AI and the Reproduction of Judgment

Fikret Bilenkisi
University of Huddersfield

Filippos Maraziotis*
IQS School of Management
Universitat Ramon Llull

Ramon Xifré
IQS School of Management
Universitat Ramon Llull,
& IESE Business School, Public–Private Sector Research Center

*Corresponding author: filippos.maraziotis@iqs.url.edu

October, 2026

Abstract

Knowledge organizations produce current decisions and future decision makers through the same work. We show that AI can separate these outputs because correct actions need not identify the judgment that produced them. What matters for reproducing judgment is whether a recipient's human and verified-AI experiences jointly reveal the criterion needed for future decisions, and whether that information becomes durable independent capability. Embedding this process in organizational design yields an autonomy rebound: once the organization has expanded beyond the scale at which every live relationship can remain developmental, greater value from AI-originated decisions expands current activity but reduces developmental relationships and successful succession, even though decision quality and learning within preserved relationships are unchanged and current surplus rises. In multidimensional problems, AI substitutes for particular diagnostic directions rather than for human work in general, leaving a residual set of experiences that organizations optimally preserve. In equilibrium, an inward shift in successor supply raises the scarcity value of reproduced judgment and attenuates but does not reverse the aggregate contraction. The broader implication is that scalable execution and the reproduction of judgment are distinct economic margins: AI can make existing judgment easier to deploy while making independently formed decision makers scarcer.

Keywords: Artificial intelligence; judgment; organizational design; human capital; knowledge hierarchies; succession.
JEL classification: D23, D83, J24, O33.

1 Introduction

Organizations often do two things at once. They produce future decision makers while solving today's difficult problems. Consider two consulting teams with the same

analyst–partner pair, the same case, the same AI capability, the same scheduled partner time, and the same final recommendation. In one team, the analyst commits to an assessment before seeing the answer and the partner responds to that assessment. In the other, AI originates the recommendation; the analyst checks or implements it and the partner reviews the completed work. The case can be solved equally well in both teams. Yet they do not necessarily leave the analyst with the same understanding of how the partner would judge the next case.

That happens because an action can contain less information than the judgment behind it. "Approve," "operate," or "proceed" may each be the right action for many different underlying assessments. Observing the action tells us what was done. It does not tell us which considerations mattered, how they were weighed, or what would have changed the decision:

$$\text{same action} \;\not\Rightarrow\; \text{same criterion.}$$

This distinction is familiar in organizations, but organizational economics typically treats the decision as the output of the problem-solving process. We study a second output: the information that process leaves with a future decision maker about the criterion used to reach the decision.

The relevant environment lies between two polar cases. At one extreme, experienced knowledge can be written down completely as a codifiable rule that anyone can apply. Then there is little reason to reproduce it through case-by-case interaction: the organization can transmit the rule directly. At the other extreme, experience is completely inarticulate: even after seeing a case, an expert cannot communicate anything about why one assessment is better than another. Then the developmental channel studied here is unavailable. Our interest is the intermediate case in which a complete rule is unavailable but realized cases make parts of experienced judgment communicable. A senior physician may be unable to write a complete diagnostic algorithm yet can explain, after seeing a patient, why a resident put too much weight on one symptom and too little on another. A deal partner can explain why a contractual provision changes the assessment of a transaction; an experienced engineer why a design that passes ordinary checks becomes unsafe under a particular combination of constraints. Particular cases reveal pieces of judgment even when no single case reveals the whole criterion.

This observation creates an organizational object that is absent from standard measures of human work or AI use. Suppose a less experienced decision maker forms an assessment before receiving the experienced source's answer. The subsequent disagreement can reveal how the source evaluates that case differently. We call the source and sequence that make such information interpretable decision provenance. Across sufficiently varied cases, the resulting restrictions form a recipient-specific diagnostic record. What matters for reproducing judgment is not the number of human decisions observed, the amount of senior contact, or the share of work formally kept human. It is whether the record contains the distinctions needed to recover the part of the criterion the organization wants future decision makers to retain.

The paper develops the organizational consequences of treating this diagnostic record as an output of current work. Human information has no privileged status in the theory. Verified AI information enters in exactly the same way when it provides independently validated information about the maintained criterion. AI can therefore supply the operative action without supplying the relevant diagnostic information, supply both, or make some human diagnostic episodes redundant. The economically relevant

distinction is not human versus machine decision making. It is between technologies that produce the current action and technologies that also produce information from which another decision maker can be formed.

Information alone is not enough to produce a successor because the recipient must retain what the record reveals and later use it without returning to the experienced source. We therefore distinguish a developmental opportunity from a successful successor. Whether the record contains the distinctions needed for future decisions is a question of diagnostic coverage; how effectively the recipient turns that record into durable independent capability is a question of assimilation. Organizations accordingly choose not only whether current work generates the information needed for development, but also who receives those opportunities, which cases enter their experience, and how many resources are devoted to turning that experience into independent judgment.

This framework yields a simple organizational tradeoff. Preserving a developmental relationship can produce a future decision maker who no longer requires the experienced source. But it also uses scarce senior capacity and can require foregoing the current gains from allowing AI to originate the decision immediately. Consider an improvement in AI that raises the value of autonomous action origination while holding current decision quality, AI's diagnostic contribution, senior-time requirements, the technology of assimilation, and the value of successors fixed. When the organization is already large enough that not every relationship can remain developmental, greater autonomy expands current activity and reduces developmental pipelines. Assimilation within the relationships that remain developmental is unchanged, so the organization produces fewer successful successors even as current operating surplus rises. We call this the autonomy rebound.

The same informational object determines where development is preserved. In multidimensional judgment problems, different cases reveal different directions of the maintained criterion. Conditional on verified information already supplied by AI or other technologies, the human developmental role is residual. In a frictionless benchmark, the organization preserves the least costly set of human-first experiences needed to complete the missing diagnostic directions—a residual diagnostic basis. With costly assimilation, full coverage is not enough: two records that identify the same criterion can differ in how effectively they can be converted into durable capability. The organization therefore chooses diagnostic geometry together with assimilation resources. When one durable record is unambiguously more informative than another, it generates a higher probability of successful succession; if it is also more costly, the organization adopts it only when successors are sufficiently valuable.

This distinction also changes equilibrium and measurement. Successful successors are an organizational output whose scarcity value feeds back into developmental investment. When an AI improvement shifts organizational successor supply inward, the supply of reproduced judgment falls and its scarcity value rises. The price response attenuates but does not reverse the contraction. If the improvement makes development more costly for some organizations than for others, the resulting scarcity can also reallocate successor production toward the less affected organizations. Higher scarcity value can increase assimilation within preserved pipelines or make a more effective developmental design worth its additional cost. Empirically, conventional measures can miss the mechanism. Two workflows can have the same current decisions, AI use, human-task shares, and review time while leaving recipients with different diagnostic records and different subsequent independent capability. A direct test would randomize

whether AI is revealed before or after a sealed human assessment, hold cases and eventual AI access fixed, and measure later independent performance on new cases.

The paper builds on the organizational economics of knowledge and on work studying how AI changes the production and transmission of expertise. Knowledge-hierarchy models show how scarce expertise, knowledge acquisition, and communication shape organizational span and specialization (Garicano 2000; Garicano and Rossi-Hansberg 2006, 2015); task models study how technology reallocates activities across inputs (Acemoglu and Autor 2011; Acemoglu and Restrepo 2018). Ide and Talamàs (2025) introduce autonomous and nonautonomous AI into a knowledge economy, while Autor and Thompson (2025) study expertise. J. S. Gans (2025) studies the cognitive division of labor between specialists who generate tacit knowledge and generalists who use codified knowledge, with the knowledge base itself evolving; broader work considers AI and the organization of firms and the economy (Agrawal, Brynjolfsson, and Korinek 2026; Chatterji, Rock, and Talamàs 2025; J. Gans 2025). We add a different informational margin: current work can leave another person with information about the criterion used to resolve a problem, and that output need not be summarized by the action that was taken.

Learning and apprenticeship models study training investment, learning by doing, learning from coworkers, and relational knowledge transfer (Arrow 1962; Becker 1964; Fudenberg and Rayo 2019; Garicano and Rayo 2017; Herkenhoff et al. 2024; Jarosch, Oberfield, and Rossi-Hansberg 2021). The closest AI-learning papers study career and apprenticeship viability, senior interaction and teamwork, the configuration of autonomy and review, intergenerational tacit-knowledge transmission, and career dynamics in which learning-by-doing, task automation, managerial frontier maintenance, and worker-to-manager transitions are jointly determined (Afrouzi et al. 2026; Bárány and Koren 2026; Garicano and Rayo 2025; Ide 2026; Petropoulos 2026). Those mechanisms concern how AI changes learning, developmental interaction, career progression, or the viability of apprenticeship. The mechanism here is different: workflow itself determines the diagnostic information about judgment that current work produces for a particular recipient. A related set makes cognition or effort endogenous: Acemoglu, Kong, and Ozdaglar (2026) study agentic recommendations that substitute for costly human learning effort and can erode collective knowledge, Agarwal, Moehring, and Wolitzky (2025) allow AI-induced effort crowd-out in human–AI collaboration without imposing a structural model of human decision making, and Caplin (2026) studies costly refinement of which diagnostic questions to ask. Here developmental resources determine how a recipient-specific diagnostic record becomes durable independent judgment, and their return depends on the geometry produced by organizational workflow. Work on verification and managerial span studies complementary margins of confidence, escalation, and expert workload (Lamba 2026; Xu et al. 2026).

The paper thus studies an old organizational activity through a different economic object. Difficult work does not only produce decisions. In many organizations, it is also the setting in which one person's judgment becomes another person's capability. AI can change these two outputs differently.

2 Decision Provenance and Diagnostic Identification

This section isolates the informational requirement for reproducing judgment. The central distinction is between the action produced by current work and the information

that work leaves about the criterion behind the action. We show how realized cases can generate restrictions on that criterion, when human and verified-AI information jointly identify what the organization wants to preserve, and why identification is still not sufficient for durable independent capability. The organizational choice of how much of this process to operate comes in the next section.

2.1 From judgment to action

Consider an experienced decision maker whose judgment is already formed. We call this person the *source*. A less experienced decision maker whom the organization may want to develop is a *recipient*.

