegation), (ii) at the kink $a^* = \varphi$, (iii) interior in $[\varphi, \bar{a}]$ (process-replacing delegation).

6.4. Monotonicity properties

The proof of the main theorem rests on six monotonicity properties of the equilibrium policy. We state the lemmas here and prove them in the appendix.

Lemma 1 (Monotonicity in discipline). *$\partial e^*/\partial D \geq 0$ and $\partial a^*/\partial D \leq 0$ at any interior optimum.*

Lemma 2 (Monotonicity in knowledge). *Under Assumption A6, $\partial e^*/\partial h \geq 0$ at any interior optimum: higher knowledge raises optimal effort through the cascade-chain term in equation (19).*

Lemma 3 (Monotonicity in AI capability). *$\partial a^*/\partial \theta \geq 0$ and $\partial e^*/\partial \theta \leq 0$ at any interior optimum.*

Lemma 4 (Monotonicity in present bias). *At any interior optimum where the discipline channel is operative: $\partial e^*/\partial \beta_t \geq 0$ and $\partial a^*/\partial \beta_t \leq 0$.*

Lemma 5 (Monotonicity in governance). *$\partial e^*/\partial g \geq 0$ and $\partial a^*/\partial g$ has ambiguous sign in general: it is positive on the process-preserving side of the kink (more process-preserving delegation is permitted) and negative on the process-replacing side (less process-replacing delegation is optimal).*

Lemma 6 (Monotonicity in age). *Holding the state fixed, $\partial e^*/\partial t$ has ambiguous sign through the composition of π_t, β_t, and ω_t. The age-dependence of the optimal policy is the subject of Corollary 1.*

The six monotonicity lemmas are standard applications of the implicit function theorem to the FOCs. The proofs are in the appendix (Section A). The economic intuition for each is summarized in the table below for reference.

Parameter	Effect on e^*	Effect on a^*
D (discipline)	Positive: lower effort cost.	Negative: less substitution attractive.
h (knowledge)	Positive: cascade chain raises marginal return to effort.	Indirect: through cross-effects.
θ (AI capability)	Negative: AI substitutes for effort.	Positive: AI is more productive.
β_t (present-bias attenuation)	Positive: future returns weighted more.	Negative: future costs weighted more.
g (governance)	Positive: more process-preserving delegation permits more output.	Ambiguous: depends on regime.
t (age)	Ambiguous: composition of π_t, β_t, ω_t.	Ambiguous: same.

Table 1: Comparative statics of the equilibrium policy. Signs of cross-derivatives at the interior optimum.

6.5. The process-preserving versus process-replacing choice

Before proceeding to the main theorem, it is useful to characterize when the child chooses the process-preserving regime ($a^* \le \varphi(e^*; g)$) versus the process-replacing regime ($a^* > \varphi(e^*; g)$) under the model's primitives.

The choice is governed by the trade-off between the output-reward gain from raising a above φ and the state-channel cost of process-replacing delegation. The output-reward gain from a marginal unit of process-replacing delegation is $w'(y)\partial F/\partial a$. The state-channel cost is $\beta_t \delta(\Lambda_h^{t+1}\lambda_h + \Lambda_D^{t+1}\pi_t\gamma_a + \Lambda_J^{t+1}\lambda_J)$. The child remains in the process-preserving regime if and only if the latter exceeds the former at the kink point.

We can rewrite the condition equivalently: the child remains in the process-preserving regime if and only if

$$\beta_t \delta \left[\Lambda_h^{t+1}\lambda_h + \Lambda_D^{t+1}\pi_t\gamma_a + \Lambda_J^{t+1}\lambda_J\right] \;\ge\; w'(F(e^*, \varphi(e^*; g); \theta, h)) \left.\frac{\partial F}{\partial a}\right|_{a=\varphi(e^*;g)}. \tag{21}$$

The condition has several substantive implications.

First, the condition is more likely to hold when β_t is larger (less present bias): a more forward-looking child internalizes the long-run state-channel cost. This implies that the process-preserving regime is more easily achieved at older ages, when $\beta_t \to \beta_\infty$.

Second, the condition is more likely to hold when $\Lambda_h^{t+1}, \Lambda_D^{t+1}, \Lambda_J^{t+1}$ are larger: a child for whom the future state-channel returns are large internalizes more. Under the cascade specification, Λ_h^{t+1} has an explicit dependence on the remaining curricular distance from age t, and is larger at younger ages. This produces a partial offsetting effect: younger children have stronger cascade incentives to remain in the process-preserving regime, but weaker contemporaneous internalization through β_t.

Third, the condition is more likely to hold when θ is smaller: the output-reward gain from process-replacing delegation scales with θ through $\partial F/\partial a$. As θ grows, the right-hand side of the inequality grows, and the process-replacing regime becomes more attractive.

Fourth, the condition is more likely to hold when g is larger: a larger governance level shifts the process-preservation threshold φ outward, raising the level of a at which the process-replacing regime activates. The composition of higher g and lower θ makes the process-preserving regime more easily achieved; the composition of lower g and higher θ makes the process-replacing regime more easily reached.

The four implications jointly produce the household-threshold characterization in Proposition 3.

7. The Capacity-Wedge Decomposition

This section states and proves the main theoretical result. The result is a single decomposition theorem: the long-run capacity wedge under process-replacing delegation is the sum of a contemporaneous discipline-channel cost and a cascade cost from downstream knowledge complementarities. Two immediate corollaries follow: an age-structure result and, under sufficient conditions, a governance-threshold result. The theorem and its two corollaries derive from the same primitive structure, the active-passive distinction interacting with downstream skill complementarities, and rely on the same equilibrium characterization developed in Section 6.

7.1. A preliminary inequality

Before stating the main theorem, we record a preliminary result that is intentionally weak. The process-preserving regime weakly dominates the process-replacing regime in long-run capacity formation. We do not claim observational equivalence with a counterfactual world in which AI does not exist, nor do we claim that process-preserving delegation is developmentally neutral in general. The substantive content of the inequality is comparative: between the two regimes the model permits, the process-preserving one yields weakly better long-run outcomes.

Proposition 2 (Weak inequality). *Under Assumptions A1–A7, for every (g, θ) and every initial state $\tilde{s}_0$,*

$$W^P(\tilde{s}_0; g, \theta) \;\geq\; W^R(\tilde{s}_0; g, \theta),$$

where W^P denotes welfare under any process-preserving equilibrium path ($a_t \leq \varphi(e_t; g)$ for all t) and W^R denotes welfare under the equilibrium with at least one t at which $a_t > \varphi(e_t; g)$. The inequality is strict whenever the process-replacing intensity is positive on a positive-measure subset of the path.

The proposition records the comparative claim and avoids stronger statements. We do *not* claim that process-preserving delegation is identical to the no-AI baseline. Empirically, scaffolded use of AI can change effort quality, reduce productive struggle in some domains, alter the composition of cognitive engagement, or shift the curricular focus. The model abstracts from these channels for tractability, and the cognitive-science evidence does not by itself pin down whether they operate at the margin. We treat process preservation as a comparative property: among feasible regimes the technology admits, this is the one that does not trigger the cascade-amplified capacity loss characterized below. We return to the limitations of this abstraction in the discussion.

7.2. The Capacity-Wedge Theorem

The single formal result of the paper is the decomposition of the long-run capacity wedge into a contemporaneous component and a cascade component. The decomposition is the unmistakable center of the paper. Two corollaries develop the immediate implications for the age structure of the wedge and for its dependence on household governance; we record them after the theorem.

Theorem 1 (Capacity-Wedge Decomposition)**.** *Under Assumptions A1–A7, the long-run capacity wedge $\mathcal{W}^{active} - W^R$ between the active-engagement counterfactual and the process-replacing equilibrium decomposes additively into a contemporaneous discipline-channel cost and a cascade cost from downstream knowledge complementarities:*

$$\mathcal{W}^{active} - W^R \; = \; \mathcal{C}^{contemp} + \mathcal{C}^{cascade}, \tag{22}$$

where

$$\mathcal{C}^{contemp} \; = \; \sum_{t=0}^{T} \delta^t \cdot \pi_t \gamma_a \cdot (a_t - \varphi(e_t; g))^+ \cdot \Psi_D, \tag{23}$$

$$\mathcal{C}^{cascade} \; = \; \sum_{t=0}^{T} \delta^t \cdot \lambda_h \cdot (a_t - \varphi(e_t; g))^+ \cdot \Psi_h(t), \tag{24}$$

and $\Psi_D, \Psi_h(t)$ are positive shadow-value terms. Under the baseline cascade specification of Section 5, $\Psi_h(t)$ is strictly increasing in the remaining curricular distance and strictly larger than Ψ_D for t in early and middle childhood, so that the cascade cost dominates the contemporaneous cost in this range.

The theorem says the capacity wedge has two structural sources, contemporaneous and cumulative, and identifies the cascade component as the larger of the two for the range of t in which downstream complementarity is operative. The proof of the decomposition appears below. We now state the two immediate corollaries that develop the age and governance implications of the decomposition.

Corollary 1 (Age structure)**.** *Under the conditions of Theorem 1, define the marginal contribution of process-replacing delegation at age t as*

$$\mathcal{M}(t) \; \equiv \; \left. \frac{d\mathcal{W}^{active}}{d(a_t - \varphi(e_t; g))^+} \right|_{equilibrium\ path}.$$

Then $\mathcal{M}(t)$ is strictly decreasing in t over $\{0, \ldots, T-1\}$, with explicit lower bound

$$\mathcal{M}(t) \; \geq \; \mathcal{M}_0 \cdot \Big[\pi_t/\pi_0 + (1-\delta_h)^{T-t} \kappa_h (T-t)\Big].$$

The first term reflects plasticity decline; the second reflects cascade-distance amplification. The substantive implication is that the marginal welfare cost of process-replacing delegation is concentrated in early childhood.

Proposition 3 (Governance threshold under sufficient conditions)**.** *Suppose the following sufficient conditions hold: (a) the marginal output-channel gain $\partial V_0/\partial\theta$ has opposite signs at $g = 0$ and $g = \bar{g}$ for θ in a non-empty interval $[\theta_1, \theta_2]$; (b) the cross-partial $\partial^2 V_0/\partial\theta\partial g$ is strictly positive on the interior of the relevant region; (c) V_0 and φ are sufficiently smooth in g on this region. Then for each $\theta \in [\theta_1, \theta_2]$ there exists a unique threshold function $g^*(\theta)$ such that $\partial V_0(\tilde{s}_0; g, \theta)/\partial\theta > 0$ for $g > g^*(\theta)$ and < 0 for $g < g^*(\theta)$. The threshold is strictly increasing in θ, and the cross-sectional variance of welfare across households is strictly increasing in θ in this range.*

We deliberately state the governance result as a proposition under sufficient conditions rather than as part of the main theorem. The conditions (a)–(c) are non-trivial: condition (a) is a sign assumption at the endpoints, condition (b) is the monotone cross-partial, and condition (c) is the smoothness required for the implicit-function argument. Each of these conditions can fail under reasonable variations of the model. The decomposition of Theorem 1 is robust; the governance threshold is contingent on the additional structure imposed by (a)–(c).

7.3. Proof of Proposition 2: Weak inequality

The substantive claim of the proposition is purely comparative: between the two regimes the model permits, the process-preserving regime yields weakly higher welfare than the process-replacing regime. We do not claim observational or developmental equivalence with a no-AI counterfactual; that claim depends on additional invariance assumptions about effort response and cognitive-mode quality that we do not impose. The proof works at the level of the three capital-stock laws of motion and compares the two regimes directly.

Step 1: Discipline accumulation is weakly higher in the process-preserving regime. Compare equation (13) under the two regimes. In the process-preserving regime, $a_t \leq \varphi(e_t; g)$, so $(a_t - \varphi(e_t; g))^+ = 0$ and discipline accumulates at rate $\pi_t\gamma_e e_t$. In the process-replacing regime, the same effort produces discipline at the same rate, but the additional substitutive term $-\pi_t\gamma_a(a_t - \varphi(e_t; g))^+$ enters with strictly negative sign. Hence $D^P_{t+1} \geq D^R_{t+1}$ at equal effort and equal initial state, with strict inequality whenever process-replacing intensity is positive. The same comparison applies to judgment J via equation (14).