Judgment is the capacity to make and bear a consequential determination when the operative objective, the relevant features, or the relative costs of errors cannot be fully specified in advance. The model does not represent that capacity exhaustively. Instead, let

$$\boldsymbol{\theta} \in \mathbb{R}^d$$

be a local diagnostic representation of the source's judgment on a contextual domain $\mathcal{X} \subset \mathbb{R}^d$. A realized case is described by $\boldsymbol{x} \in \mathcal{X}$, and the source's diagnostic index on that case is

$$s_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}'\boldsymbol{x}.$$

The coordinates of $\boldsymbol{x}$ are considerations that become describable once the case is encountered; $\boldsymbol{\theta}$ records how the source weighs them locally. Thus $\boldsymbol{x}$ describes the case, while $\boldsymbol{\theta}$ describes how the source evaluates its features. Tacit and practical knowledge are broader related objects (Polanyi 1966; Ryle 1949), while Knightian uncertainty describes circumstances in which such judgment may be exercised (Knight 1921).

The organization need not want to reproduce every component of $\boldsymbol{\theta}$. It may care only about distinctions that matter for a particular set of future decisions or for accountability. Let $M \in \mathbb{R}^{q \times d}$ have full row rank and define

$$\phi = M\boldsymbol{\theta} \in \mathbb{R}^q.$$

We call ϕ the *maintained criterion*: the part of the source's local criterion that the organization wants future decision makers to retain. Two values of $\boldsymbol{\theta}$ that generate the same $M\boldsymbol{\theta}$ are therefore equivalent for the organization's maintained purpose. Components in $\ker(M)$ may adapt without counting as a failure to reproduce the maintained criterion.

For a case relevant to this criterion, write

$$\boldsymbol{x} = M'\boldsymbol{z}, \qquad \boldsymbol{z} \in \mathbb{R}^q.$$

On such a case, the source's diagnostic index depends on the maintained criterion through

$$\boldsymbol{\theta}'\boldsymbol{x} = \phi'\boldsymbol{z}.$$

We call $\boldsymbol{z}$ the case's *target direction*. Different target directions provide different linear restrictions on the maintained criterion, revealing different aspects of how the source evaluates cases. The relevant target directions are assumed to span $\mathbb{R}^q$. Choosing a different M changes what the organization seeks to preserve; it is adaptation of the target rather than failure to reproduce the old one.

We can now separate the criterion from the action observed in current production. The operative action belongs to $\mathcal{Y}(x)$ and is

$$y^*(x;\theta) = \Gamma_x(s_\theta(x)). \tag{1}$$

The map Γ_x can be many-to-one. If a project is approved whenever the diagnostic index exceeds a threshold, for example, different assessments on the same side of that threshold generate the same action.

For a fixed case, call the set of diagnostic-index values that generate the same action an *action cell*. The benchmark assumes that the source's index is not exactly on the boundary between cells.

Assumption 1 (Locally coarse action)**.** *For every context in the finite action record, the source index lies in the interior of an action cell: there exists $\epsilon_x > 0$ such that*

$$\Gamma_x(s) = \Gamma_x(s_\theta(x)) \qquad \textit{whenever } |s - s_\theta(x)| < \epsilon_x.$$

The assumption simply says that a small change in the underlying assessment need not change the observed action. This is what makes the action coarser than the criterion that generates it.

Proposition 1 (Action equivalence)**.** *Under Assumption 1, any finite record of correct operative actions is consistent with a nontrivial neighborhood of source criteria. Hence, for any maintained criterion of positive dimension, correct action observations alone need not identify $\phi = M\theta$.*

Correct actions need not be uninformative: they can rule out some possible criteria. The result is narrower. A finite record of correct coarse actions can leave a range of criteria consistent with what was observed and therefore cannot eliminate a missing target direction merely by being correct. Current production can consequently deliver the right action without reproducing the judgment behind it.

We next specify the benchmark role of AI in producing the current action. A *live context* is a case the organization must resolve as part of current production. An action is *source-consistent* when it equals $y^*(x;\theta)$, and let a index AI capability.

Assumption 2 (Scalable AI action origination)**.** *At capability a, AI can originate the source-consistent operative action on any live context assigned to it without requiring a formed-human source to originate that action. AI action origination has no separate capacity constraint proportional to the stock of formed human judgment. Human accountability or oversight may still consume senior capacity.*

This benchmark gives AI the correct current action, so the mechanism does not depend on contemporaneous AI error. It also separates the scalability of action origination from the stock of formed human judgment: producing another source-consistent action does not itself require another formed source. The assumption concerns action origination only. AI may also provide information about the maintained criterion, which we define next.

2.2 Diagnostic information and provenance

Proposition 1 says that seeing the final action may leave too much of the criterion unresolved. A different workflow can reveal more.

Suppose recipient i initially evaluates cases using criterion $\boldsymbol{\theta}_i^0$. On a target context $\boldsymbol{x} = M'\boldsymbol{z}$, a human-first workflow requires the recipient to form an attributable assessment before the source or AI answer is revealed:

$$t_i(\boldsymbol{x}) = \boldsymbol{\theta}_i^{0\prime}\boldsymbol{x}.$$

The timing matters because an assessment supplied after the answer is known need not reveal what the recipient would have believed independently.

After observing the attempt, the formed source can provide the recipient-specific correction

$$m_i(\boldsymbol{x}) = s_{\boldsymbol{\theta}}(\boldsymbol{x}) - t_i(\boldsymbol{x}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_i^0)'\boldsymbol{x}.$$

If the correction reaches the same recipient, then

$$t_i(\boldsymbol{x}) + m_i(\boldsymbol{x}) = \boldsymbol{\theta}'\boldsymbol{x} = \phi'\boldsymbol{z}.$$

The interaction therefore gives recipient i a known linear restriction on the maintained criterion: it reveals the value of ϕ in direction $\boldsymbol{z}$.

This channel requires the source to be able to articulate a discrepancy on a realized case even when the complete criterion cannot be written down in advance.

Assumption 3 (Contextual articulability)**.** *After observing a realized context and an attributable pre-correction attempt, formed judgment can produce the recipient-specific local correction $m_i(\boldsymbol{x})$ on that context. The complete maintained criterion is not costlessly contractible or communicable as an operative mapping independently of contextual episodes.*

The assumption places the model between two polar cases. If the complete criterion could be written down and transferred directly, case-by-case interaction would be unnecessary. If the source could communicate nothing about a discrepancy even after seeing the case, this developmental channel would be unavailable. The model studies the intermediate case in which realized contexts make particular pieces of judgment communicable.

We use *decision provenance* for the features of the workflow that make a diagnostic restriction interpretable: what generated it and the maintained criterion to which it is linked. For a recipient-specific correction, provenance additionally includes the prior assessment relative to which the correction was generated and correct delivery of that correction to the same recipient. Because $m_i(\boldsymbol{x})$ is defined relative to recipient i's own assessment, a correction intended for one recipient cannot in general be reassigned to another without changing its informational content.

A human-first attempt followed by a correctly coupled source correction is therefore one technology for producing a target-linked diagnostic restriction. We refer to such interactions as correction-based human diagnostic episodes.

A directly validated source-index measurement is a different technology. On a target context,

$$s_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}'\boldsymbol{x} = \phi'\boldsymbol{z}.$$

Because its target-linked content is established directly, such a measurement need not be generated from a recipient's prior attempt. If it is supplied to the recipient, it enters the diagnostic record as a verified restriction on the maintained criterion.

AI can supply the same kind of information. An explanation, critique, or counterfactual is credited as diagnostic information whenever its relation to the maintained target is independently verified. A verified AI diagnostic measurement is an AI-generated signal whose mapping to a known linear functional $\phi' z$ has been independently established. In the exact benchmark, the restriction is measured without error. In applications, verification could come from held-out calibration against source-index measurements, an audited measurement protocol, or another procedure that validates the mapping. Appendix E separately allows noise in the target-linked diagnostic record.[1]

The diagnostic technology is therefore source-neutral conditional on content. A human-generated restriction and a verified AI-generated restriction contribute in the same way when they convey the same information about ϕ. The relevant distinction is not human information versus machine information. It is between information that is verifiably linked to the maintained criterion and an operative action that need not reveal that criterion.

2.3 Diagnostic span and identification

A single diagnostic episode generally reveals only one restriction on ϕ. When the maintained criterion has several dimensions, the recipient needs enough independent restrictions to recover all of them.

Let $\mathcal{E}_i^H$ be recipient i's set of verified human diagnostic items, including correction-based human diagnostic episodes and any directly validated human source-index measurements. For each item $e \in \mathcal{E}_i^H$, let $z_e \in \mathbb{R}^q$ denote the target direction it supplies.

The recipient may also receive verified AI diagnostic measurements. Let

$$\mathcal{A}_i(a) \subseteq \mathbb{R}^q$$

be the span of the target directions supplied by those measurements. Thus $\mathcal{A}_i(a)$ records the directions of the maintained criterion for which verified AI information is already available.

Combining the two sources gives recipient i's total verified diagnostic span,

$$\mathcal{D}_i(a) = \mathcal{A}_i(a) + \operatorname{span}\{z_e : e \in \mathcal{E}_i^H\}.$$

We say that the recipient has *full diagnostic coverage* when

$$\mathcal{D}_i(a) = \mathbb{R}^q.$$

Full coverage means that no direction of the maintained criterion is missing from the recipient's verified information.

For comparison with the source-coordinate formulation, let X_i^H denote the matrix whose rows are the source-coordinate contexts associated with recipient i's verified human measurements.

1. A finite outside human specialist is treated symmetrically. If the organization adopts that specialist's criterion as the maintained target and the interaction produces verified target-linked measurements, those measurements enter the diagnostic record. If the specialist supplies only actions, the informational problem can resemble AI-first execution, but action origination remains tied to finite expert capacity. The AI primitive studied here combines scalable action origination with a potentially distinct diagnostic-transmission technology.

Proposition 2 (Diagnostic span)**.** *Suppose the verified diagnostic measurements are exact. The maintained criterion ϕ is identified for recipient i if and only if*

$$\mathcal{D}_i(a) = \mathbb{R}^q.$$

In the no-verified-AI full-target case $M = I_d$, this is the usual full-rank condition on the matrix of source-linked human contexts. More generally, with human measurements only, it is equivalent to $\mathrm{row}(M) \subseteq \mathrm{row}(X_i^H)$.

The relevant object is therefore coverage rather than exposure volume. Many episodes can add no new direction; a small number can suffice when they complete the target span. Counts of cases, senior contact, or human-task shares are consequently not sufficient statistics for whether the developmental record identifies the maintained criterion.

The proposition also makes the substitution margin precise. Verified AI information reduces the human diagnostic span that remains to be supplied. If AI supplies a verified restriction in one target direction, human experience need not reproduce that same direction again; it needs to supply only the directions that remain missing. At the limiting case $\mathcal{A}_i(a) = \mathbb{R}^q$, verified AI information alone covers the maintained criterion and no human diagnostic episode is required for identification.