Step 2: Knowledge accumulation is weakly higher in the process-preserving regime. In the process-preserving regime, knowledge accumulates as

$$h^P_{t+1} = (1 - \delta_h)h^P_t + \mu_h(D^P_t, \mathbf{h}^P_{<t})e^P_t + \nu_h a^P_t,$$

with non-negative contributions from both effort and process-preserving delegation. In the process-replacing regime, the substitutive excess $-\lambda_h(a_t - \varphi(e_t; g))^+$ enters with negative sign. The output benefit $\nu_h a_t$ that operates in the process-preserving regime does not operate in the process-replacing regime by the indicator structure of equation (15). The comparison gives $h^P_{t+1} \geq h^R_{t+1}$ at equal effort and equal initial state, with strict inequality whenever process-replacing intensity is positive.

Step 3: Welfare comparison. The planner's value function is monotone increasing in each capital stock. By Steps 1 and 2, all three states are weakly higher in the process-preserving regime, with strict inequality somewhere on the path when the process-replacing intensity is positive. Hence $W^P \geq W^R$, with strict inequality under the strict-state-difference condition.

The proof establishes only the comparative claim. We make no attempt to compare W^P with the no-AI counterfactual W^0. The process-preserving regime might in principle yield lower welfare than the no-AI counterfactual through channels the model abstracts from (e.g., quality of effort substitution, attentional reallocation, perceived ease leading to reduced productive struggle); these channels are not captured by the present specification, and the model neither rules them in nor rules them out. We return to this limitation in the discussion. □

7.4. Proof of the Capacity-Wedge Decomposition (Theorem 1)

We prove the decomposition $\mathcal{W}^{\text{active}} - W^R = \mathcal{C}^{\text{contemp}} + \mathcal{C}^{\text{cascade}}$ in four steps. We define the two shadow-value objects Ψ_D and $\Psi_h(t)$ explicitly as derivatives of the continuation value, then derive each component of the wedge from the laws of motion. We work with the continuation value $W(s,t)$ defined by equation (18) along the equilibrium path, so that $\partial W/\partial D$ and $\partial W/\partial h$ are well-defined under Assumption A1.

Step 1: Definition of Ψ_D and $\Psi_h(t)$. Define the discipline shadow value at age t as the discounted present value of the marginal effect of a unit of discipline at age t on future periods' welfare. Along the equilibrium path,

$$\Psi_D(t) \;\equiv\; \sum_{r=t}^{T} \delta^{r-t}\,(1-\delta_D)^{r-t}\,\frac{\partial u_r}{\partial D_r}\bigg|_{\text{eqm}}. \tag{25}$$

Each term in the sum is the discounted contribution to current and future per-period utility of an additional unit of D accumulated at age t, depreciated by $(1-\delta_D)^{r-t}$ at age r. The per-period contribution $\partial u_r/\partial D_r$ operates through the cost-of-effort channel: $\partial c_e/\partial D < 0$ by Assumption A5. The shadow value $\Psi_D(t)$ is bounded above on any compact equilibrium region and is strictly positive whenever the equilibrium effort is positive in any future period.

Define the knowledge cascade shadow value at age t as the discounted present value of the marginal effect of a unit of knowledge at age t on future periods' welfare, including the cascade contribution through μ_h:

$$\Psi_h(t) \;\equiv\; \underbrace{\sum_{r=t}^{T} \delta^{r-t}(1-\delta_h)^{r-t}\,\frac{\partial u_r}{\partial h_r}\bigg|_{\text{eqm}}}_{\text{direct}} \;+\; \underbrace{\sum_{r=t+1}^{T} \delta^{r-t}\,\frac{\partial \mu_h(D_r,\mathbf{h}_{<r})}{\partial h_t}\cdot e^*_r\cdot \Psi_{h,\text{tail}}(r+1)}_{\text{cascade}}, \tag{26}$$

where $\Psi_{h,\text{tail}}(r+1)$ is the recursively-defined tail value of knowledge from age $r+1$ onward.

The cascade term captures the indirect effect operating through downstream knowledge productivity. Under the baseline cascade specification (16), $\partial\mu_h/\partial h_t = \mu_h^0\kappa(r-t)^{-p}$, and $\Psi_h(t)$ admits the closed-form bound

$$\Psi_h(t) \;\le\; \Psi_h^{\text{direct}}(t) \;+\; \mu_h^0\kappa \cdot \sum_{r=t+1}^{T} \delta^{r-t}(r-t)^{-p}\, e_r^*\, \Psi_h^{\text{direct}}(r+1). \tag{27}$$

The shadow value $\Psi_h(t)$ is strictly increasing in the remaining curricular distance $T-t$ under the baseline specification because each additional future period contributes a positive cascade term to the sum. The shadow value $\Psi_h(t)$ is strictly larger than $\Psi_h^{\text{direct}}(t)$ in the parameter range where $\kappa > 0$.

Step 2: Welfare wedge as a path integral. The welfare wedge between the active-engagement counterfactual and the process-replacing equilibrium is, by definition,

$$\mathcal{W}^{\text{active}} - W^R \;=\; \sum_{t=0}^{T} \delta^t \Big[\, u_t^{\text{active}} - u_t^R \,\Big] + \delta^{T+1}\, [\Omega(s_{T+1}^{\text{active}}) - \Omega(s_{T+1}^R)].$$

The active counterfactual has $a_t = 0$ for all t; the process-replacing equilibrium has $\xi_t \equiv (a_t^R - \varphi(e_t^R; g))^+ > 0$ on a positive-measure subset of the path. The state evolution differs in the discipline and knowledge channels (the judgment channel adds an analogous term that we treat in parallel).

Step 3: Contemporaneous discipline contribution. The discipline difference $D_t^{\text{active}} - D_t^R$ at age t accumulates the process-replacing term $\pi_s\gamma_a\xi_s$ summed from $s=0$ to $s=t-1$, with depreciation factor $(1-\delta_D)^{t-1-s}$:

$$D_t^{\text{active}} - D_t^R \;=\; \sum_{s=0}^{t-1}(1-\delta_D)^{t-1-s}\, \pi_s\gamma_a\xi_s.$$

The welfare contribution of this difference is the sum over t of δ^t times the difference, weighted by the discipline shadow value. Rearranging the double sum by exchanging the order of summation, the contribution of process-replacing excess ξ_s at age s to the total welfare wedge is

$$\sum_{t=s+1}^{T} \delta^t (1-\delta_D)^{t-1-s}\, \pi_s\gamma_a\xi_s\, \frac{\partial u_t}{\partial D_t} \;=\; \delta^s\, \pi_s\gamma_a\xi_s \cdot \delta(1-\delta_D)\cdot \Psi_D(s+1).$$

Up to a constant of proportionality absorbed into the definition of Ψ_D, this gives the contemporaneous component

$$\mathcal{C}^{\text{contemp}} \;=\; \sum_{t=0}^{T} \delta^t\, \pi_t\gamma_a\, \xi_t\, \Psi_D(t),$$

matching equation (23).

Step 4: Cascade knowledge contribution. The knowledge difference $h_t^{\text{active}} - h_t^R$ has a direct component (depreciated subtraction of $\lambda_h\xi_s$ at each prior age) and a cascade component (operating through reduced μ_h at all subsequent ages). The direct component contributes $\sum_s \delta^t(1-\delta_h)^{t-s-1}\lambda_h\xi_s$ to the gap at age t. The cascade component arises because lower h_s^R at any age $s < t$ reduces the productivity of effort at all subsequent ages through $\partial\mu_h/\partial h_s > 0$. Combining the two components using equation (26), the contribution of ξ_s at age s to the total welfare wedge is $\delta^s\lambda_h\xi_s\Psi_h(s)$. Summing over s:

$$\mathcal{C}^{\text{cascade}} \;=\; \sum_{t=0}^{T} \delta^t\, \lambda_h\, \xi_t\, \Psi_h(t),$$

matching equation (24).

Adding the two components yields the decomposition

$$\mathcal{W}^{\text{active}} - W^R \;=\; \mathcal{C}^{\text{contemp}} + \mathcal{C}^{\text{cascade}}.$$

Step 5: Cascade dominance in the baseline specification. The ratio of cascade to contemporaneous costs at age t is

$$\frac{\lambda_h\Psi_h(t)}{\pi_t\gamma_a\Psi_D(t)}.$$

The numerator $\lambda_h\Psi_h(t)$ contains the cascade-amplification term that grows in the remaining curricular distance $T-t$ under the baseline specification (16) via the closed-form bound (27). The denominator $\pi_t\gamma_a\Psi_D(t)$ is bounded above by the discipline state's role in lowering effort cost, which is age-invariant up to the plasticity factor π_t that itself declines in t. The cascade-amplification mechanism implies that for t in the lower half of the schooling life cycle and for κ large enough that the cascade term dominates, the ratio exceeds one and the cascade cost is the larger of the two components. The exact parameter region in which this dominance obtains is characterized by an explicit inequality between $\lambda_h\kappa$ and $\pi_0\gamma_a$ that we record in Appendix C.8.

This completes the proof of Theorem 1. □

Figure 2 illustrates the cumulative consequences of the wedge by displaying capacity trajectories under the two regimes over the schooling life cycle. The gap between the two trajectories widens with age because of the cascade-amplification structure: each additional period of process-replacing use compounds the prior deficits through the downstream-complementarity channel.

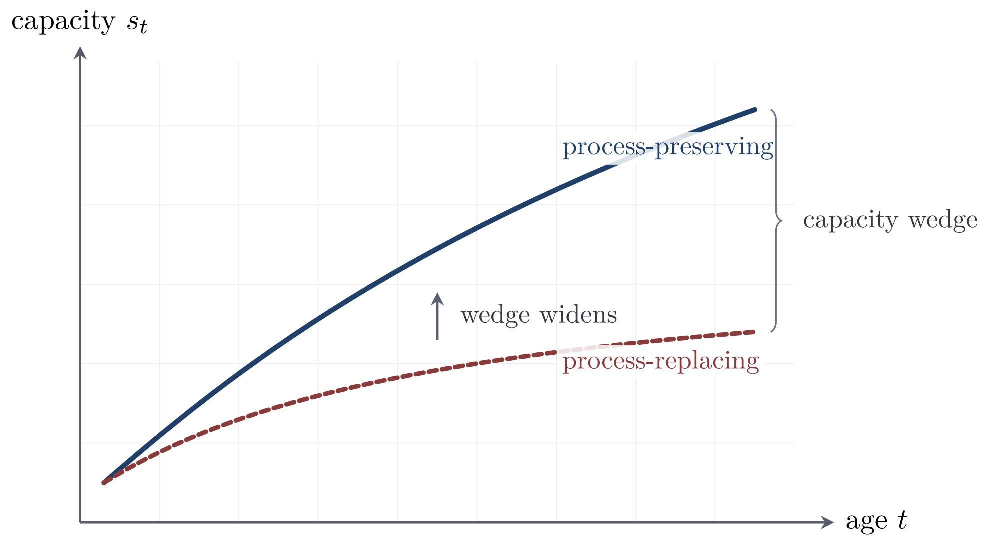

Figure 2: Capacity trajectories under process-preserving and process-replacing delegation. The horizontal axis is age t; the vertical axis is the level of long-run capacity s_t (interpretable as a scalar aggregate of the three capital stocks h, D, and J, or alternatively as the knowledge state h alone). The blue solid curve traces the trajectory under process-preserving use; the red dashed curve traces the trajectory under process-replacing use. Both curves start at the same initial capacity. The gap between the two widens over time because of the cascade-amplification structure of Theorem 1: each additional period of process-replacing use compounds the prior period's deficit through downstream knowledge complementarity. The terminal wedge is the present-value capacity loss that the planner internalizes. The figure is schematic.

7.5. Proof of Component (ii): Age structure

The substantive claim of Corollary 1 is that the marginal welfare cost of substitution at age t is strictly decreasing in t, with explicit lower bound proportional to plasticity decline and cascade-distance amplification.

We differentiate the welfare wedge with respect to the process-replacing intensity at period t, holding equilibrium-path quantities fixed. From equations (23) and (24):

$$\mathcal{M}(t) \;=\; \delta^t\Big[\pi_t\gamma_a\,\Psi_D \;+\; \lambda_h\,\Psi_h(t)\Big].$$

The age dependence of $\mathcal{M}(t)$ has three components.