Correct AI actions are different. Because the action map is coarse, repeatedly observing source-consistent actions does not by itself close a missing target direction. What matters is the diagnostic content of the information, not whether the final action happened to be correct.

Appendix A records the equivalent source-coordinate rank statements, and Appendix E gives the corresponding noisy-precision results.

2.4 From diagnostic coverage to successful succession

Diagnostic identification tells us what information is available to the recipient. It does not tell us whether that information becomes durable independent capability. We call the process through which a recipient converts the available diagnostic record into capability that can later be used without returning to the source or replaying the original episodes *assimilation*.

Let $\widehat{\phi}_i$ denote the criterion recipient i can deploy independently after assimilation. It need not be numerically identical to ϕ. What matters is how well it performs on the decisions the recipient is expected to face.

Suppose future target contexts z are drawn from a distribution F_i with full-rank second moment. Define

$$\Omega_i \equiv \mathbb{E}_{F_i}[zz'] \succ 0.$$

The matrix Ω_i reflects the recipient's future decision environment. Payoff weights on different contexts can be incorporated into the same second-moment matrix, so it is determined by the decisions the successor is expected to face rather than chosen as a free statistical metric.

For a future target context, the difference between the diagnostic index generated by the recipient and the maintained criterion is

$$z'(\widehat{\phi}_i - \phi).$$

Expected squared diagnostic-index error is therefore

$$\mathbb{E}_{F_i}\left[\left(z'(\widehat{\phi}_i - \phi)\right)^2\right] = (\widehat{\phi}_i - \phi)'\Omega_i(\widehat{\phi}_i - \phi).$$

Define the recipient's independent-performance loss as

$$\Delta_i = (\widehat{\phi}_i - \phi)'\Omega_i(\widehat{\phi}_i - \phi). \tag{2}$$

Let $\overline{L}_i > 0$ be the maintained performance standard: the largest loss consistent with the organization's required level of independent performance.

Recipient i is a *successful successor* if three conditions hold:

(i) her diagnostic record has full coverage,

$$\mathcal{D}_i(a) = \mathbb{R}^q;$$

(ii) after assimilation, her independent-performance loss satisfies

$$\Delta_i \leq \overline{L}_i;$$

(iii) she can implement and bear the corresponding target decisions without returning to the source.

Coverage and performance are deliberately separate requirements. A recipient who never acquires one maintained direction is not counted as successfully reproduced merely because she happens to perform well on a finite set of realized cases. The performance threshold evaluates the quality of independent capability conditional on coverage; favorable realized outcomes cannot substitute for a missing part of the maintained criterion.

For the same reason, successful succession does not require exact numerical equality,

$$\widehat{\phi}_i = \phi.$$

It requires a covered criterion that the recipient can deploy independently within the economically relevant performance standard.

The distinction between coverage and assimilation is important for what follows. Diagnostic coverage determines whether the developmental process supplies the information the successor needs. Assimilation determines how effectively that information becomes durable independent capability. An organization can provide a complete diagnostic record but devote too few resources to assimilating it; it can also devote substantial developmental effort to a record that still leaves a target direction uncovered.

The next section embeds these margins in a simple organizational problem. It uses a two-direction benchmark in which developmental relationships provide complete diagnostic coverage and introduces a scalar assimilation technology that determines the probability of successful succession. Section 4.2 then returns to the multidimensional setting and derives that probability from the geometry of the combined human-and-AI diagnostic record. Appendix A retains the frictionless exact-acquisition benchmark as a polar case.

3 Organization and the Autonomy Rebound

Section 2 separates action production, diagnostic coverage, and assimilation at the recipient level. We now turn those informational requirements into an organizational choice. The organization has a scarce input—senior capacity—that supports both current production and the development of future decision makers. A two-direction benchmark keeps this allocation problem transparent while making successful succession an endogenous output of developmental opportunities and assimilation.

3.1 A simple organization

Set $q = d = 2$, $M = I_2$, and suppose there is no verified AI diagnostic span. The organization therefore seeks to reproduce the full two-dimensional maintained criterion through human diagnostic information. The benchmark uses the recipient-specific correction technology of Section 2, so complete diagnostic coverage requires two linearly independent target directions.

A live relationship can be organized in one of two ways. In a *human-first live relationship*, the recipient makes an attributable assessment before the answer is revealed and then receives the correctly coupled source correction. This interaction supplies one target direction while also producing the source-consistent current action.

In an *AI-first live relationship*, AI originates the source-consistent current action before the recipient makes an independent diagnostic assessment. In the present benchmark, which has no verified AI diagnostic span, this mode does not supply the correction-based live diagnostic restriction. Human accountability or review may still be present; the distinction concerns who originates the current action and whether the recipient-first diagnostic interaction is preserved.

A single human-first live relationship supplies only one direction of the two-dimensional maintained criterion. To complete diagnostic coverage, the organization can reserve a separate, non-identical counterfactual case on which the recipient again makes an independent attempt and receives source correction. The counterfactual case supplies the second target direction, assumed to be linearly independent of the first.

A recipient enters the *developmental pipeline* when the organization provides both pieces of this diagnostic process: a human-first live relationship and the additional rank-completing counterfactual episode. Entry into the pipeline provides complete diagnostic coverage in this benchmark. It does not yet guarantee successful succession, because the resulting record must still be assimilated. Appendix A gives the corresponding correction-based four-function implementation and shows why each component is needed for an open set of criteria. The organization can commit to AI-revelation timing and reserve senior capacity for the counterfactual episode.

We can now define the organization's quantities. Let n denote total live span, the mass of current relationships served. Let $h \leq n$ denote the mass of those relationships organized human-first, so $n - h$ is the mass organized AI-first. Finally, let T denote the mass of recipients entering the developmental pipeline. Because every developmental recipient requires a human-first live relationship,

$$0 \leq T \leq h \leq n.$$

The scarce organizational input is senior capacity. Let $B > 0$ denote the total amount available. An AI-first live relationship uses $\tau_A > 0$ units of senior capacity, including any

accountability or review that remains after AI originates the action. A human-first live relationship uses $\tau_H \geq \tau_A$ units. The difference $\tau_H - \tau_A$ captures any additional senior involvement required to preserve the recipient-first diagnostic interaction. Completing the rank-completing counterfactual episode for one developmental recipient requires a further $\rho_R > 0$ units of senior capacity.

Total senior-capacity use is therefore

$$\tau_A(n - h) + \tau_H h + \rho_R T \leq B. \tag{3}$$

The constraint separates three uses of senior capacity: supporting AI-first current relationships, preserving human-first live diagnosis, and completing the additional diagnostic direction required for development. The organizational trade-off arises because the same senior capacity supports current scale and the production of future decision makers.

3.2 Current production, assimilation, and succession

We now assign values to current production and to successful succession. Both live modes produce the source-consistent current action, so their difference does not come from current decision quality.

Let $b > 0$ be current surplus common to the two modes. AI-first execution adds an autonomy value $g(a) \geq 0$, with $g'(a) \geq 0$. This is the current value of allowing AI to originate the action rather than waiting for an independent human assessment and the associated source-linked diagnosis. It can reflect lower decision latency, less pre-decision junior effort or synchronization, greater parallelism, or related gains from origination. It does not represent a better current decision, and it does not require lower senior-time use.

Only AI-first live relationships receive this additional value. Since their mass is $n - h$, total autonomy surplus is $g(a)(n - h)$.

Expanding live span may also be costly independently of the senior-capacity constraint. Let $C(n)$ denote this span cost.

Assumption 4 (Span cost)**.** *$C : [0, \infty) \to \mathbb{R}_+$ is twice continuously differentiable, $C(0) = C'(0) = 0$, $C''(n) > 0$, and $C'(n) \to \infty$.*

The convexity of C makes further expansion increasingly costly and ensures a well-behaved choice of organizational span.

A recipient who enters the developmental pipeline has complete diagnostic coverage in this two-direction benchmark. Coverage, however, is not yet successful succession. The recipient must still assimilate the diagnostic record into durable independent capability. Let $p \geq 0$ denote assimilation investment per developmental recipient. Given p, let $\chi(p) \in [0, 1]$ be the probability that the recipient subsequently meets the maintained independent-performance standard. Let $\zeta c(p)$ be the real assimilation-resource cost per pipeline recipient.

Thus T and $\chi(p)$ describe different margins. T is the mass of recipients who receive a complete developmental opportunity; $\chi(p)$ determines how effectively each such opportunity is converted into successful succession. The scalar success technology is a fixed-design benchmark nested in the source-neutral diagnostic-geometry model in

Section 4.2.[2]

Assumption 5 (Benchmark assimilation technology)**.** *χ and c are twice continuously differentiable on $[0,\infty)$. The success function satisfies $\chi(0)=0$, $\chi'(0)>0$, $\chi'(p)>0$ for $p>0$, and $\chi''(p)\le 0$. The resource cost satisfies $c(0)=c'(0)=0$, $c'(p)>0$ for $p>0$, $c''(p)>0$, and $c(p)\to\infty$. Let $\zeta>0$.*

More assimilation resources therefore raise the probability of successful succession at a weakly diminishing rate, while their resource cost is strictly convex. For every $A>0$, the conditions below will imply a unique interior assimilation choice.

Each successful successor creates continuation value $V>0$. The organization captures a share $\alpha\in[0,1]$ of that value. Write

$$A=\alpha V.$$

Thus A is the private continuation value the organization assigns to one successful successor.

Before reducing the live-mode choice, the organization solves

$$\max_{0\le T\le h\le n,\ p\ge 0}\left\{bn-C(n)+g(a)(n-h)+[A\chi(p)-\zeta c(p)]T\right\}\quad\text{s.t.}\quad(3). \tag{4}$$

The objective has three parts. The organization obtains ordinary current surplus from live span, the autonomy payoff on relationships operated AI-first, and the expected private value of successful succession net of assimilation resources. The autonomy payoff appears only once, on AI-first live relationships; assimilation resources are used only for recipients who enter the developmental pipeline.

The first simplification concerns human-first live relationships that do not lead to a developmental pipeline.

Lemma 1 (Minimal human-first live mass)**.** *Every optimal allocation can be replaced, holding (n,T,p) fixed, by an optimal allocation with $h=T$. If $g(a)>0$, every optimum satisfies $h=T$.*

The reason is direct. Holding (n,T,p) fixed, a human-first live relationship beyond those needed for the T developmental recipients produces no additional successor. Relative to AI-first execution, it weakly uses more senior capacity and, when $g(a)>0$, gives up the autonomy payoff. The organization therefore preserves human-first live work only where it supports a developmental pipeline.