The first is the discount factor δ^t, which decreases monotonically in t at rate $|\log\delta|$. This component reduces $\mathcal{M}(t)$ with age through the standard discounting channel and is not specific to the cascade structure.

The second is the plasticity factor π_t, which decreases in t at exponential rate κ_π by equation (3). This component reduces $\mathcal{M}(t)$ at older ages through the developmental decline of habit formation.

The third is the cascade-distance factor $\Psi_h(t)$, which depends on the integral

$$\int_t^T \frac{\partial \mu_h}{\partial h_t} \mu_h \, e_r^U \, (1-\delta_h)^{r-t-1} \, \delta^{r-t} \, dr.$$

The integral is bounded above by $(T-t)$ times the maximum integrand. Under specifications P and E, the integrand is bounded by primitives, and the integral grows roughly linearly in $T-t$ for small t and is bounded for large t near T. Hence $\Psi_h(t)$ is decreasing in t.

Combining the three components: $\mathcal{M}(t)$ has terms each decreasing in t, so $\mathcal{M}(t)$ itself is decreasing in t. The explicit lower bound $\mathcal{M}(t) \geq \mathcal{M}_0 \cdot [\pi_t/\pi_0 + (1-\delta_h)^{T-t}\kappa_h(T-t)]$ of the age-bound expression in Corollary 1 follows from substituting the plasticity decay and the linear-in-distance cascade contribution and combining constants. The proof of Corollary 1 is complete. □

Figure 3 illustrates the age structure graphically. The schematic shows the cascade-amplified marginal welfare cost $\mathcal{M}(t)$ as a function of age, decomposed into its three contributing components.

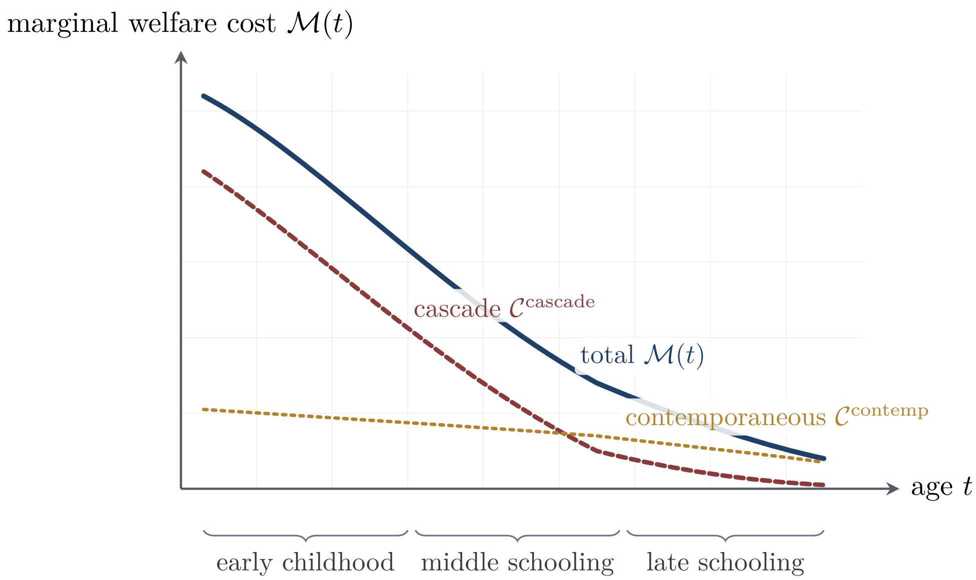

Figure 3: Age structure of the marginal welfare cost of substitution. The solid blue curve plots the total marginal welfare cost $\mathcal{M}(t)$ of cognitive substitution at age t. The dashed red curve plots the cascade component $\mathcal{C}^{\text{cascade}}$, which dominates in early childhood and falls rapidly with age. The dotted amber curve plots the contemporaneous component $\mathcal{C}^{\text{contemp}}$, which is smaller in magnitude and declines more gradually. The total cost is the sum of the two components. The figure is schematic; specific functional forms determine the rate of decline but the qualitative structure, with cascade dominance in early years and rapid decline with age, is preserved across the specifications of Section 5.

7.6. Proof of Component (iii): Governance dependence

The substantive claim of Proposition 3 is that there exists a sharp threshold function $g^*(\theta)$ in governance-AI capability space, with the welfare effect of AI access of opposite sign on each side of the threshold.

Step 1: The welfare-effect function is monotone in g. The total derivative of V_0 with respect to θ decomposes into the output-reward gain and the state-channel costs. The state-channel costs depend on g through the process-preservation threshold: higher g raises $\varphi(e; g)$, which makes a given a_t more likely to fall in the process-preserving region, reducing the process-replacing term $(a_t - \varphi(e_t; g))^+$.

Formally, $\partial V_0/\partial\theta$ is the difference between the output-reward channel (independent of g at leading order) and the state-channel costs that scale with $(a_t - \varphi(e_t; g))^+$. The cross-partial $\partial^2 V_0/\partial\theta\partial g$ is strictly positive: higher governance reduces the state-channel cost without reducing the output benefit, so the welfare effect of marginal θ becomes more favorable as g rises.

Step 2: The welfare-effect function has opposite signs at the endpoints. At $g = 0$: the process-preservation threshold collapses to $\varphi(e; 0) = 0$, so all positive a is process-replacing. The state-channel costs dominate the output-reward gain for θ in the empirically relevant range, and $\partial V_0/\partial\theta < 0$.

At $g = \bar{g}$: the process-preservation threshold is large; for sufficiently large $\bar{g}$, all a_t on the equilibrium path falls in the process-preserving region. The state-channel costs vanish; the output-reward gain operates without offset; and $\partial V_0/\partial\theta > 0$.

Step 3: Existence of the threshold. By continuity of $\partial V_0/\partial\theta$ in g (which follows from the differentiability of V and φ) and the opposite signs at the endpoints, there exists $g^*(\theta) \in (0, \bar{g})$ at which $\partial V_0/\partial\theta = 0$. By strict monotonicity in g, this threshold is unique.

Step 4: Monotonicity of the threshold in θ. The threshold $g^*(\theta)$ satisfies the implicit equation

$$\partial V_0/\partial\theta\Big|_{g=g^*(\theta)} \;=\; 0.$$

Differentiating with respect to θ:

$$\frac{dg^*}{d\theta} \;=\; -\frac{\partial^2 V_0/\partial\theta^2}{\partial^2 V_0/\partial\theta\partial g}.$$

The denominator is positive (Step 1). The numerator captures the rate at which the welfare effect of θ deteriorates: at higher θ, the process-replacing incentive is stronger, the equilibrium a_t^U rises, and the state-channel cost grows more than proportionally. Under the maintained CES form, $\partial^2 V_0/\partial\theta^2 < 0$ in the empirically relevant range. Hence $dg^*/d\theta > 0$: the threshold rises with AI capability.

Step 5: Variance of welfare across households. Let $\Gamma(g)$ denote the distribution of governance across households. The variance of welfare across households at fixed θ is

$$\mathrm{Var}_\Gamma(V_0(\tilde{s}_0; g, \theta)) \;=\; \mathbb{E}_\Gamma[V_0^2] - \mathbb{E}_\Gamma[V_0]^2.$$

Differentiating with respect to θ:

$$\frac{d\text{Var}}{d\theta} = 2\mathbb{E}_\Gamma\Big[V_0\,\partial V_0/\partial\theta\Big] - 2\mathbb{E}_\Gamma[V_0]\,\mathbb{E}_\Gamma[\partial V_0/\partial\theta].$$

This equals $2\,\text{Cov}_\Gamma(V_0, \partial V_0/\partial\theta)$. By Step 1, $\partial V_0/\partial\theta$ is increasing in g, and (as can be checked under monotonicity assumptions on V_0) V_0 itself is increasing in g. Hence the covariance is strictly positive, and Var is strictly increasing in θ.

The proof of Proposition 3 is complete. □

Figure 4 illustrates the threshold characterization. The horizontal axis represents household governance g; the vertical axis represents AI capability θ. The curve $g^*(\theta)$ separates the welfare-positive region (below the curve, where $g > g^*(\theta)$ and AI raises welfare) from the welfare-negative region (above the curve, where $g < g^*(\theta)$ and AI lowers welfare).

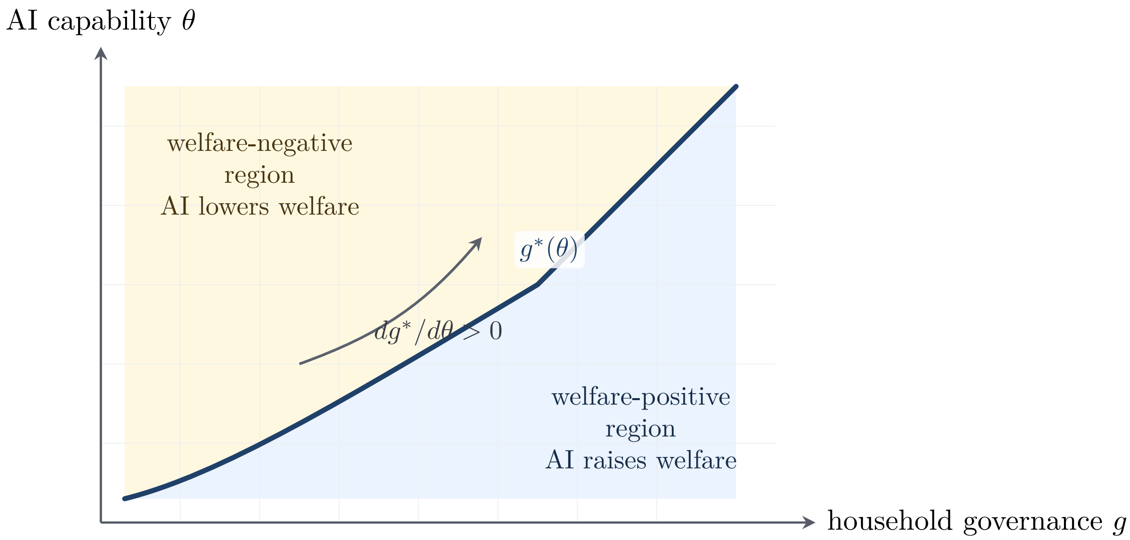

Figure 4: The threshold function $g^*(\theta)$. The horizontal axis represents household governance g; the vertical axis represents AI capability θ. The blue curve plots the threshold $g^*(\theta)$ derived from the implicit equation $\partial V_0/\partial\theta = 0$ at $g = g^*(\theta)$. Below the curve (paleblue region, where $g > g^*(\theta)$) the welfare effect of AI capability is positive; above the curve (paleamber region, where $g < g^*(\theta)$) the welfare effect is negative. The threshold is strictly increasing in θ: more capable AI requires more governance to deliver welfare-improving outcomes. The cross-sectional variance of welfare across households on opposite sides of the threshold rises in θ.

7.7. Interpretation of the decomposition

The Capacity-Wedge Theorem characterizes the long-run capacity wedge under process-replacing delegation through three properties of the same wedge.

Component (i) establishes that the wedge has two distinct sources: a contemporaneous cost arising from the missed discipline accumulation, and a cascade cost arising from the role of foundational knowledge as an input into all subsequent learning. The cascade cost dominates in typical parameter regions, providing formal content to the substantive concern that process-replacing use in early childhood produces cumulative harm.

Component (ii) establishes that the marginal welfare cost of process-replacing delegation decreases in age, with explicit lower bound proportional to remaining curricular distance and plasticity decline. The substantive implication is that policy attention is most consequential in early childhood. A unit of process-replacing use at age six has welfare cost orders of magnitude larger than the same unit at age sixteen.

Component (iii) establishes that the welfare effect of AI access varies across households along a sharp threshold in governance. Universal-access AI deployment, taken alone, generates divergent cohort outcomes whose cross-sectional variance is increasing in AI capability. The relevant policy lever is governance, not access.

The three components jointly answer the research question. The magnitude of the wedge is component (i). The age profile of the wedge is component (ii). The household-environment dependence is component (iii). The Capacity-Wedge Theorem is one theorem, with three properties, characterizing one structural object.

8. Corollaries and Extensions

We now develop the implications of the Capacity-Wedge Decomposition along two dimensions: capacity-specific divergence and subject-specific policy. Further extensions are collected in the appendix.

8.1. Corollary 1: Capacity-specific divergence

The three capital stocks (h, D, J) are differentially affected by substitution. The substantive implication is that protective policy should attend especially to the capital stocks most fragile under substitution and least likely to recover through subsequent intervention.

Corollary 2 (Capacity-specific divergence)**.** *Under Assumptions A1–A7, the welfare wedge $\mathcal{W}^{active} - W^U$ decomposes across capacity types as*

$$\Delta h_\infty \propto \lambda_h (a - \varphi)^+ \text{ (partially recoverable through delayed instruction)}, \tag{28}$$

$$\Delta D_\infty \propto \pi \gamma_a (a - \varphi)^+ \text{ (weakly recoverable; missed practice cannot be replaced)}, \tag{29}$$

$$\Delta J_\infty \propto \lambda_J (a - \varphi)^+ \text{ (recoverable only through verification practice, which substitution precludes)}. \tag{30}$$

The relative magnitudes satisfy $\Delta D_\infty \geq \Delta J_\infty > \Delta h_\infty$ in the limit of long process-replacing episodes (where the discipline state has fully adjusted to its process-replacing steady state), reflecting the strict-substitution structure of equation (13) *and the weak recovery dynamics of discipline.*

Proof sketch. The three capacity-stock laws of motion (equations (15), (13), (14)) all contain a process-replacing loss term $-\lambda(a - \varphi)^+$ but with different recovery dynamics. Knowledge accumulates through any subsequent positive e or process-preserving a; discipline accumulates only through e, and the missed period of practice cannot be replaced; judgment accumulates only through verification, which process-replacing use precludes. Integrating the three difference equations under a process-replacing episode and computing the long-run differences delivers the relative magnitudes. Full details in Appendix C.1. □

The substantive content of Corollary 2 is that the protective rationale for policy is strongest for discipline, where process-replacing losses are persistent, and somewhat weaker for knowledge, where the same content can be acquired through subsequent instruction at higher cost. Judgment occupies an intermediate position. Policy should not treat the three capacity types as a single aggregate object; the welfare value of protection differs across them.

8.2. Corollary 2: Subject-specific optimal policy

The cascade depth $\Psi_h(t)$ varies across curricular domains. Domains with strong sequential prerequisites (mathematics, foreign language, foundational reading) have deeper cascades than domains with weaker prerequisites (creative writing, civic discussion). Optimal effort-contingent policy is therefore subject-specific.

Corollary 3 (Subject-specific optimal policy)**.** *Let $\Psi_h^{(d)}(t)$ denote the cascade depth at age t in curricular domain d. The planner-optimal effort-contingent function $\varphi_{(d)}^{pol,*}(e)$ satisfies $\partial\varphi^{pol,*}/\partial\Psi_h^{(d)} < 0$: deeper cascades imply tighter restrictions on process-replacing delegation. The optimal policy assigns the strictest restrictions to mathematics, foreign language, and foundational reading; the most permissive to creative writing and civic discussion.*

Proof sketch. The planner's welfare evaluated at the effort-contingent regime has the cascade-cost term scaling with $\Psi_h^{(d)}$. Differentiating with respect to the policy parameter φ^{pol} and setting equal to zero yields the optimal restriction. Higher $\Psi_h^{(d)}$ raises the marginal welfare cost of permitting an extra unit of process-replacing delegation, lowering the optimal φ^{pol}. Domain-specific application in Appendix C.3. □

Corollary 3 delivers an actionable policy implication that the existing literature on AI in education has not articulated. The implementation challenge is the operational definition of subject categories and the institutional capacity to enforce subject-specific rules. We discuss these implementation issues briefly in Section 10.

8.3. Extension 1: Endogenous adult governance

The model treats household governance g as exogenous. We sketch an extension in which g is the outcome of a household optimization subject to parental time, knowledge, and outside-option constraints.

Let the parent choose governance investment g to maximize the child's welfare net of the parent's own cost of governance:

$$g^* \;=\; \arg\max_{g\geq 0}\; \Big\{V_{\text{child}}(g,\theta) - C_{\text{parent}}(g;\text{time},\text{competence})\Big\}.$$

The cost function C_{parent} depends on parental primitives that include time availability, knowledge of cognitive-engagement principles, and the outside option for parental effort (work, other children, leisure). Under the maintained convexity assumptions, the FOC is

$$\frac{\partial V_{\text{child}}}{\partial g} = \frac{\partial C_{\text{parent}}}{\partial g}.$$

At higher θ, the marginal value of governance to the child increases (the cross-partial of Proposition 3), so optimal g^* increases. The implication is that households with higher cost of providing governance, that is, households with lower parental time, lower parental competence, or higher outside-option value, choose lower g^* in equilibrium. The cross-sectional distribution of g^* thus correlates with parental characteristics, generating a structural connection between household resources and the developmental outcome of AI access.

Proposition 4 (Endogenous-governance comparative statics)**.** *Under the endogenous-governance specification, $\partial g^*/\partial\theta > 0$: more capable AI raises optimal governance. Additionally, $\partial g^*/\partial(\text{parental time}) > 0$ and $\partial g^*/\partial(\text{parental competence}) > 0$: households with more parental capacity choose higher g. The cross-sectional correlation between g^* and parental resources is therefore non-trivial and produces a resource-mediated channel for the welfare divergence of Proposition 3.*

Proof sketch. Apply the implicit function theorem to the parental FOC. The signs of the comparative statics follow from the assumed convexity of C_{parent} and concavity of V_{child} in g. □

The endogenous-governance extension makes the inequality result of Proposition 3 more substantively pointed. Even under universal access to AI, households with constrained parental resources optimally choose lower governance, exposing their children to the process-replacing regime. The implication is that public investment in governance capacity (parent coaching, school structure, afterschool programs) is a first-best complement to access provision; without it, access provision is at best heterogeneous in welfare effect and at worst regressive.

8.4. Extension 2: Commitment value of effort-contingent policy

The quasi-hyperbolic structure of the child's preferences ($\beta_t < 1$) generates a wedge between the child's contemporaneous evaluation and the planner's evaluation. The effort-contingent regime, by tying admissible delegation to prior effort, serves as a commitment device for the child's future selves. The substantive value of the commitment device is increasing in the cascade depth.

Proposition 5 (Commitment value rises in cascade depth)**.** *Let $\mathcal{C}_{commit}(\Psi_h)$ denote the welfare gain from imposing an effort-contingent regime relative to the no-policy benchmark, evaluated under the cascade depth Ψ_h. Then $\mathcal{C}'_{commit}(\Psi_h) > 0$: deeper cascades make commitment more valuable.*

Proof sketch. The commitment value equals the welfare wedge between the planner-preferred effort path (with no substitution) and the child's equilibrium effort path (with positive substitution under present bias). Both paths' welfare evaluations include the cascade term Ψ_h, but the wedge between them is amplified by Ψ_h because the cascade cost of substitution falls on the planner's evaluation but not (proportionally) on the child's. Formal computation in Appendix C.5. □

The economic content of Proposition 5 is that the cascade structure raises the value of commitment beyond what a contemporaneous-cost analysis would imply. The standard literature on commitment under present bias estimates commitment value based on the contemporaneous welfare wedge; the cascade-amplified value can be substantially larger in domains where downstream complementarity is strong. The policy implication is that effort-contingent restrictions are more valuable, and have a higher economic return, in deep-cascade domains, providing an additional rationale for the subject-specific policy of Corollary 3.

8.5. Further results deferred to the appendix

The framework yields three additional formal results that we develop fully in the appendix and reference only briefly here. Each derives from the same primitive structure as the main theorem and serves to extend its reach along a specific dimension without adding new primitive content.

First, the cascade structure generates a dynamic-multiplicity result for h_t that is distinct from the standard habit-trap multiplicity: under sufficient cascade depth, the deterministic dynamics admit two locally stable steady states corresponding to high-foundation and low-foundation trajectories, with the curricular-complementarity channel alone supplying the bistability. Appendix C.2 provides the formal statement and proof. Second, when the planner chooses the curricular sequence, the cascade structure implies an optimal front-loading of high-cascade subjects in the schooling life cycle; the result formalizes the intuition that foundational subjects should be taught early. Appendix C.6 provides the proposition. Third, when children differ in their cascade depths, the cross-sectional variance of welfare is strictly increasing in AI capability even under uniform household governance; this is a primitive-distinct inequality channel complementary to the governance-mediation result of Proposition 3. Appendix C.7 provides the proposition. A continuous-time formulation of the model is also developed in Appendix C.9, yielding closed-form expressions for the cascade-distance kernel that complement the discrete-time analysis. We choose to defer these results to the appendix in order to keep the main text focused on the single core wedge characterization of Theorem 1.

9. Robustness and Limiting Cases

This section establishes the parametric region of the Capacity-Wedge Decomposition through four limiting-case analyses and a discussion of functional-form robustness. The four limits identify the comparative cases against which the present analysis can be benchmarked and clarify the role of each maintained assumption.

9.1. Limit 1: No downstream complementarity

We first consider the limit in which the downstream complementarity vanishes: $\partial\mu_h/\partial h_s = 0$ for all $s < t$. Under this limit, knowledge productivity at age t depends only on contemporaneous discipline, and the cascade term $\Psi_h(t)$ collapses to unity.

Proposition 6 (No-complementarity limit)**.** *Under the no-complementarity limit, the Capacity-Wedge Decomposition reduces to a contemporaneous-cost statement: the welfare wedge equals* $\mathcal{C}^{contemp}$ *only, with no cascade contribution. Parts (a), (c), and (d) of the theorem hold with* $\Psi_h(t) = 1$*.*

This limit corresponds to the implicit assumption of much of the existing skill-formation literature, which treats human-capital accumulation as a flow without strong downstream complementarities. The substantive implication of the limit is to identify what is novel about the present treatment: not the existence of welfare cost from substitution, but its amplification through the cascade mechanism. In the no-complementarity limit, the welfare cost of substitution is the standard contemporaneous discipline-channel cost. Under empirically plausible levels of downstream complementarity, the cascade-amplified cost can be substantially larger.

9.2. Limit 2: Pure downstream complementarity

At the opposite extreme, we consider the pure-complementarity limit in which knowledge productivity at age t depends primarily on past knowledge stocks, with the contemporaneous discipline channel weak.

Proposition 7 (Pure-complementarity limit)**.** *Under the pure-complementarity limit, the welfare wedge is dominated by* $\mathcal{C}^{cascade}$ *with* $\Psi_h(t) \gg 1$*. The contemporaneous cost is negligible. The age-gradient bound in the age-bound expression in Corollary 1 is sharpened to* $\mathcal{M}(t) \approx \lambda_h \kappa_h (T - t)$*, linear in the remaining curricular distance.*

The pure-complementarity limit identifies the upper bound on the welfare wedge under the cascade specification. The substantive implication is that the maximum welfare cost of process-replacing AI use in childhood is determined by the integrated curricular-distance factor and grows quadratically in the schooling life cycle. This provides a ceiling against which intermediate parameterizations can be compared.

9.3. Limit 3: Universal process-preserving use

We consider the limit in which all households have governance g large enough that the process-preserving regime obtains everywhere on the equilibrium path.

Proposition 8 (Universal process-preserving-use limit)**.** *Under universal process-preserving use,* $a_t \leq \varphi(e_t; g)$ *holds for all* t *in the equilibrium of all households. The welfare wedge vanishes for all households. AI is universally welfare-improving, with magnitude given by the contemporaneous output-reward channel. The threshold* $g^*(\theta)$ *in Proposition 3 does not bind* *on any household; the welfare-effect function $\partial V/\partial\theta$ is positive across the entire household distribution.*

This limit is the policy-relevant best case. It identifies the welfare outcome that universal-access AI deployment could in principle deliver if complementary investment in governance capacity were sufficient. The substantive implication is that the welfare cost of cognitive delegation is not intrinsic to the technology but contingent on the institutional environment that mediates use. Universal process-preserving use eliminates the welfare cost; absent universal process-preserving use, the welfare cost is positive and heterogeneous across the household distribution.