We can consequently eliminate h. Define

$$\tau=\tau_A,\qquad \rho=\tau_H-\tau_A+\rho_R.$$

Here τ is the senior-capacity requirement common to the AI-first baseline live relationship, while ρ is the additional senior capacity required to convert one such live relationship into a complete developmental pipeline. The latter combines the extra senior requirement of preserving the human-first live interaction with the capacity required for the rank-completing counterfactual episode.

2. For example, in the two-dimensional isotropic case $G=pI_2$ and $\Omega=I_2$, the Gaussian criterion error implies $p\Delta$ has a chi-square distribution with two degrees of freedom. Hence $\chi(p)=1-\exp(-\overline{L}p/2)$, which is strictly increasing and concave. The scalar curvature conditions below are therefore exact for a transparent special case, not asserted for every multidimensional diagnostic geometry.

The reduced physical production set is

$$0 \leq T \leq n, \qquad \tau n + \rho T \leq B. \tag{5}$$

This constraint contains the central physical trade-off of the benchmark. Every live relationship uses τ units of senior capacity. Making one of those relationships developmental requires an additional ρ units.

Given T and p, successful successor output is

$$P = \chi(p)T. \tag{6}$$

The distinction between T and P is important. Developmental entry creates opportunities to reproduce judgment; assimilation determines how many of those opportunities become successful successors.

Conditional on operating a positive developmental pipeline, assimilation is chosen to maximize

$$A\chi(p) - \zeta c(p).$$

The first-order condition is

$$A\chi'(p^*) = \zeta c'(p^*). \tag{7}$$

Under Assumption 5, the objective is strictly concave, with positive derivative at $p = 0$ and negative derivative for sufficiently large p. Hence the solution is unique and interior for every $A > 0$, with $p^*(0) = 0$. Differentiating the first-order condition gives

$$\frac{dp^*}{dA} = \frac{\chi'(p^*)}{\zeta c''(p^*) - A\chi''(p^*)} > 0. \tag{8}$$

A more valuable successor therefore induces greater assimilation investment.

It is useful to collect this optimized developmental return in one object. Define

$$\mathcal{J}(A) = A\chi(p^*(A)) - \zeta c(p^*(A)). \tag{9}$$

The term $\mathcal{J}(A)$ is the optimized developmental value per pipeline recipient before accounting for the autonomy payoff forgone by keeping that relationship developmental. By the envelope theorem,

$$\mathcal{J}'(A) = \chi(p^*(A)) > 0.$$

A *pure autonomy change* raises $g(a)$ while holding A and the assimilation technology fixed. Such a change does not alter the assimilation problem, so $p^*_g = 0$. Pure autonomy will therefore change how many developmental opportunities the organization preserves, not the assimilation intensity within those that remain.

After optimizing assimilation, the organization chooses only (n, T):

$$U_\alpha(n, T, a) = [b + g(a)]n - C(n) + [\mathcal{J}(A) - g(a)]T \tag{10}$$

subject to equation (5).

The reduced objective exposes the autonomy mechanism. The same primitive $g(a)$ appears on both sides of the organizational allocation. It raises the return to live scale through $[b + g(a)]n$, while it lowers the net return to a developmental pipeline through $[\mathcal{J}(A) - g(a)]T$. These are not separate assumptions: both follow from the fact that preserving a developmental relationship means foregoing AI-first origination on that relationship.

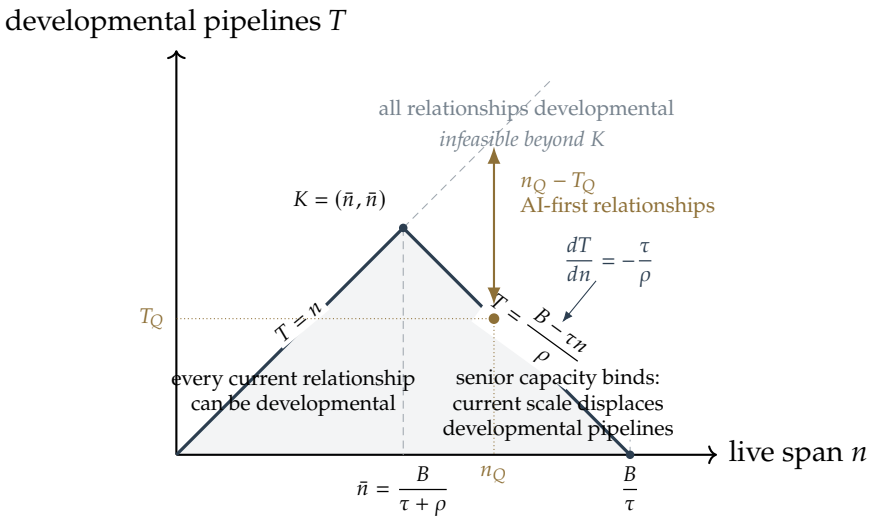


Figure 1: The developmental-pipeline frontier. Before the kink K, every live relationship can be developmental. Beyond K, senior capacity binds and the organization must trade developmental pipelines against current scale along $T = (B - \tau n)/\rho$: one additional unit of live span requires τ/ρ fewer developmental pipelines. At the representative over-span allocation Q, the gap $n_Q - T_Q$ is the mass of live relationships organized AI-first. Successful successor production on the frontier is $P = \chi(p^*)T$.

3.3 The developmental-pipeline frontier and organizational regimes

The reduced capacity constraint determines how much development can accompany a given level of current scale. At live span n, developmental entry cannot exceed the number of live relationships, $T \leq n$. It also cannot exceed the amount permitted by remaining senior capacity,

$$T \leq \frac{B - \tau n}{\rho}.$$

The relevant limit is therefore whichever of these two bounds is tighter.

Proposition 3 (Developmental-pipeline possibility frontier)**.** *A pair (n, T) is implementable in the reduced benchmark if and only if it satisfies equation (5). The upper frontier is*

$$T^{\max}(n) = \min\left\{n, \frac{B - \tau n}{\rho}\right\}, \qquad 0 \leq n \leq \frac{B}{\tau}, \tag{11}$$

with kink $\bar{n} = B/(\tau + \rho)$. At the optimized assimilation choice, successful succession on this frontier is $P^{\max}(n) = \chi(p^*)T^{\max}(n)$.

The frontier has two economically distinct segments. For $n \leq \bar{n}$, every live relationship can remain developmental and the upper frontier is $T = n$. Current scale can expand without sacrificing developmental entry.

Beyond the kink, $n > \bar{n}$, there is no longer enough senior capacity to keep every live relationship developmental. The frontier becomes

$$T = \frac{B - \tau n}{\rho},$$

with slope

$$\frac{dT}{dn} = -\frac{\tau}{\rho}.$$

One additional unit of live span therefore displaces τ/ρ developmental pipelines. This is the physical source of crowd-out in the model: once senior capacity binds on the declining part of the frontier, current scale and developmental entry compete directly for the same scarce input.

To characterize the organization's choice along this frontier, let

$$W = \mathcal{J}(A) - g$$

be the private net value of converting an AI-first relationship into a developmental pipeline after assimilation has been optimized. The developmental pipeline creates value $\mathcal{J}(A)$, but preserving it requires giving up the autonomy payoff g.

When $W > 0$, development has positive direct value before accounting for its use of scarce senior capacity, so at any chosen n the organization selects the maximum feasible T.

On the declining segment of the frontier, an additional unit of live span yields $b + g$ in current return but displaces τ/ρ pipelines, each worth W. Define the resulting marginal return to live span by

$$R_A = b + g - W\frac{\tau}{\rho}.$$

Proposition 4 (Organizational regimes)**.** *Suppose $W > 0$. Under Assumptions 4–5, the optimum is unique and belongs to exactly one regime:*

(i) ***Under-span.*** *If $C'(\bar{n}) > b + \mathcal{J}(A)$, then $n^* < \bar{n}$, $T^* = n^*$, and $C'(n^*) = b + \mathcal{J}(A)$.*

(ii) ***Balanced.*** *If $R_A \le C'(\bar{n}) \le b + \mathcal{J}(A)$, then $n^* = T^* = \bar{n}$.*

(iii) ***Over-span.*** *If $C'(\bar{n}) < R_A < C'(B/\tau)$, then $\bar{n} < n^* < B/\tau$,*

$$C'(n^*) = R_A, \qquad T^* = \frac{B - \tau n^*}{\rho} < n^*.$$

(iv) ***Exhausted development.*** *If $R_A \ge C'(B/\tau)$, then $n^* = B/\tau$ and $T^* = P^* = 0$.*

In every positive-development regime, $P^* = \chi(p^*(A))T^*$.

The regimes differ in whether senior capacity forces current production and development apart. In the under-span regime, the organization is small enough that every live relationship remains developmental. In the balanced regime, it sits at the kink: every live relationship is still developmental, but further expansion would require sacrificing development.

The over-span regime is different. The organization deliberately operates beyond the scale at which all live relationships can remain developmental. Some relationships are developmental and the rest are AI-first, so additional current scale directly displaces developmental pipelines. This is the region in which the autonomy rebound operates locally.

In the exhausted-development regime, no developmental pipeline remains. Importantly, exhaustion need not mean that development has negative direct value: even with $W > 0$, scarce senior capacity can be more valuable at the margin in current production. The quantitative illustration below makes this distinction explicit.

Below the kink, autonomy has no direct quantity margin because every live relationship can remain developmental. At the other extreme, sufficiently high autonomy can eliminate developmental entry altogether; Appendix A gives the exact exit and regime-incidence statements.

3.4 The autonomy rebound

We now isolate the effect of greater AI autonomy. The experiment raises the value of autonomous action origination while holding current action quality, verified diagnostic technology, assimilation technology, senior capacity, and all senior-time requirements fixed.

Suppose the organization is in the interior over-span regime. A higher g affects its choice in two reinforcing ways. It raises the current return to an AI-first live relationship, and it lowers $W = \mathcal{J}(A) - g$, the net value of preserving that relationship as developmental. Because the organization is already on the declining part of the developmental frontier, the resulting expansion of current scale must displace developmental pipelines.

Theorem 1 (Autonomy rebound holding senior time fixed)**.** *Suppose the unique optimum lies in the interior over-span regime on an open capability interval. Hold $b, B, \tau, \rho, A, \zeta, c, \chi$ fixed and suppose $g'(a) > 0$. Then $p^{*\prime}(a) = 0$ and*

$$n'(a) = \frac{g'(a)(1 + \tau/\rho)}{C''(n(a))} > 0, \tag{12}$$

$$T'(a) = -\frac{\tau}{\rho} n'(a) < 0, \tag{13}$$

$$P'(a) = \chi(p^*)T'(a) < 0. \tag{14}$$

Current operating surplus net of developmental-resource costs,

$$Y(a) = [b + g(a)]n(a) - C(n(a)) - [g(a) + \zeta c(p^*)]T(a),$$

strictly rises, with

$$Y'(a) = g'(a)[n(a) - T(a)] + A\chi(p^*)\frac{\tau}{\rho} n'(a) > 0. \tag{15}$$

The result holds at $\alpha = 1$ and on a nonempty open set of primitives.