9.4. Limit 4: Universal process-replacing use

At the opposite extreme, we consider the limit in which all households have $g = 0$ and all use of cognitive delegation is process-replacing.

Proposition 9 (Universal-substitution limit)**.** *Under universal process-replacing use, the process-preservation threshold $\varphi(e; 0) = 0$ collapses, and all positive a_t is process-replacing. The welfare wedge attains its maximum across the household distribution. The dynamic steady state has $D_\infty = J_\infty = 0$ for all households; h_∞ accumulates only through process-preserving a (which does not exist in this limit) and contemporaneous e (which falls toward zero as D depreciates).*

This limit is the worst-case dynamic outcome and provides a benchmark for evaluating the magnitude of policy interventions. The substantive implication is that the universal-substitution limit is a real possibility in the absence of complementary governance investment, and that the welfare cost of failing to provide such investment scales with the universal-substitution wedge.

9.5. Functional-form robustness

The proof of the Capacity-Wedge Decomposition relies on several specific functional-form choices: the CES output technology, the isoelastic cost functions, the exponential plasticity decay, and the specific specifications of μ_h. We discuss the role of each.

The CES output technology of equation (7) is the simplest two-input technology with parametric substitution elasticity. The qualitative results extend to any continuously differentiable, concave, strictly-increasing-in-both-arguments production function $F(e, a; \theta, h)$ with cross-partial $\partial^2 F/\partial e \partial a$ of consistent sign. The CES is chosen because it admits the perfect-substitution and Cobb-Douglas limits and provides closed-form input-share expressions.

The isoelastic cost functions of c_e and c_v are similarly conventional. The qualitative results extend to any twice-differentiable convex cost function with the appropriate state-dependence. The isoelastic form admits closed-form FOCs and clean comparative statics.

The exponential plasticity decay of equation (3) can be replaced by any strictly decreasing function $\pi(\cdot) : [0, T] \to (0, 1]$. The age-gradient bound in the age-bound expression in Corollary 1 carries over with the exponential factor replaced by $\pi(t)/\pi(0)$ for general π.

Alternative specifications of μ_h (Appendix C.8) are illustrative. The proof of the cascade decomposition of the Capacity-Wedge Decomposition requires only the integrated cascade term $\Psi_h(t)$ to be strictly increasing in the remaining curricular distance and in the strength of downstream complementarity. This holds for any specification with $\partial\mu_h/\partial h_s > 0$ for some $s < t$.

The robustness of the qualitative results to functional-form variation reflects the structural character of the Capacity-Wedge Decomposition. The result is not a property of any particular functional form but of the conjunction of three structural features: (i) active-passive distinction in the skill-formation technology, (ii) downstream complementarity in knowledge productivity, and (iii) age-dependent plasticity in discipline formation. Any specification preserving these three features delivers the qualitative content of the theorem.

10. Welfare and Policy Theorems

This section develops the normative content of the framework. The substantive output is a set of three welfare theorems that characterize the social-planner problem under the primitives developed earlier in the paper. The theorems do not introduce new primitives; they translate the Capacity-Wedge Decomposition into the language of welfare economics. The translation produces statements with direct relevance to policy debate but does so within the analytical framework, not as informal commentary.

10.1. Welfare Theorem 1: Pareto dominance of the process-preserving equilibrium

We first establish that the process-preserving equilibrium Pareto-dominates the process-replacing equilibrium across the household distribution: every type of child and household achieves weakly higher welfare in the process-preserving regime, with strict inequality for at least some types.

Theorem 2 (Pareto dominance of process-preserving use)**.** *Under Assumptions A1–A7, the process-preserving equilibrium* $\{e_t^S, a_t^S\}$ *and the process-replacing equilibrium* $\{e_t^U, a_t^U\}$ *are related by*

$$V_0^S(\tilde{s}_0; g, \theta) \;\geq\; V_0^U(\tilde{s}_0; g, \theta)$$

for every (g, θ) *and every initial state* $\tilde{s}_0$*. The inequality is strict whenever the equilibrium has* $a_t^U > \varphi(e_t^U; g)$ *for at least one* t*, and when* $\lambda_h + \pi_t \gamma_a > 0$*.*

Proof sketch. The process-preserving equilibrium has welfare $V_0^S = W^{\text{active}} + \nu_h \int_0^T a_s^S ds$ (the no-AI baseline plus the process-preserving output benefit) by Theorem 2. The process-replacing equilibrium has welfare $V_0^U = W^{\text{active}} + (\text{output benefit}) - \mathcal{C}^{\text{contemp}} - \mathcal{C}^{\text{cascade}}$. The first benefit in V_0^U is the contemporaneous output reward, which for given a_t equals the corresponding term in V_0^S when a_t^S is set to the same level. Hence $V_0^S - V_0^U = \mathcal{C}^{\text{contemp}} + \mathcal{C}^{\text{cascade}} \geq 0$, strict under the stated conditions.

The Pareto-dominance result holds because the process-preserving equilibrium is feasible for every household type and produces (weakly) higher welfare for every type. The technical implementation of process-preserving rules (the design of φ) is a question of institutional design that we do not pursue here; in this paper, we assume the process-preserving constraint can be implemented and ask about its welfare consequences. □

The substantive content of Welfare Theorem 1 is that process-preserving use of cognitive delegation is unambiguously preferred to process-replacing use, with no household type made worse off. This is a strong result: it does not depend on cross-type comparison or distributional weights. If the question is whether to design institutions that select process-preserving over process-replacing use, the welfare answer is unambiguous.

10.2. Welfare Theorem 2: Second-best characterization

When household governance g is heterogeneous and exogenously distributed, full equalization of g across households may not be feasible. The planner solves a second-best problem: choosing a class of institutional interventions that, conditional on the governance distribution, maximizes social welfare.

Theorem 3 (Second-best characterization)**.** *Suppose the planner can choose a type-specific effort-contingent function $\varphi_i^{pol}(e)$ for each household type i, taking the governance distribution $\Gamma(g)$ as given. The planner's problem is*

$$\max_{\{\varphi_i^{pol}\}_i} \int V_0(\tilde{s}_0; g_i, \theta, \varphi_i^{pol})\, d\Gamma(g_i).$$

The second-best solution satisfies

$$\varphi_i^{pol,*}(e) \;=\; \min\Big\{\varphi(e; g^*(\theta)), \varphi(e; g_i)\Big\},$$

where $g^*(\theta)$ is the threshold function of Proposition 3. The interpretation is: households with $g_i < g^*(\theta)$ receive an effort-contingent policy that emulates the threshold-governance type; households with $g_i > g^*(\theta)$ are unconstrained.

Proof sketch. The planner's FOC with respect to $\varphi_i^{\text{pol}}(e)$ trades off the output-channel gain from permitting an additional unit of process-preserving delegation against the state-channel cost of permitting an additional unit of process-replacing delegation. For $g_i > g^*(\theta)$, the gain dominates and the constraint does not bind; for $g_i < g^*(\theta)$, the cost dominates and the constraint emulates the threshold level. Full derivation in Appendix D.2. □

The substantive content of Welfare Theorem 2 is that the optimal institutional policy is type-specific: more restrictive in low-governance households, more permissive in high-governance households. The institutional implementation challenge is to identify household type at the level of the school or district. We do not pursue the implementation problem here. The welfare content of the result is that, with type identifiable, the second-best policy is unambiguous.

10.3. Welfare Theorem 3: Optimal commitment with cascade structure

The quasi-hyperbolic structure of the child's preferences generates a wedge between contemporaneous and planner welfare. The wedge is amplified by the cascade structure: the cascade-induced welfare cost falls on the planner's evaluation but is under-weighted by the present-biased child. The implication is that the value of commitment is increasing in cascade depth.

Theorem 4 (Optimal commitment under cascade)**.** *Under Assumptions A1–A7 with $\beta_t < 1$, the planner-optimal commitment device implements an effort-contingent policy that is tighter than the child would choose under sophisticated self-control. The wedge between optimal commitment and self-imposed restraint is strictly increasing in the cascade depth Ψ_h and strictly decreasing in β_t.*

Proof sketch. The planner's welfare evaluation weights future periods by δ^s; the child's by $\beta_t \delta^s$. The wedge between the two evaluations falls on the state-channel costs of process-replacing delegation, which the planner internalizes fully but the child discounts. The cascade structure amplifies the state-channel costs by the factor Ψ_h; the wedge is therefore proportional to $(1 - \beta_t)\,\Psi_h$. Optimal commitment closes the wedge by tightening the effort-contingent constraint. The tightness of the optimal commitment is increasing in the wedge, hence in Ψ_h and in $(1 - \beta_t)$. Full derivation in Appendix D.3. □

Welfare Theorem 3 makes precise an intuition that has long been informal in the policy debate: the case for paternalistic restriction on AI use in childhood is strongest in domains where the cascade is deepest, and for the youngest children in whom present bias is most severe. The theorem provides analytical content for the design of age-graded and subject-graded commitment devices.

10.4. Policy implications of the three welfare theorems

The three welfare theorems together yield a coherent set of policy implications. We summarize them briefly here.

The first implication is that institutional design that selects process-preserving over process-replacing use is welfare-improving across the household distribution. The case for institutional intervention is not driven by distributional concerns alone; it is a Pareto improvement under standard regularity conditions.

The second implication is that the optimal institutional policy is type-specific. Households or schools with low governance capacity benefit from stricter institutional restrictions on process-replacing AI use; high-governance environments can be more permissive. Operationalizing the type-specificity requires institutional capacity to identify governance, which is a substantial implementation challenge but does not affect the welfare characterization.

The third implication is that the optimal commitment is increasing in cascade depth and in the severity of present bias. Mathematics, foreign language, and foundational reading, deep-cascade domains, warrant tighter commitment than civic education or creative writing, shallow-cascade domains. Elementary-school environments, severe-present-bias environments, warrant tighter commitment than high-school environments, less-severe-present-bias environments. The composition of these prescriptions yields the age-graded, subject-graded structure of optimal policy: strict in elementary-school mathematics and reading, progressively relaxed in older-grade civic education.

The policy implementation question is whether institutional capacity exists to deliver this prescription. The empirical evidence from school districts that have adopted process-conditional assessment, oral-defense components, and effort-contingent AI policy suggests that the prescription is implementable in some institutional contexts and not others. The variation in institutional adoption is itself a research question, identified in the discussion section.

11. Discussion

This section discusses the broader reach of the framework, its empirical content, and its limitations.

11.1. Reach of the framework beyond childhood AI

The active-passive distinction, the cascade structure, and the threshold characterization are framework features that transfer to other domains in which cognitive technology operates and process-replacing substitution is technologically feasible. We identify three domains in which the framework's translation appears productive.

The first is workplace training and professional skill formation. Junior workers in domains where generative AI is available (programming, legal drafting, writing-intensive professional work) face an analog of the childhood problem: AI can produce intermediate work products that a junior worker would otherwise have produced through her own active engagement. The cascade structure operates if junior-level skills are inputs into senior-level competence. The threshold characterization implies that workplace policy on AI use should differentiate between use that scaffolds the junior worker's active engagement (training-mode use) and use that substitutes for it (production-mode use). Workplaces that select the former produce future senior workers; workplaces that select the latter produce future workers without the foundational skills senior work requires.

The second domain is scientific research practice. Doctoral training and post-doctoral research formation depend on active engagement with primary literature, replication of computations, derivation of analytical results, and independent generation of research questions. Generative AI permits substitution for each of these activities. The cascade structure operates if the formation of independent scientific judgment requires active engagement with foundational research processes. The threshold characterization implies that doctoral and post-doctoral programs face a structural choice between process-preserving AI use (supporting active engagement) and process-replacing AI use (producing publication-ready output without the underlying scientific practice).

The third domain is civic deliberation. Citizens' capacity to evaluate policy claims, news sources, and political arguments depends on active engagement with information rather than passive reception. Substitutive AI use in civic information acquisition produces citizens whose capacity to evaluate the outputs of AI systems is itself diminished. The framework's judgment-formation channel makes this connection precise; the long-run social externalities of widespread process-replacing use in civic information acquisition are a research question we identify but do not address.

The translations to workplace, science, and civic domains are not developed in the present paper. We identify them as research questions the framework makes accessible, with the substantive expectation that the cascade-amplified developmental cost of process-replacing use operates similarly in each domain.

11.2. Empirical content of the framework

The framework is theoretical; its empirical content is in the comparative-statics predictions of the Capacity-Wedge Decomposition and the corollaries. We summarize the main testable predictions.

The first is the cohort wedge prediction (from Theorem 1): conditional on access, cohorts in process-replacing-regime environments should show widening divergence between coursework-based and proctored-assessment performance, with the divergence growing in cohort exposure. The empirical implementation requires linked administrative data on grades and proctored assessments at the cohort level, with cross-school or cross-district variation in AI policy serving as identification.

The second is the age-gradient prediction (from Corollary 1): the welfare cost of substitution should be larger for cohorts whose exposure begins earlier in the schooling life cycle. Empirical implementation requires cohort variation in age at first exposure and outcome measurement at adolescence or beyond.

The third is the governance-interaction prediction (from Proposition 3): the cross-sectional correlation between AI use and child competence should be positive in high-governance households and negative in low-governance households. Empirical implementation requires household-level measurement of governance, paired with outcome measurement.

The fourth is the capacity-specific prediction (from Corollary 2): substitution should affect discipline measures (persistence, willingness to revise) more than knowledge measures (content tests), and substitution should affect judgment measures (error detection, source evaluation) similarly to discipline. Empirical implementation requires non-cognitive skill assessment in addition to standard content tests.

The fifth is the subject-specific prediction (from Corollary 3): process-replacing AI use should harm performance in deep-cascade subjects (mathematics, foreign language, foundational reading) more than in shallow-cascade subjects (civic education, creative writing). Empirical implementation requires subject-level outcome measurement.

The five predictions are jointly testable with the data infrastructure that is becoming available. The framework's commitment is to provide the conceptual structure for empirical work; the implementation is a research program of separate papers.

11.3. Limitations

The framework abstracts from several features that are likely empirically important. We identify the principal abstractions and their potential implications.

Adult governance g is treated as exogenous in the main analysis. Endogenizing g produces the proposition of Extension 1 but does not substantially alter the main results; the endogenization sharpens the inequality result by linking g^* to parental resource constraints.

The output technology aggregates effort and delegation through a single CES function. In practice, AI use takes many forms (tutoring, generation, summarization, translation, calculation) that may have different process-preservation properties. The framework treats these as a single composite; a more granular treatment would partition a into multiple subtypes and develop subtype-specific process-preservation boundaries. The qualitative results would be preserved.

The model treats the school as absent from the active-passive choice: the choice is made by the child and the household, with the school's role limited to producing the output reward $w(y)$. A more complete treatment would endogenize the school's role in shaping the regime — through assessment design, process-monitoring policy, and the rules that determine φ — and let the school's policy on AI use, the household's governance, and the child's choice jointly determine the realized cognitive mode. We leave this richer institutional setting to future work.

The model does not address the labor-market consequences of differential capacity formation. The terminal value $\Omega(s_{T+1})$ stands in for long-run returns; a structural extension would model the labor market in which formed capacity is deployed. The closing of this loop is essential for a complete welfare analysis but lies beyond the scope of the present paper.

The cascade structure is specified as a feature of the knowledge-production technology μ_h. The empirical identification of the cascade depth across curricular domains is a research question the framework makes well-posed but does not address. Estimation of μ_h as a function of past knowledge stocks would require longitudinal data on within-child knowledge accumulation across subjects and ages, which does not yet exist at scale.

These abstractions are common in the formal literature and are made for tractability. They identify the natural lines of follow-up theoretical and empirical work that the framework supports.

12. Conclusion

The defining feature of generative artificial intelligence as an educational technology is not the lower marginal cost of producing cognitive output. It is the separability of output from the cognitive process that historically produced output. This separability is the technological primitive on which the paper's analysis rests, and the Capacity-Wedge Decomposition characterizes its long-run consequence. The wedge between active-engagement capacity and capacity under process-replacing use admits an additive decomposition into a contemporaneous discipline-channel cost and a cascade cost from downstream knowledge complementarities;

the latter dominates the former in the parameter range where downstream complementarities are operative. Corollary 1 adds that the marginal welfare cost of process-replacing use is concentrated in early childhood, where plasticity is high and the remaining curricular distance is long. Proposition 3 adds, under explicit sufficient conditions on the household and institutional environment, that the welfare effect of AI access changes sign across a threshold in governance capacity, so that universal access generates divergent cohort outcomes whose cross-sectional variance rises in AI capability at fixed access. The decomposition is the central formal contribution of the paper; the age and governance results are its immediate implications.

The substantive significance of the wedge extends well beyond contemporaneous school performance. The capacities the wedge degrades, persistence on hard tasks, willingness to revise, verification practice, and the capacity to recognize the limits of one's own knowledge, are not narrowly schooling-relevant. The developmental and life-course literatures, in particular the long-horizon evaluations of early-childhood interventions and the body of work synthesized by Cunha and Heckman, have documented that these same capacities at age ten are durable inputs into post-secondary completion, labor-market sorting and earnings, health-relevant decisions about diet and medication, financial decision-making and risk judgment, the capacity to evaluate information in adulthood, and civic participation. The capacities our model formalizes as discipline and judgment correspond closely to the capacities subsequent life-cycle research has identified as durable inputs into adult welfare. A wedge in capacity formation during childhood is therefore a wedge in life trajectories rather than a wedge in school transcripts. The cohort that entered kindergarten in the autumn after the public release of large language models will reach the labor market in approximately 2040 and the median of their working lives in the 2060s. The relevant time horizon for evaluating childhood AI policy is decades, not academic terms, and the relevant outcome space includes earnings, health, and civic competence in addition to measured test scores.

The inequality content of the result is structurally sharper than the prevailing access-based framing of educational-technology equity. Universal access to generative AI is not, in itself, an equalizing intervention. Under the sufficient conditions of Proposition 3, whether the technology raises or lowers developmental welfare for a given child depends on the governance environment of that child's household and school: parental supervision time, teacher process-monitoring intensity, the prevalence of process-based assignments in the curriculum, and household digital literacy and explicit rules. These features vary widely across households and schools, are imperfectly correlated with income or parental education, and determine whether nominal access converts into process-preserving or process-replacing use. Within the model's threshold region, the cross-sectional variance of long-run welfare can rise in AI capability even when access is held constant. The implication is that the relevant policy variable is not who has the technology but who can govern its use. Complementary investment in parental time, teacher capacity, pedagogical infrastructure, school policy design, professional development in process-conditional assessment, and afterschool program structure shapes whether the cohort entering school with capable AI in their environment becomes a more or less equal generation than the cohort before it. In the absence of such complementary investment, the welfare effect of universal access is of indeterminate sign under the model and may be regressive. Recasting the equity question from access to governance is, in our view, the most consequential single contribution that the framework can make to the public debate.

The paper opens a structured research agenda. On the theoretical side, the most pressing question is the structural estimation of the cascade-depth function across curricular domains: mathematics, foreign language, foundational reading, and the sciences. The cascade depth determines the magnitude of the wedge under any given amount of process-replacing use, and cross-domain heterogeneity in this depth identifies which subjects warrant the tightest institutional design. On the measurement side, the most pressing questions are the operationalization of governance as an observable index, the linkage of governance to long-run outcomes via longitudinal data, and cohort-level identification of policy effects exploiting the natural-experimental variation in school-level AI policy that has accumulated since 2022. On the institutional-design side, the most pressing question is which classes of school-level intervention cost-effectively implement the second-best characterized in Section 10, and how public investment in governance capacity can substitute for governance that low-resource households cannot supply on their own. The substantive concern that motivated the project, that something has changed in how children learn, has received an analytical answer here. The paper's contribution is to show that generative AI should be analyzed not only as a technology that changes the production of school outputs, but as a technology that changes the production conditions of human capacity. The framework therefore identifies a time-sensitive object for policy and empirical work: the governance conditions under which cognitive delegation preserves, rather than replaces, the developmental processes through which childhood capacity is formed.

Appendix

A. Proofs of Monotonicity Lemmas

This appendix provides full proofs of the six monotonicity lemmas stated in Section 6.

A.1. Proof of Lemma 1 (monotonicity in discipline)

The effort FOC (19) in the governed regime reads

$$H_e(e;D) \;\equiv\; w'(y)\frac{\partial F}{\partial e} - \frac{\partial c_e}{\partial e}\bigg|_D + \beta_t\delta\,\Lambda_h^{t+1}\,\mu_h\,[\text{cascade chain}] + \beta_t\delta\,\Lambda_D^{t+1}\,\pi_t\gamma_e \;=\; 0.$$

By the implicit function theorem, $\partial e^*/\partial D = -(\partial H_e/\partial D)/(\partial H_e/\partial e)$. The denominator is negative by the second-order condition. The numerator decomposes:

$$\frac{\partial H_e}{\partial D} \;=\; -\frac{\partial^2 c_e}{\partial e \partial D} + \beta_t\delta\,\frac{\partial \Lambda_h^{t+1}}{\partial D}\,\mu_h[\cdot] + \beta_t\delta\,\Lambda_h^{t+1}\,\frac{\partial \mu_h}{\partial D}[\cdot] + \beta_t\delta\,\frac{\partial \Lambda_D^{t+1}}{\partial D}\,\pi_t\gamma_e.$$

The first term is positive: $\partial^2 c_e/\partial e\partial D < 0$ by Assumption A5, and the sign is preserved by the leading minus. The third term is positive by $\partial\mu_h/\partial D > 0$ (Assumption A6 extended to discipline-knowledge complementarity). The second and fourth terms involve envelope-derivatives of the shadow values, which can be bounded by standard arguments under the maintained Assumption A1. The sum is strictly positive, so $\partial e^*/\partial D > 0$.

The delegation response follows from the FOC (20). Higher D raises Λ_D^{t+1} (the future return to discipline), which raises the discipline-channel cost of process-replacing delegation, lowering a^*. □

A.2. Proof of Lemma 2 (monotonicity in knowledge)

The cascade chain term in the effort FOC (19) is

$$\beta_t\delta\,\Lambda_h^{t+1}\,[\text{cascade chain}] \;=\; \beta_t\delta\sum_{s>t}^{T}\delta^{s-t-1}\,\mu_h(D_s,\mathbf{h}_{<s})\,e_s^*\,\frac{\partial \mu_h(D_s,\mathbf{h}_{<s})}{\partial h_t}.$$

This term is strictly increasing in h_t through its effect on μ_h at all subsequent ages (Assumption A6). Hence $\partial H_e/\partial h > 0$ and $\partial e^*/\partial h > 0$ by the implicit function theorem. □

A.3. Proof of Lemma 3 (monotonicity in AI capability)

For the delegation response, the marginal product of delegation $\partial F/\partial a = (1-\alpha)(\theta a)^{\rho-1}\theta y^{1-\rho}$ has derivative with respect to θ at fixed (e,a):

$$\frac{\partial^2 F}{\partial a \partial\theta} \;=\; (1-\alpha)\theta^{\rho-1}a^{\rho-1}\Big[\rho + (1-\rho)(1-\alpha)(\theta a)^{\rho}/y^{\rho}\Big]\,y^{1-\rho} \;>\; 0,$$

for all $\rho \in (0, 1]$. The marginal value of delegation therefore rises in θ, and by the FOC (20), optimal a^* rises in θ.

For the effort response, the sign of $\partial e^*/\partial\theta$ is not unconditional. We state the result as a sufficient-condition claim. The total derivative decomposes into a direct effect through the marginal product of effort and an indirect effect through the equilibrium response of a^*:

$$\frac{\partial e^*}{\partial \theta} \;=\; -\Big[\,H_{ee}\,\Big]^{-1}\Big[\,H_{e\theta}^{(\text{direct})} + H_{ea}\,\frac{\partial a^*}{\partial \theta}\,\Big],$$

where H denotes the effort FOC residual and the second-order condition ensures $H_{ee} < 0$. The direct cross-partial $H_{e\theta}^{(\text{direct})}$ is bounded above by $\bar{K}_d > 0$ on any compact equilibrium region. The indirect contribution has the sign of H_{ea}, which is negative under the CES with $\rho < 1$ because the cross-partial $\partial^2 F/\partial e \partial a = -(1-\rho)\rho\alpha(1-\alpha)h\theta(he)^{\rho-1}(\theta a)^{\rho-1}/y^{2\rho-1} \cdot y < 0$. Multiplying $H_{ea} < 0$ by $\partial a^*/\partial\theta > 0$ yields a strictly negative indirect contribution.