The theorem changes the number of developmental opportunities, not their effectiveness. Pure autonomy leaves the assimilation problem unchanged, so p^* and the success probability $\chi(p^*)$ are constant. Successful succession falls because the organization operates fewer developmental pipelines, not because recipients in the remaining pipelines learn less effectively.

The result therefore has three linked quantity implications. Current live span rises; developmental entry falls because senior capacity binds; and successful successor production falls proportionally with T. At the same time, current operating surplus net of developmental-resource costs rises.

This is the autonomy rebound. Greater AI autonomy makes current execution more valuable and induces the organization to expand current activity. Once senior capacity

already limits how many relationships can remain developmental, that expansion reduces the organization's production of future independent decision makers.

The result is not driven by AI error, deterioration in assimilation, or a reduction in senior time. It also does not require incomplete appropriation of successor value: it holds at $\alpha = 1$. Even an organization that fully internalizes the continuation value of successful successors can optimally choose more current scale and less succession.

The quantitative illustration of Figure 2 makes the global organizational reallocation behind the local comparative static transparent. For clarity, normalize $B = \tau = \rho = \mathcal{J}(A) = 1$, let $C(n) = n^2$ and $b = 3/2$, and vary only the autonomy value g, holding A, p^*, and hence $\chi(p^*)$ fixed. The exact regime boundaries are $g/\mathcal{J} = 1/4$ and $g/\mathcal{J} = 3/4$, and the optimal policies are

$$n^*(g) = \begin{cases} \frac{1}{2}, & 0 \le g \le \frac{1}{4}, \\ \frac{1}{4} + g, & \frac{1}{4} < g < \frac{3}{4}, \\ 1, & g \ge \frac{3}{4}, \end{cases} \qquad T^*(g) = \begin{cases} \frac{1}{2}, & 0 \le g \le \frac{1}{4}, \\ \frac{3}{4} - g, & \frac{1}{4} < g < \frac{3}{4}, \\ 0, & g \ge \frac{3}{4}. \end{cases}$$

For $g/\mathcal{J} \le 1/4$, every live relationship can remain developmental, so greater autonomy has no quantity effect. Once $g/\mathcal{J} > 1/4$, the organization enters the over-span region and senior capacity binds on the declining developmental frontier. Under this normalization, each additional unit of live span displaces one developmental pipeline, while assimilation and the probability of success within every preserved pipeline remain unchanged.

At $g/\mathcal{J} = 1/2$, current live span is 50 percent above its low-autonomy level while successful successor production is 50 percent below it. At $g/\mathcal{J} = 3/4$, live span has doubled and development is exhausted.

The last threshold illustrates an important feature of the mechanism. Development disappears while

$$\mathcal{J} - g = \frac{1}{4} > 0.$$

The organization therefore does not stop producing successors because a developmental relationship has become directly worthless. It stops because the senior capacity required to preserve development is more valuable at the margin in supporting the larger AI-first organization.

3.5 What is specific to autonomy?

A decline in succession does not by itself identify the autonomy mechanism. Ordinary productivity can also expand organizational span; senior-time saving can relax the capacity constraint; diagnostic progress can change what development requires; and greater successor scarcity can raise the return to development. The channels differ in where they enter the organizational problem.

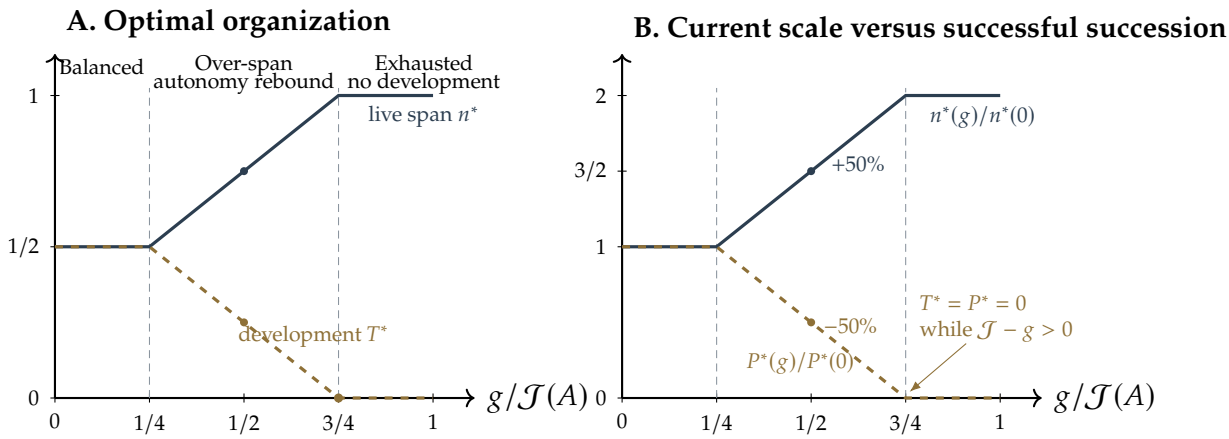


Figure 2: The autonomy rebound over the full organizational optimum. The normalization is chosen for transparency rather than to match empirical moments. Panel A plots current live span and developmental entry. Panel B normalizes current live span and successful successor production to one at $g = 0$. Before the capacity kink, autonomy has no quantity effect because every live relationship can remain developmental. In the over-span region, higher autonomy expands current scale and reduces developmental pipelines while p^* and $\chi(p^*)$ remain fixed. Development is exhausted at $g/\mathcal{J} = 3/4$ even though its direct net value remains positive, $\mathcal{J} - g = 1/4$.

Primitive change	**Direct organizational incidence**
Ordinary productivity, $b \uparrow$	Raises the return to current span; it does not directly change the marginal value of preserving a developmental pipeline.
Autonomous origination, $g \uparrow$	Raises the return to AI-first live execution and lowers the net value of preserving the diagnostic process on a developmental relationship.
Senior-time saving, $\tau \downarrow$ or $\rho \downarrow$	Relaxes the physical developmental-pipeline frontier at a fixed live span.
Diagnostic progress	Changes the verified information supplied by AI and therefore the residual human geometry, developmental design, or success probability.
Successor scarcity, $V \uparrow$	Raises $A = \alpha V$, increasing the return to developmental entry and, under the benchmark second-order condition, assimilation intensity.

The distinction between ordinary productivity and autonomy is especially useful. Holding assimilation at its optimum, ordinary productivity b raises the return to current span but does not directly change the marginal value of preserving a developmental pipeline:

$$\frac{\partial^2 U_\alpha}{\partial b \, \partial T} = 0.$$

Autonomy g, by contrast, acts on both margins:

$$\frac{\partial^2 U_\alpha}{\partial g \, \partial n} = 1, \qquad \frac{\partial^2 U_\alpha}{\partial g \, \partial T} = -1.$$

It raises the return to current AI-first scale and simultaneously raises the opportunity cost of preserving development.

On an interior over-span branch with B, τ, ρ fixed,

$$\frac{dn}{db} = \frac{1}{C''(n)}, \qquad \frac{dn}{dg} = \frac{1+\tau/\rho}{C''(n)}, \qquad -\frac{dT}{dg} = \left(1+\frac{\tau}{\rho}\right)\left(-\frac{dT}{db}\right).$$

Thus an equally sized autonomy shift has an additional local incidence because the same primitive that rewards AI-first execution also makes developmental relationships more costly to preserve.

Senior-time saving works differently again. On the positive over-span frontier and at fixed n,

$$T_\tau = -\frac{n}{\rho} < 0, \qquad T_\rho = -\frac{T}{\rho} < 0.$$

A reduction in either τ or ρ therefore expands feasible development. AI that saves senior time can offset the crowd-out generated by greater autonomy.

These distinctions matter for interpretation. Observing that AI expands organizational span does not by itself imply that succession must fall. The answer depends on whether the underlying technological change mainly raises autonomy value, saves senior time, changes diagnostic information, improves assimilation, or moves several of these margins together.

General capability changes

To keep these channels separate, allow capability a to affect several primitives simultaneously. On an interior over-span branch, optimal live span satisfies

$$C'(n(a)) = b(a) + g(a) - [\mathcal{J}(A(a)) - g(a)]\frac{\tau(a)}{\rho(a)}. \tag{16}$$

The expression separates the main forces affecting scale. Higher b or g raises the return to current activity; higher A raises the value of development through $\mathcal{J}(A)$; and lower τ or ρ changes the senior-capacity technology.

For positive variables define

$$\varepsilon_n = \frac{d\log n}{d\log a}, \qquad \varepsilon_L = -\frac{d\log \tau}{d\log a}, \qquad \varepsilon_R = -\frac{d\log \rho}{d\log a},$$
$$\varepsilon_\chi = \frac{d\log \chi(p^*)}{d\log a}, \qquad \ell = \frac{\tau n}{B} \in (0,1).$$

Here ε_n is the elasticity of live span; ε_L and ε_R measure proportional reductions in the live and incremental developmental senior requirements; ε_χ measures the response of success within a developmental pipeline; and ℓ is the share of senior capacity absorbed by the baseline live requirement.

Log differentiation of the physical frontier gives

$$\frac{d\log T}{d\log a} = \varepsilon_R - \frac{\ell}{1-\ell}(\varepsilon_n - \varepsilon_L). \tag{17}$$

The decomposition has a direct interpretation. A reduction in the incremental developmental requirement, $\varepsilon_R > 0$, raises developmental entry directly. A reduction in the senior requirement of live production, $\varepsilon_L > 0$, protects capacity that can be used

for development. Expansion of live span, $\varepsilon_n > 0$, works in the opposite direction. Its crowd-out effect is amplified by

$$\frac{\ell}{1-\ell},$$

so the same proportional expansion in span has a larger effect when current activity already absorbs a large share of senior capacity.

Because $P = \chi T$, successful succession satisfies

$$\frac{d \log P}{d \log a} = \varepsilon_\chi + \varepsilon_R - \frac{\ell}{1-\ell}(\varepsilon_n - \varepsilon_L). \tag{18}$$

Successful succession therefore adds a fourth margin: changes in the probability that a developmental pipeline produces an independent successor. The equation separates changes in assimilation success, developmental senior requirements, live senior requirements, and endogenous organizational span.

For a pure autonomy improvement,

$$\varepsilon_\chi = \varepsilon_L = \varepsilon_R = 0.$$

The success response then comes entirely through span-induced crowd-out of developmental entry. This is the elasticity version of the autonomy rebound.