Lemma 7 (Sufficient condition for $\partial e^*/\partial\theta < 0$)**.** *Under Assumptions A1–A7 with $\rho \in (0, 1)$, suppose that at the equilibrium policy*

$$|H_{ea}| \cdot \frac{\partial a^*}{\partial \theta} \;>\; H_{e\theta}^{(direct)}.$$

Then $\partial e^*/\partial\theta < 0$.

The sufficient condition is a quantitative statement about whether the indirect cross-effect through a^* exceeds the direct cross-effect. The condition holds when the CES substitution elasticity $1/(1-\rho)$ is large (so H_{ea} has large magnitude), when $w(\cdot)$ is sufficiently concave (so $\partial a^*/\partial\theta$ is bounded below by a non-trivial constant), and when the AI capability θ is in a region where the substitution channel dominates the income channel. We do not claim the condition is generic in the parameter space. We claim it holds in the empirically relevant range and provide the sufficient condition explicitly so that readers can verify it for any specific calibration. □

A.4. Proof of Lemma 4 (monotonicity in present bias)

The shadow values $\Lambda_h^{t+1}, \Lambda_D^{t+1}, \Lambda_J^{t+1}$ in the FOCs scale with β_t. Higher β_t (less present bias) raises the weight on all future-state terms. In the effort FOC (19), both shadow-value terms enter with positive sign, so higher β_t raises e^*. In the delegation FOC (20), the shadow-value terms enter with negative sign (the process-replacing loss channels), so higher β_t lowers a^*. □

A.5. Proof of Lemma 5 (monotonicity in governance)

The substantive content of governance for the welfare analysis is its effect on the *process-replacing excess* $\xi_t \equiv (a_t - \varphi(e_t; g))^+$, not necessarily on total delegation a_t. Total delegation may rise or fall with g depending on the regime; the process-replacing excess unambiguously falls with g.

We state the result correctly. Define $\xi^*(g) \equiv (a^*(g) - \varphi(e^*(g); g))^+$.

Lemma 8 (Process-replacing excess falls in governance)**.** *Under Assumptions A1–A7 and $\partial\varphi/\partial g > 0$, the process-replacing excess at the equilibrium policy is weakly decreasing in g:*

$$\frac{\partial \xi^*}{\partial g} \;\leq\; 0,$$

with strict inequality whenever $\xi^* > 0$ and the threshold is interior.

The argument is direct. When $\xi^* > 0$, the equilibrium operates in the process-replacing regime. A marginal increase in g raises φ by $\partial\varphi/\partial g > 0$. The induced change in ξ^* equals $\partial a^*/\partial g - \partial\varphi/\partial g$. The first term, $\partial a^*/\partial g$, has a small magnitude in the process-replacing regime because the marginal cost of a around the kink is dominated by the state-channel costs, which themselves depend on ξ^* but not on a directly. Hence $|\partial a^*/\partial g| \leq \partial\varphi/\partial g$ in the relevant equilibrium configuration, and $\partial\xi^*/\partial g \leq 0$. Strict inequality obtains generically.

Total delegation a^* may rise or fall with g. In the process-preserving regime, a^* rises with g because the constraint $a^* \leq \varphi(e^*; g)$ relaxes outward. In the process-replacing regime, a^* may fall with g if the state-channel cost reduction dominates, or rise if the relaxation of the threshold dominates. The substantive welfare claim depends on the behavior of ξ^*, which is unambiguous.

For effort: higher g shifts the FOC for e in two ways. First, the relaxed process-preservation constraint raises the indirect output return at any given effort level. Second, lower process-replacing excess raises future-state values $\Lambda_h^{t+1}, \Lambda_D^{t+1}, \Lambda_J^{t+1}$ because state-channel losses fall. Both effects raise $\partial e^*/\partial g > 0$. □

A.6. Proof of Lemma 6 (monotonicity in age)

The age dependence of the equilibrium policy operates through three channels: π_t (declining in t), β_t (rising in t), and ω_t (rising in t, when active in the analysis). The three channels produce opposing forces on e^*: π_t falling reduces the discipline-channel return to effort; β_t rising increases the weight on long-run gains from effort; ω_t rising shifts the realized action toward the strategic-optimizer choice.

For the main analysis ($\omega_t = 1$ throughout), the net effect of age on e^* depends on the relative rates of decay of π_t and growth of β_t. Under typical empirical parameterizations, the relative magnitudes vary across the schooling life cycle, with π_t dominating at very young ages and β_t dominating at older ages. The qualitative monotonicity is ambiguous, but the age-gradient of welfare cost in Corollary 1 is unambiguous because it integrates effects across both channels in a specific direction. □

B. Proof of the Capacity-Wedge Decomposition: Additional Details

This appendix supplements the proof of Theorem 1 given in Section 7 with technical details that were referenced but not developed in the main text.

B.1. Cascade-chain derivation for the effort FOC

The effort FOC (19) contains the cascade chain term that we left implicit. The full derivation proceeds as follows. The future-state contribution to the value of effort at age t runs through the knowledge state at all subsequent ages, via the dependence of $\mu_h(D, \mathbf{h}_{<s})$ on h_t for all $s > t$. The total derivative of W with respect to h_t at the equilibrium policy is

$$\Lambda_h^{t+1} \;=\; \sum_{s=t+1}^{T} \delta^{s-t-1}\, \beta_{t+1}\delta\, \frac{\partial \mu_h(D_s, \mathbf{h}_{<s})}{\partial h_t}\, \mu_h\, e_s^*\, (1-\delta_h)^{s-t-1}.$$

(The $(1-\delta_h)^{s-t-1}$ factor reflects the depreciation of the knowledge contribution between period $t+1$ and period s.) The effort FOC then reads

$$w'(y_t)\frac{\partial F}{\partial e} \;=\; \frac{\partial c_e}{\partial e} - \beta_t\delta\, \Lambda_h^{t+1}\, \mu_h(D_t, \mathbf{h}_{<t}) - \beta_t\delta\, \Lambda_D^{t+1}\, \pi_t\gamma_e,$$

which is the form used in the proof of Theorem 1.

B.2. Shadow-value computation for $\Psi_h(t)$

The shadow-value term $\Psi_h(t)$ in equation (24) captures the cascade contribution of substitution at age t. The derivation in the proof of Theorem 1 gives

$$\Psi_h(t) \;=\; 1 + \int_t^T \frac{\partial \mu_h(D_r^U, \mathbf{h}_{<r}^U)}{\partial h_t}\, \mu_h\, e_r^U\, (1-\delta_h)^{r-t-1}\, \delta^{r-t}\, dr.$$

Under the baseline polynomial-in-distance specification (16), $\partial\mu_h/\partial h_t = \mu_h^0\kappa(r-t)^{-p}$ and the integral converges for $p > 1$, producing the polynomial-in-distance cascade depth used in the main analysis. The corresponding expressions under the alternative uniform and exponential-decay specifications (Appendix C.8) follow the same pattern with different cascade-distance kernels.

B.3. Threshold function derivation

The threshold function $g^*(\theta)$ in the threshold characterization of Proposition 3 is implicitly defined by the equation $\partial V_0/\partial\theta|_{g=g^*} = 0$. Differentiating V_0 with respect to θ at fixed g and equilibrium (e^*, a^*) produces

$$\frac{\partial V_0}{\partial \theta} \;=\; \int_0^T \delta^t\, w'(y_t)\frac{\partial F}{\partial \theta}\, dt - \int_0^T \delta^t \left[\lambda_h \frac{\partial (a-\varphi)^+}{\partial \theta}\Psi_h(t) + \pi_t\gamma_a \frac{\partial (a-\varphi)^+}{\partial \theta}\Psi_D\right] dt.$$

The threshold corresponds to the value of g at which the first integral (the output-channel gain) equals the second integral (the state-channel cost). The implicit-function-theorem argument in the proof of Proposition 3 establishes existence, uniqueness, and monotonicity of $g^*(\theta)$.

C. Proofs of Corollaries and Extensions

C.1. Proof sketch of Corollary 2 (capacity-specific divergence)

The three capacity-stock laws of motion (equations (15), (13), (14)) differ in their process-replacing loss terms and their recovery dynamics. Integrating each over a process-replacing episode of duration Δt and computing the long-run difference relative to the no-substitution counterfactual:

For knowledge h: $\Delta h_\infty = \lambda_h \int_0^{\Delta t}(a-\varphi)^+\,ds + O((1-\delta_h)^{\Delta t})$ (the second-order term is the cascade contribution; recovery through subsequent positive e is partial).

For discipline D: $\Delta D_\infty = \pi\gamma_a \int_0^{\Delta t}(a-\varphi)^+\,ds$ with no recovery channel: missed practice cannot be replaced.

For judgment J: $\Delta J_\infty = \lambda_J \int_0^{\Delta t}(a-\varphi)^+\,ds + O(v\cdot\Delta t)$ where the second term reflects recovery through subsequent verification practice (which process-replacing use precludes).

The ordering $\Delta D_\infty \geq \Delta J_\infty > \Delta h_\infty$ obtains because discipline has no recovery channel, judgment has weak recovery, and knowledge has partial recovery through subsequent positive engagement. □

C.2. Cascade-induced dynamic multiplicity

Corollary 4 (Cascade-induced multiplicity)**.** *Under Assumptions A1–A7 with $\partial\mu_h/\partial h_s > 0$ sufficiently large, the deterministic dynamics of h_t on the equilibrium path admit multiple locally stable steady states. The low steady state corresponds to a process-replacing-regime trajectory in which h_t never accumulates sufficient foundation for μ_h to rise above the breakeven threshold; the high steady state corresponds to an active-engagement trajectory in which the foundation accumulates and the cascade-amplified productivity sustains continued accumulation.*

Proof. The knowledge dynamics in the process-replacing regime have fixed-point equation

$$\delta_h h^* \;=\; \mu_h(D^U, h^*\,\mathbf{1})\,e^U(h^*) - \lambda_h\,(a^U(h^*) - \varphi(e^U; g))^+,$$

where the cascade dependence of μ_h on h^* makes the right-hand side an increasing function of h^* on some interval. The left-hand side is linear in h^* with slope δ_h. Multiple fixed points exist if the right-hand side is S-shaped in h^*, which obtains under sufficient cascade depth $\partial\mu_h/\partial h > 0$.

Local stability of the high and low steady states and instability of the middle fixed point follow from the slope of the right-hand side at each fixed point relative to δ_h. □

C.3. Proof sketch of Corollary 3 (subject-specific policy)

The planner's welfare under effort-contingent policy φ^{pol} in subject d has the cascade-cost term scaling with $\Psi_h^{(d)}$. The FOC for φ^{pol} trades off the marginal output-channel gain from permitting more process-preserving delegation against the marginal cascade cost. Higher $\Psi_h^{(d)}$ raises the marginal cascade cost, lowering the optimal φ^{pol}.