For two allocations on the positive over-span segment generated by a pure autonomy change, with B, τ, ρ, and χ fixed and $\ell_0 = \tau n_0 / B$,

$$\frac{T_1 - T_0}{T_0} = \frac{P_1 - P_0}{P_0} = -\frac{\ell_0}{1-\ell_0}\frac{n_1 - n_0}{n_0}.$$

This finite-change identity is exact on the positive over-span segment. It makes the quantity trade-off transparent: once the organization is capacity constrained, the percentage contraction in development and successful succession is mechanically tied to the percentage expansion of current scale. Figure 2 shows how that segment fits into the full organizational optimum.

3.6 Other capability changes

The pure-autonomy comparative static isolates one channel. Actual AI progress can move several primitives at the same time. It may raise ordinary productivity, raise the value of autonomous origination, reduce senior-time requirements, provide new verified diagnostic information, or change the effectiveness of assimilation. These effects can reinforce or offset one another.

Equations (16)–(18) keep those margins separate. The autonomy rebound is the response to a particular experiment: greater value from AI-originated current action holding current decision quality, verified diagnostic technology, assimilation technology, senior capacity, and senior-time requirements fixed. It is therefore not a claim that every improvement in AI capability must reduce succession. Time saving, diagnostic progress, or improvements in developmental effectiveness can offset that channel.

Appendix A records the full derivative and also shows that incomplete human-first exposure may retain positive developmental value without changing the mode reduction or the autonomy rebound, provided that value remains below the autonomy gain. The appendix additionally allows the organization to purchase senior capacity at convex cost.

4 Diagnostic Geometry and Developmental Design

The benchmark determines how much development accompanies current scale. In a multidimensional domain, the organization must also decide which diagnostic experiences to preserve and how intensively to assimilate them.

4.1 The residual diagnostic basis

Fix recipient i and a q-dimensional maintained criterion ϕ. Candidate human decision contexts are indexed by e, with target direction $z_e \in \mathbb{R}^q$, and the complete menu together with the verified AI diagnostic span $\mathcal{A}_i(a)$ can identify ϕ. Preserving context e in a human-first, source-diagnosed workflow supplies the restriction $\phi' z_e$ but forgoes autonomy value $g_{ie}(a) > 0$. Candidate preserved episodes are initially normalized to use the same unit of senior diagnostic capacity. Throughout this subsection, $\mathcal{A}_i(a)$ is the verified diagnostic span already available to recipient i. Producing or validating that span may itself use resources; the residual-basis problem conditions on it and characterizes the allocation of the human diagnostic work that remains.

Let $r_i^A = \dim \mathcal{A}_i(a)$ and define the family of residual human bases

$$\mathfrak{B}_i^H(\mathcal{A}_i(a)) = \left\{H : |H| = q - r_i^A, \quad \mathcal{A}_i(a) + \operatorname{span}\{z_e : e \in H\} = \mathbb{R}^q\right\}.$$

When this family is nonempty, define

$$K_i^H(a) = \min_{H \in \mathfrak{B}_i^H(\mathcal{A}_i(a))} \sum_{e \in H} g_{ie}(a). \tag{19}$$

Proposition 5 (Residual diagnostic basis)**.** *Suppose $\mathfrak{B}_i^H(\mathcal{A}_i(a)) \neq \varnothing$ and all candidate human-first opportunity costs are positive. Under the equal-capacity normalization above, conditional on exact diagnostic coverage and frictionless developmental conversion, the current-surplus-maximizing workflow preserves human-first provenance on a minimum-cost set $H^* \in \mathfrak{B}_i^H(\mathcal{A}_i(a))$ and assigns every human episode redundant relative to $\mathcal{A}_i(a) + \operatorname{span}\{z_e : e \in H^*\}$ to AI-first execution. Every exact-coverage workflow incurs human autonomy cost at least $K_i^H(a)$. If $\mathcal{A}_i(a) = \mathbb{R}^q$, no human diagnostic episode is required for coverage.*

The human residual is thus a complement to diagnostic information already supplied by other technologies. A large human-task share can fail to identify the criterion when its cases are redundant, while a small share can suffice when those cases complete the missing directions. The spanning argument is elementary; the organizational implication is not: human-task shares are not sufficient statistics for succession, and AI progress can reallocate human decision making toward technologically unchanged contexts.

Corollary 1 (Diagnostic progress)**.** *Hold the candidate human contexts and their positive autonomy costs fixed, and write $K_i^H(\mathcal{A})$ for the residual basis cost at verified AI span $\mathcal{A}$. If*

$$\mathcal{A}_i^0 \subseteq \mathcal{A}_i^1,$$

then the number of human diagnostic directions required for exact coverage weakly falls and

$$K_i^H(\mathcal{A}_i^1) \leq K_i^H(\mathcal{A}_i^0).$$

If the inclusion is strict, the residual dimension and the minimum cost both fall strictly. In particular, one new independent verified AI direction reduces the cardinality of a residual human basis by one.

The corollary concerns the extensive margin of diagnostic coverage. Verified AI progress can also change the geometry and developmental effectiveness of the record without changing its rank. The next subsection makes that margin explicit.

4.2 Endogenous assimilation and developmental design

The residual-basis result is the frictionless developmental benchmark. More generally, all verified diagnostic information—human or AI—must be converted into durable human capability before a recipient can later decide independently. Let $\mathcal{E}_i^A(a)$ be a finite set of verified AI diagnostic measurements delivered to a developmental recipient, with directions z_{ij} spanning $\mathcal{A}_i(a)$. Let H be a chosen set of human diagnostic episodes. Write

$$\mathcal{R}_i(H, a) = \mathcal{E}_i^A(a) \cup H$$

for the complete verified record, where every item $j \in \mathcal{R}_i(H, a)$ has a known target direction $z_{ij} \in \mathbb{R}^q$.

The objective diagnostic restriction is exact when the episode occurs. Durable internal encoding is imperfect. For an item with assimilation precision $p_{ij} > 0$, let

$$\widetilde{s}_{ij} = z_{ij}'\phi + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0, p_{ij}^{-1}),$$

independently across encoded items; $p_{ij} = 0$ means that the item is omitted from the durable record. The noise is therefore not error in the senior or AI diagnostic message. It is imperfect durable conversion of an exact contextual lesson.

The durable information matrix is

$$G_i(H, p, a) = \sum_{j \in \mathcal{R}_i(H,a)} p_{ij} z_{ij} z_{ij}'. \tag{20}$$

When $G_i > 0$, efficient reconstruction satisfies

$$\widehat{\phi}_i - \phi \sim N(0, G_i^{-1}).$$

For the performance loss in equation (2), define the probability of successful succession from one developmental pipeline as

$$\chi_i(H, p, a) = \Pr\left\{ (\widehat{\phi}_i - \phi)' \Omega_i (\widehat{\phi}_i - \phi) \le \overline{L}_i \;\middle|\; G_i(H, p, a) \right\}.$$

If $G_i(H, p, a)$ is singular, set $\chi_i(H, p, a) = 0$. This hard zero implements the constitutive coverage requirement in the definition of successful succession; it is not a claim that a rank-deficient learner could never happen to perform well on a finite realized test. Cognitive effort cannot repair a missing durably encoded target direction in the maintained succession benchmark. Conditional on coverage, more precise encoding can raise the probability of meeting the performance standard.

Let $c_i(H, p, a) = \sum_{j \in \mathcal{R}_i(H,a)} c_{ij}(p_{ij})$ be the real assimilation-resource cost per developmental recipient. For each record item, c_{ij} is continuous on $[0, \infty)$, continuously differentiable and strictly increasing for $p_{ij} > 0$, convex, satisfies $c_{ij}(0) = 0$, and is coercive: $c_{ij}(p) \to \infty$ as $p \to \infty$. These costs apply to human and AI diagnostic information symmetrically conditional on content, although their functions may differ if one source is easier to encode. Let $K_i^C(H, a) \ge 0$ be the current surplus forgone by

preserving the chosen human diagnostic episodes rather than using AI-first execution, with $K_i^C(\varnothing, a) = 0$, and let $R_i(H) \geq 0$ be their additional senior-capacity requirement per developmental recipient, with $R_i(\varnothing) = 0$. For organization i, let $b_i, g_i^L(a), C_i, \tau_i, B_i$ denote the analogues of the scalar live-production primitives, and let $\zeta_i > 0$ scale assimilation-resource costs. The integrated organizational problem is

$$\max_{n_i, T_i, H, p} \left\{ [b_i + g_i^L(a)]n_i - C_i(n_i) + [A\chi_i(H, p, a) - K_i^C(H, a) - \zeta_i c_i(H, p, a)]T_i \right\} \tag{21}$$

subject to

$$0 \leq T_i \leq n_i, \qquad \tau_i n_i + R_i(H)T_i \leq B_i. \tag{22}$$

Successful successor output is $P_i = \chi_i(H, p, a)T_i$.

This formulation keeps senior scarcity in the primal problem rather than inserting a shadow price into an exogenous cost index. At an interior optimum with multiplier λ_i on equation (22), the within-pipeline design solves

$$\max_{H, p} \left\{ A\chi_i(H, p, a) - K_i^C(H, a) - \zeta_i c_i(H, p, a) - \lambda_i R_i(H) \right\}.$$

For any continuously chosen precision $p_{ij} > 0$, the first-order condition is

$$A \frac{\partial \chi_i(H, p, a)}{\partial p_{ij}} = \zeta_i c'_{ij}(p_{ij}). \tag{23}$$

The organization therefore ranks experiences by four margins: their contribution to successful succession, their current autonomy opportunity cost, their assimilation-resource cost, and their use of scarce senior capacity.

Proposition 6 (Coverage before assimilation)**.** *For the developmental technology above, any design whose durably encoded directions fail to span $\mathbb{R}^q$ has $\chi_i(H, p, a) = 0$. A design with full span and strictly positive encoding precision on a spanning subset has $G_i > 0$ and therefore a strictly positive probability of meeting any performance standard $\overline{L}_i > 0$. Conditional on an interior precision choice, equation* (23) *characterizes optimal assimilation.*

The proposition preserves the hierarchy of the paper: coverage first, assimilation second. In the frictionless developmental limit, covered designs produce success with probability one and assimilation costs vanish. Under the equal-capacity normalization and positive human opportunity costs, equation (21) then collapses to Proposition 5.

Proposition 7 (Developmental dominance and design upgrading)**.** *Fix organization i, capability a, future-task matrix $\Omega_i > 0$, and performance standard $\overline{L}_i > 0$. Consider two fully specified full-coverage developmental designs (H_k, p_k), $k \in \{0, 1\}$, with durable information matrices $G_k = G_i(H_k, p_k, a) > 0$ and success probabilities $\chi_k = \chi_i(H_k, p_k, a)$.*

(i) *If $G_1 \geq G_0$, then $\chi_1 \geq \chi_0$. If $G_1 \succ G_0$, then $\chi_1 > \chi_0$.*

(ii) *Suppose $H_1 \neq H_0$, $R_i(H_1) = R_i(H_0)$, $\chi_1 > \chi_0$, and let the difference in direct per-pipeline developmental cost be*

$$\Delta_i^c \equiv \left[K_i^C(H_1, a) + \zeta_i c_i(H_1, p_1, a) \right] - \left[K_i^C(H_0, a) + \zeta_i c_i(H_0, p_0, a) \right] > 0.$$

Conditional on operating a positive developmental pipeline and choosing between these two designs, design 1 is preferred if and only if

$$A \geq A_i^\dagger \equiv \frac{\Delta_i^c}{\chi_1 - \chi_0}. \tag{24}$$

At equality the organization is indifferent between the two designs.

The probability ordering is a standard Gaussian comparison; the economic content is the organizational ranking. A more informative durable record can be costlier yet preferred once successor value exceeds the threshold. Because $A = \alpha V$, scarcity can therefore induce an upgrade in developmental architecture as well as greater assimilation within a fixed design. Equal senior requirements make equation (24) global for this fixed pair. With unequal senior use, the comparison adds the endogenous term $\lambda_i[R_i(H_1) - R_i(H_0)]$; if assimilation precision is reoptimized as A changes, the switch is instead determined by the designs' optimized value functions and need not retain the closed-form threshold.

Loewner dominance is deliberately a partial order. When designs redistribute durable information across directions, their ranking should depend on the future-task environment Ω_i, assimilation costs, senior requirements, and successor value rather than on a forced one-dimensional quality index. Positive assimilation frictions can therefore make a diagnostically redundant episode worth retaining.

Proposition 8 (Benchmark assimilation and scarcity)**.** *In the scalar fixed-design benchmark of Section 3, the unique interior assimilation choice satisfies equation (7) and $dp^*/dA > 0$. Hence, whenever $\alpha > 0$, a rise in the scarcity value V raises assimilation intensity and the success probability $\chi(p^*)$, holding the diagnostic design fixed.*

This intensive-margin response is deliberately modest. In a general multidimensional design, developmental pipelines, individual precisions, and the discrete diagnostic configuration can substitute for one another, so the theory does not require every margin to rise with V. The aggregate object that is monotone in successor value is optimal successful-successor supply.

4.3 AI placement and substitution

The exact exchange result below is the frictionless developmental benchmark. Hold $\mathcal{A}_i$ fixed and write $\mathfrak{B} = \mathfrak{B}_i^H(\mathcal{A}_i)$. Let B^* be the unique least-cost residual basis. For episode $e \in B^*$, define the gap to the cheapest residual basis that excludes it,

$$\Delta_e = \min_{B \in \mathfrak{B}: e \notin B} \left[\sum_{j \in B} g_j - \sum_{j \in B^*} g_j \right],$$

with $\Delta_e = +\infty$ if every residual basis contains e.

Theorem 2 (Basis substitution and bottlenecks)**.** *Start from a unique least-cost residual basis B^* for a fixed verified AI diagnostic span and increase only the autonomy value of episode e by $t \geq 0$.*

(i) *If $e \notin B^*$, the optimal residual basis and its cost do not change.*

(ii) *If $e \in B^*$ and $\Delta_e < \infty$, then B^* remains uniquely optimal for $0 \leq t < \Delta_e$. At $t = \Delta_e$, a basis excluding e ties; for $t > \Delta_e$, every minimum-cost basis excludes e. The residual basis cost rises by t before the switch and by Δ_e thereafter.*

(iii) *If $\Delta_e = +\infty$, episode e belongs to every residual basis. Conditional on exact diagnostic coverage it remains human-first for every t, and the residual basis cost rises one-for-one with t.*

When AI becomes more valuable on a protected but replaceable type of case, the organization can move the recipient's independent decision to another type whose

technology has not changed. This is basis substitution. With endogenous assimilation, the same change can also alter precision choices or make a non-basis design preferable; the exact thresholds above remain the clean frictionless benchmark.

4.4 Diagnostic bottlenecks and measurement noise

A protected direction that belongs to every residual basis is a diagnostic bottleneck. AI may perform its operative action perfectly, yet the organization retains human-first provenance there because no other available context completes the missing diagnostic direction. Progress on that task then raises the opportunity cost of coverage rather than displacing the diagnostic role to another case.

The assimilation friction above concerns durable conversion of an exact verified diagnostic record. Appendix E separately allows the target-linked record itself to be noisy and shows that weakest-direction diagnostic precision is governed by the smallest eigenvalue of its information matrix. The two frictions are conceptually distinct: measurement quality determines what the record contains, while assimilation determines how effectively that content becomes durable human capability.

5 The Market for Reproduced Judgment

The organizational problem leaves one price outside the firm: V, the continuation value of a successful successor. For the equilibrium closure, we work within a single maintained judgment domain and measure successful succession in deployment-equivalent units. Organizations may differ in their diagnostic technologies, future-task environments, performance standards, and costs of producing such units. The aggregation assumes that, once counted as one unit of successful-successor output P_i, a unit provides the same marginal service to the continuation sector. This scalar normalization isolates the equilibrium scarcity of reproduced judgment; heterogeneous judgment domains or successor types that are imperfect substitutes would instead require type-specific quantities and scarcity values.

5.1 Organizational successor supply

Consider a continuum of organizations $i \in [0, 1]$ with common appropriation share $\alpha \in (0, 1]$. Organization i chooses

$$x_i = (n_i, T_i, H_i, p_i)$$

from the feasible set defined by its live technology and equation (22). Its successful successor output is

$$P_i(x_i, a) = \chi_i(H_i, p_i, a)T_i.$$

Given scarcity value V, the organization solves

$$\max_{x_i} \{F_i(x_i, a) + \alpha V P_i(x_i, a)\}, \tag{25}$$

where F_i collects current operating surplus net of autonomy opportunity costs and real assimilation costs. Equation (21) is the explicit benchmark form. Let $P_i(V, a)$ denote optimal successful-successor supply.

Revealed preference gives the key monotonicity without requiring each internal choice to move in the same direction. For $V_2 > V_1$, optimality at the two values implies

$$(V_2 - V_1)[P_i(V_2, a) - P_i(V_1, a)] \geq 0.$$

Thus successful-successor supply is weakly increasing in V, even though developmental entry, assimilation intensity, and diagnostic design can substitute for one another.

Let aggregate successful succession be

$$P^{\text{agg}} = \int_0^1 P_i(V, a)\, di,$$

and define aggregate organizational supply

$$\mathcal{S}(V, a) = \int_0^1 P_i(V, a)\, di.$$

A competitive continuation sector obtains present-value deployment surplus $\Phi(P^{\text{agg}})$, where $\Phi' > 0$ and $\Phi'' < 0$. Its marginal value

$$V = \Phi'(P^{\text{agg}})$$

closes the market.

Proposition 9 (Unique aggregate equilibrium)**.** *Suppose aggregate organizational supply $\mathcal{S}(V, a)$ is continuous and bounded on the feasible successor interval, and Φ' is continuous and strictly decreasing there. Then an aggregate equilibrium*

$$P^{\text{agg}} = \mathcal{S}(V, a), \qquad V = \Phi'(P^{\text{agg}}) \tag{26}$$

exists and is unique.

The result needs only the monotonicity of optimal successful-successor supply in its value. The internal developmental technology can include discrete design changes, endogenous assimilation, and developmental entry or exit.

5.2 Scarcity feedback

Consider two capability states with organizational supply schedules $\mathcal{S}^0(V)$ and $\mathcal{S}^1(V)$, and corresponding equilibria $(P^{\text{agg},0}, V^0)$ and $(P^{\text{agg},1}, V^1)$.

Theorem 3 (Scarcity feedback)**.** *If*

$$\mathcal{S}^1(V) \leq \mathcal{S}^0(V) \qquad \text{for every relevant } V,$$

then $P^{\text{agg},1} \leq P^{\text{agg},0}$ and $V^1 \geq V^0$. If the inward shift is strict at the initial equilibrium, $\mathcal{S}^1(V^0) < \mathcal{S}^0(V^0)$, then both inequalities are strict.

At a differentiable equilibrium, for a scalar capability shifter a,

$$\frac{dP^{\text{agg}}}{da} = \frac{\mathcal{S}_a}{1 - \mathcal{S}_V \Phi''(P^{\text{agg}})}. \tag{27}$$

Hence the aggregate quantity response has the same sign as the fixed-V organizational supply shift; if $\mathcal{S}_a < 0$, then $dV/da > 0$, and scarcity attenuates the contraction whenever $\mathcal{S}_V > 0$.

At an equilibrium with positive P^{agg} and V, define

$$\eta = -\frac{d \log V}{d \log P^{\text{agg}}} > 0, \qquad \varepsilon_{\mathcal{S},V} = \frac{\partial \log \mathcal{S}}{\partial \log V} \geq 0.$$

Equation (27) becomes

$$\frac{d \log P^{\text{agg}}}{d \log a} = \frac{\left.\frac{\partial \log \mathcal{S}}{\partial \log a}\right|_V}{1 + \eta \varepsilon_{\mathcal{S},V}}. \tag{28}$$

Market clearing therefore cannot reverse a benchmark contraction. The attenuation factor has a direct sufficient-statistic interpretation: if $\eta \varepsilon_{\mathcal{S},V}$ equals 1/2, 1, or 2, market clearing leaves respectively two-thirds, one-half, or one-third of the fixed-V contraction in aggregate successful succession. These values illustrate the multiplier rather than calibrate it. Scarcity raises the return to producing successful successors. In the fixed-design benchmark, Proposition 8 identifies one intensive response: higher V raises assimilation effort and success probability among developmental recipients who remain in the pipeline. Proposition 7 identifies a second, discrete response: when two designs use the same senior capacity and one is both more effective and more costly, a sufficiently high successor value makes the organization willing to upgrade the developmental architecture itself.

5.3 Diagnostic exposure and where succession moves

The general model permits organizations to respond to a targeted AI change through developmental entry, assimilation, or diagnostic design. A clean cross-organizational assignment result survives in the frictionless residual-basis benchmark, where $\chi = 1$, assimilation costs vanish, all admissible residual bases use the same senior requirement ρ_i, and the per-recipient current opportunity cost is $K_i(a) = K_i^H(a)$. The firm problem then reduces to

$$\max_{0 \leq T_i \leq n_i,\ \tau_i n_i + \rho_i T_i \leq B_i} \left\{ [b_i + g_i^L(a)] n_i - C_i(n_i) + [\alpha V - K_i(a)] T_i \right\}, \tag{29}$$

with $P_i = T_i$.