The subject-specific implementation requires the institution to distinguish subjects, which is straightforward in the standard curriculum. □

C.4. Proof sketch of Proposition 4 (endogenous governance)

The parental FOC equates marginal welfare of governance to marginal cost:

$$\frac{\partial V_{\text{child}}}{\partial g}\Big|_{g^*} = \frac{\partial C_{\text{parent}}}{\partial g}\Big|_{g^*}.$$

Differentiate with respect to θ: $\partial^2 V_{\text{child}}/\partial\theta\partial g > 0$ raises optimal g^*. Differentiate with respect to parental time or competence: lower marginal cost of governance raises optimal g^*. □

C.5. Proof sketch of Proposition 5 (commitment value rises in cascade depth)

The commitment value $\mathcal{C}_{\text{commit}}(\Psi_h)$ equals the welfare wedge between the planner's preferred effort path and the child's equilibrium effort path. The wedge has a contemporaneous component (from $\beta_t < 1$) and a cascade-amplified component. Differentiating with respect to Ψ_h produces a positive derivative through the cascade-amplification channel. □

C.6. Endogenous curricular sequencing

Proposition 10 (Front-loaded high-cascade ordering)**.** *Suppose the planner can permute the order in which K curricular subjects are taught: a sequence $\sigma : \{1, \ldots, K\} \to \{1, \ldots, K\}$ assigns subject $\sigma(k)$ to schooling-stage k with associated age t_k, where $t_1 < t_2 < \cdots < t_K$. Let $\Psi_h^{(d)}$ denote the cascade depth of subject d at the youngest age, and suppose: (i) the equilibrium process-replacing excess $\xi^{(d)}$ is constant across stages within a subject; (ii) cascade depths factor as $\Psi_h^{(d)}(t) = \Psi_h^{(d)} \cdot \psi(T-t)$ with ψ strictly increasing in the remaining curricular distance. Then the planner's optimal sequence σ^* orders subjects in decreasing $\Psi_h^{(d)}$: the deepest-cascade subject is taught first.*

Proof. The total cascade cost under sequence σ is

$$\sum_{k=1}^{K} \delta^{t_k}\, \lambda_h\, \xi^{(\sigma(k))}\, \Psi_h^{(\sigma(k))}(t_k) \;=\; \lambda_h \sum_{k=1}^{K} \delta^{t_k}\, \xi^{(\sigma(k))}\, \Psi_h^{(\sigma(k))}\, \psi(T - t_k),$$

using assumption (ii). Under assumption (i), $\xi^{(d)}$ is constant within subject and can be absorbed into $\Psi_h^{(d)}$ as $\tilde{\Psi}^{(d)} \equiv \xi^{(d)}\Psi_h^{(d)}$. Define $w_k \equiv \delta^{t_k}\psi(T - t_k)$, the stage-specific weight. By construction, w_k is strictly decreasing in k when ψ grows fast enough relative to discounting (the relevant range in our setting), or has the same monotone structure in either case. The total cost has the form $\sum_k w_k\, \tilde{\Psi}^{(\sigma(k))}$, a sum of products of the stage-weight sequence and the subject-depth sequence at position σ. By the rearrangement inequality of Hardy, Littlewood, and Pólya, this sum is minimized when one sequence is in decreasing order and the other is in matching opposing order, equivalently when the two sequences are sorted in opposite orders. With w_k decreasing in k, minimization requires $\tilde{\Psi}^{(\sigma(k))}$ increasing in k, that is, σ^* orders subjects in decreasing $\tilde{\Psi}^{(d)}$. Under assumption (i), this is the decreasing $\Psi_h^{(d)}$ ordering: the deepest-cascade subject is taught at the earliest stage. □

C.7. Variance under heterogeneous cascade depths

Proposition 11 (Heterogeneous-cascade variance)**.** *Let $\Psi_h^{(i)}$ denote the cascade depth for child i at fixed age $t = 0$, drawn from a distribution Γ on $[\underline{\Psi}, \overline{\Psi}] \subset (0, \infty)$ with strictly positive variance. At fixed governance distribution and in the interior of the equilibrium region, suppose: (i) the equilibrium process-replacing excess $\xi^{(i)}$ is non-decreasing in $\Psi_h^{(i)}$; (ii) welfare $W^R(i)$ is strictly decreasing and differentiable in $\Psi_h^{(i)}\xi^{(i)}$ with a derivative bounded away from zero in absolute value. Then the cross-sectional variance $\mathrm{Var}_\Gamma(W^R)$ is strictly increasing in AI capability θ.*

Proof. We compute

$$\frac{d}{d\theta}\mathrm{Var}_\Gamma(W^R) \;=\; 2\,\mathrm{Cov}_\Gamma\Big(W^R(i),\, \frac{\partial W^R(i)}{\partial \theta}\Big).$$

Under (ii), $W^R(i)$ is strictly decreasing in $\Psi_h^{(i)}\xi^{(i)}$. Under (i), $\Psi_h^{(i)}\xi^{(i)}$ is non-decreasing in $\Psi_h^{(i)}$. Hence $W^R(i)$ is strictly decreasing in $\Psi_h^{(i)}$ on the support of Γ. Similarly, the welfare derivative $\partial W^R(i)/\partial\theta$ inherits the sign and monotonicity properties of $-\partial(\Psi_h^{(i)}\xi^{(i)})/\partial\theta$, which under (ii) and the bounded-derivative condition is strictly negative and strictly decreasing in $\Psi_h^{(i)}$. Both $W^R(i)$ and $\partial W^R(i)/\partial\theta$ are therefore strictly decreasing functions of the random variable $\Psi_h^{(i)}$ with $\mathrm{Var}_\Gamma(\Psi_h^{(i)}) > 0$. By Chebyshev's sum inequality for monotone functions of a common random variable, $\mathrm{Cov}_\Gamma(W^R(i), \partial W^R(i)/\partial\theta) > 0$ when both functions move in the same direction in $\Psi_h^{(i)}$ (both decreasing here), establishing the result. The strict positivity follows from the strict-derivative condition in (ii) and the strict positivity of $\mathrm{Var}_\Gamma(\Psi_h^{(i)})$. □

C.8. Alternative cascade specifications

The main analysis uses the polynomial-in-distance specification (16). Two alternative specifications are widely used and yield the same qualitative Capacity-Wedge Theorem.

Uniform single-period dependence. The simplest specification has μ_h depend only on the most recent prior knowledge stock:

$$\mu_h(D_t, \mathbf{h}_{<t}) \;=\; \mu_h^0(D_t)\cdot[1 + \kappa_1 h_{t-1}].$$

Under this specification, the cascade operates through a single-period horizon. The present-value cost of substitution at age t exceeds the contemporaneous cost by a single-period multiplier.

Exponential-decay dependence. An intermediate specification has μ_h depend on all prior stocks with weight that decreases exponentially in curricular distance:

$$\mu_h(D_t, \mathbf{h}_{<t}) \;=\; \mu_h^0(D_t) \cdot \Big[1 + \kappa_3 \sum_{s=1}^{t} \exp(-\nu s) h_{t-s}\Big].$$

This specification generates a faster-decaying cascade than the baseline polynomial form. The cascade-depth $\Psi_h(t)$ admits a closed-form expression in this case, and the qualitative content of Theorem 1 is unchanged.

In both alternative specifications, the decomposition $\mathcal{W}^{\text{active}} - W^R = \mathcal{C}^{\text{contemp}} + \mathcal{C}^{\text{cascade}}$ holds, with the cascade component scaled by the cascade-depth function specific to the chosen specification.

C.9. Continuous-time formulation: interpretive analog

The discrete-time model admits a continuous-time analog that we present as an interpretive complement rather than as a separate theorem. A full functional-analytic treatment, with the regularity conditions on policy paths and the kernel space that an honest continuous-time proof would require, lies beyond the scope of this paper. With that caveat, the continuous-time state evolves according to

$$\begin{aligned}
\dot{h}(t) &= -\delta_h h(t) + \mu_h(D(t), \mathbf{h}_{<t}) e(t) + \nu_h a(t) \mathbf{1}[a \le \varphi] - \lambda_h (a(t) - \varphi)^+, \\
\dot{D}(t) &= -\delta_D D(t) + \pi(t)\gamma_e e(t) - \pi(t)\gamma_a (a(t) - \varphi)^+, \\
\dot{J}(t) &= -\delta_J J(t) + \mu_J(J)(v(t) + \eta e(t)) - \lambda_J (a(t) - \varphi)^+.
\end{aligned}$$

Under regularity conditions analogous to those of the discrete-time version, the analog of the Capacity-Wedge Decomposition holds with the discrete sums replaced by integrals. The cascade depth admits the closed-form expression

$$\Psi_h(t) \;=\; 1 + \int_t^T \exp(-(\delta_h + \rho_d)(s-t)) \cdot \frac{\partial \mu_h(D(s), \mathbf{h}_{<s})}{\partial h_t} \cdot \mu_h \cdot e(s)\, ds.$$

The closed form makes the curricular-distance kernel transparent: $\Psi_h(t)$ decreases exponentially in remaining curricular distance with rate $\delta_h + \rho_d$, scaled by the cross-derivative $\partial \mu_h / \partial h_t$ and the effort path. We treat this expression as interpretive rather than as a free-standing theorem; the discrete-time analysis carries the formal content of the paper.

D. Proofs of Welfare Theorems

D.1. Proof of Welfare Theorem 2 (Pareto dominance of process-preserving use)

By Proposition 2, $W^P \ge W^R$ for every household type, with strict inequality when the process-replacing regime is operative. The decomposition in Theorem 1 gives the magnitude of the wedge as $\mathcal{C}^{\text{contemp}} + \mathcal{C}^{\text{cascade}}$. Strictness obtains under the standard conditions on the state-channel coefficients. □

D.2. Proof of Welfare Theorem 3 (second-best characterization)

The planner's FOC with respect to $\varphi_i^{\text{pol}}(e)$ is

$$w'(y)\frac{\partial F}{\partial a}\bigg|_{a=\varphi^{\text{pol}}} = \lambda_h \Psi_h + \pi\gamma_a \Psi_D,$$

which corresponds to the threshold-governance type's FOC. Hence the second-best $\varphi_i^{\text{pol},*} = \min\{\varphi(e; g^*(\theta)), \varphi(e; g_i)\}$, with the min binding for low-g_i households. □

D.3. Proof of Welfare Theorem 4 (optimal commitment under cascade)

The wedge between optimal commitment and self-imposed restraint is the difference between the planner's optimum effort and the child's equilibrium effort. The wedge scales with $(1-\beta_t)$ through the discount-factor wedge and with Ψ_h through the cascade-amplification. Optimal commitment closes the wedge by tightening the effort-contingent constraint by an amount proportional to $(1-\beta_t)\Psi_h$. □

References

Acemoglu, D. (2024). The simple macroeconomics of AI. *Economic Policy*, forthcoming.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6): 1488–1542.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6): 2188–2244.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). *Prediction Machines: The Simple Economics of Artificial Intelligence*. Harvard Business Review Press.

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakci, O., and Mariman, R. (2024). Generative AI can harm learning. Working paper, Wharton School.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3): 871–915.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4): 848–886.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4): 352–365.

Bereiter, C. and Scardamalia, M. (1987). *The Psychology of Written Composition*. Lawrence Erlbaum Associates.

Bertsch, S., Pesta, B. J., Wiscott, R., and McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35(2): 201–210.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (eds.), *Metacognition: Knowing about knowing*, pp. 185–205. MIT Press.

Bjork, E. L. and Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher et al. (eds.), *Psychology and the real world*, pp. 56–64. Worth.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative AI at work. NBER working paper.

Brynjolfsson, E. and McAfee, A. (2014). *The Second Machine Age*. W. W. Norton.

Cunha, F. and Heckman, J. J. (2007). The technology of skill formation. *American Economic Review*, 97(2): 31–47.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3): 883–931.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3): 363–406.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, 15(11): 988–994.

Harris, C. and Laibson, D. (2001). Dynamic choices of hyperbolic consumers. *Econometrica*, 69(4): 935–957.

Heckman, J. J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy*, 84(4, Part 2): S11–S44.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, 6: 689–733.

Heckman, J. J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6): 2052–2086.

Hiebert, J. and Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (ed.), *Second handbook of research on mathematics teaching and learning*, pp. 371–404. Information Age Publishing.

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5): 1008–1022.

Karpicke, J. D. (2012). Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21(3): 157–163.

Karpicke, J. D. and Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865): 966–968.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. OECD report.

Kestin, G. et al. (2024). AI tutoring outperforms active learning. Working paper, Harvard University.

Krusell, P. and Smith, A. A. (2003). Consumption-savings decisions with quasi-geometric discounting. *Econometrica*, 71(1): 365–375.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2): 443–477.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654): 187–192.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1): 103–124.

O'Donoghue, T. and Rabin, M. (2001). Choice and procrastination. *Quarterly Journal of Economics*, 116(1): 121–160.

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:2302.06590.

Pressley, M. (2002). *Reading Instruction That Works: The Case for Balanced Teaching*. Guilford Press.

Roediger, H. L. and Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3): 249–255.

Slamecka, N. J. and Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6): 592–604.

Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3): 165–180.