Between two equilibria generated by a change that affects this benchmark only through K_i, let

$$\Delta V = V^1 - V^0, \qquad \Delta K_i = K_i^1 - K_i^0.$$

Theorem 4 (Diagnostic exposure and reallocation)**.** *In the frictionless residual-basis benchmark, consider a targeted AI change that weakly raises K_i for every organization and leaves the other organizational primitives unchanged. Suppose the induced organizational supply shift is strict at the initial equilibrium. Then aggregate successful succession falls and $\Delta V > 0$. Moreover:*

(i) *Organization i weakly contracts succession if $\Delta K_i > \alpha \Delta V$, and weakly expands it if $\Delta K_i < \alpha \Delta V$, with strict inequalities wherever its policy is strictly responsive.*

(ii) *Before any basis switch, if a common increase $t > 0$ in the autonomy value of direction e raises K_i by t for exposed organizations and leaves K_i unchanged for the others, then $0 < \alpha \Delta V < t$. Exposed organizations weakly contract succession and unexposed organizations weakly expand it, with strict responses where policies are responsive.*

The theorem gives the clean comparative-advantage result in the benchmark where diagnostic geometry maps directly into a residual basis cost. With endogenous assimilation, the same economic force operates through the optimized successor-supply function, but one should not sign every internal margin separately: a firm can alter the number of developmental pipelines, assimilation intensity, or the diagnostic design itself.

5.4 Private and social value of succession

With full appropriation, $\alpha = 1$, payments in the deployment market are transfers and the competitive allocation is constrained efficient conditional on maintained targets and organizational technologies; Appendix G gives the planner argument. When $\alpha < 1$, organizations underweight the continuation value created by successful successors.

Proposition 10 (Appropriation, span, and succession)**.** *Fix V, capability, and the scalar benchmark architecture. Relative to $\alpha = 1$, an organization with $\alpha < 1$ chooses a weakly lower successful-successor mass. If both optima are interior over-span, the private organization chooses higher live span, lower developmental-pipeline mass, lower assimilation intensity, and lower successful succession. If both are under-span, it chooses lower live span, lower developmental entry, lower assimilation intensity, and lower successful succession.*

What is underweighted is successful future capability, not human employment or contact. The autonomy rebound itself can occur at $\alpha = 1$, so incomplete appropriation is a separate margin.

At a common feasible allocation, consider revealing AI before the recipient's independent assessment rather than after it. Let $G > 0$ be the current-surplus gain from AI-first timing and $\Delta P > 0$ the resulting loss of successful successors.

Proposition 11 (AI timing under incomplete appropriation)**.** *At fixed common quantities and developmental technologies, the private organization chooses AI-first timing while the constrained-efficient planner chooses human-first timing if and only if*

$$\alpha V \Delta P < G < V \Delta P. \tag{30}$$

The inequality is a fixed-allocation comparison. With endogenous span, developmental entry, assimilation, and diagnostic design, private and planner allocations must be compared using their separately optimized value functions rather than extending equation (30) mechanically.

6 Measuring the Reproduction of Judgment

The theory changes what must be observed to distinguish current production from successful succession.

6.1 What conventional observables miss

Proposition 12 (Conventional observables do not identify succession)**.** *There exist pairs of workflows with the same live span, final-action agreement, total AI quantity, scheduled human review hours, and human-in-the-loop share but different successful-successor masses. Such pairs arise when AI revelation timing, verified target linkage, recipient-level coupling, target-relevant contextual span, or assimilation resources differ.*

The proposition is an observational-equivalence result. Current output, task assignment, human contact, review, and AI exposure can all be held fixed while recipients receive records with different coverage or developmental effectiveness. Measuring succession therefore requires information about the record left with the recipient and what became of it, not only the inputs used to produce the current action.

6.2 Existing evidence and empirical predictions

Existing evidence already separates current AI-assisted performance from later independent capability in ways that motivate this measurement. Dell'Acqua et al. (2026) document sizable effects of generative AI on knowledge-worker performance and quality in consulting tasks. In a randomized consulting reskilling experiment, Wiles et al. (2024) find large gains on unfamiliar data-science tasks with ChatGPT but no corresponding improvement on subsequent technical questions without it. In high-school mathematics, Bastani et al. (2025) find that unrestricted GPT-4 raises assisted practice performance while reducing later unaided performance relative to no AI; learning-oriented guardrails largely eliminate the loss. These settings concern technical skill and education rather than the maintained judgment criterion, but they show why current performance and durable independent capability must be measured separately.

Timing evidence points to the workflow margin. Yin et al. (2025) compare physicians who receive AI advice after an initial diagnosis with physicians who receive advice concurrently with the case. Ex-post advice improves current diagnostic accuracy and calibration and induces more thorough information processing and engagement with the AI rationale. The study measures current decision making rather than later acquisition, so it establishes the relevance of decision precedence without testing succession directly.

Other deployments show that AI can improve learning. In customer support, Brynjolfsson, Li, and Raymond (2025) find productivity gains and evidence of durable learning, especially among less experienced workers. The theory accommodates that possibility: AI can transmit useful information, preserve an independent attempt, or itself provide verified diagnostic measurements. The predicted developmental effect therefore depends on workflow architecture, diagnostic content, and assimilation rather than on AI exposure alone.

Three predictions organize the empirical content.

(i) **Decision provenance.** Holding cases and eventual AI access fixed, preserving an attributable human assessment before AI revelation should improve later independent performance when the source-linked diagnostic channel is operative, even if current action quality is unchanged.

(ii) **Diagnostic design and assimilation.** A targeted increase in autonomy value on a protected but replaceable direction should move human-first decisions toward substitute contexts in the frictionless basis benchmark. If AI instead supplies a verified diagnostic direction, human provenance should fall where that information makes the human record redundant. Conditional on the same diagnostic design, greater developmental resources should raise the probability of meeting the independent-performance standard; conditional on resources, records with stronger coverage should outperform records with missing or weakly encoded directions.

(iii) **Equilibrium sorting and developmental upgrading.** A capability change that shifts successful-successor supply inward should raise the scarcity value of

reproduced judgment. More exposed organizations should contract succession relative to less exposed organizations in the residual-basis benchmark. The higher scarcity value should raise assimilation intensity among fixed-design pipelines under the benchmark second-order condition and, in pairwise comparisons covered by Proposition 7, can justify switching to a costlier but developmentally dominant architecture.

Researchers therefore need the timing of AI revelation relative to an attributable human assessment, the target linkage and recipient-specific delivery of diagnostic feedback, the contextual directions covered by the record, developmental resources devoted to assimilation, later independent performance, live span, and senior load. Human-in-the-loop shares and review hours can be large while concentrated on diagnostically redundant cases.

6.3 Identification designs

A direct design randomizes AI revelation timing. Recipients receive the same cases and eventual AI access, but one arm requires a sealed assessment before the AI action is revealed. Source feedback and total scheduled contact can be held fixed or separately randomized. Later independent performance on differentiated target contexts then provides an observable measure of one component of successful succession rather than of immediate accuracy. Recording developmental time or other assimilation resources separates the information generated by the workflow from the effort used to internalize it.

A second design exploits or randomizes task-specific AI improvement. For a replaceable protected direction, human-first decisions should migrate toward substitute contexts whose technology did not change in the residual-basis benchmark. For a bottleneck, the human-first decision should remain on that context conditional on preserving coverage while its opportunity cost rises. When AI begins supplying an independently verified measurement of the maintained target, human provenance should disappear only along directions made redundant by the new verified span. These cross-task responses distinguish diagnostic substitution from a generic change in human labor demand.

For the over-span rebound, equations (17)–(18) organize measurement. A pure autonomy improvement has $\varepsilon_\chi = \varepsilon_L = \varepsilon_R = 0$, so successful succession moves through endogenous span and developmental entry. More general AI improvements can also change assimilation success or senior requirements. A fixed-workload time study therefore cannot identify the autonomy channel by itself.

The market result supplies a cross-organizational design. The relevant treatment is the change in optimized successful-successor supply, not overall AI intensity. In the residual-basis benchmark, cross-organizational responses should sort with diagnostic exposure; in the integrated developmental model, researchers should additionally observe whether organizations respond through developmental entry, assimilation intensity, or design changes.

6.4 Organizational design

The model points to a measurement and design rule rather than a generic human-percentage mandate. Credit verified AI diagnostic measurements; preserve independent

human assessments and source-linked feedback on residual directions needed for coverage; and allocate developmental resources according to how the resulting record converts into later independent performance. Non-identical cases can repair missing directions, while additional or differently chosen experiences can improve developmental effectiveness when assimilation is costly. Monitor live span together with senior load and successful successor output. Transfers per successful successor address incomplete appropriation when that wedge is present. The object to preserve is a diagnostic and developmental process that forms future independent judgment, not human involvement for its own sake.

7 Conclusion

Knowledge organizations produce two outputs through the same difficult work: a decision today and someone capable of making related decisions tomorrow. The central point of this paper is that these outputs need not move together. A correct action can reveal less than the judgment that produced it. Different underlying criteria can generate the same observed decision, so current performance—even repeated correct performance—need not tell us whether the judgment behind it is being reproduced. What matters is whether the information left with a recipient reveals the distinctions the organization wants future decision makers to retain.

This distinction changes the organizational problem. A developmental relationship has future value because it can produce a decision maker who no longer requires the experienced source, but preserving it also uses scarce senior capacity and can require foregoing the current benefits of AI-first execution. When the value of autonomous origination rises while current decision quality, diagnostic technology, assimilation technology, senior-time requirements, and successor value remain fixed, an organization that is already beyond the developmental capacity kink expands current activity and reduces developmental pipelines. The relationships that remain developmental are no less effective, yet the organization produces fewer successful successors while current operating surplus rises. This is the autonomy rebound. In the quantitative illustration, development can disappear even while its direct net developmental value remains positive: scarce senior capacity is worth more at the margin in expanding AI-first current scale.

Which experiences remain developmental also matters. Judgment is multidimensional, so verified AI information substitutes for human diagnostic experience only along the directions it actually supplies. In the frictionless benchmark, the organization preserves a least-cost residual set of human experiences that completes what other technologies leave uncovered. Some experiences are readily replaced; others are bottlenecks because no available substitute preserves the same missing information. With costly assimilation, full coverage is not enough. Two complete diagnostic records can differ in how effectively they produce independent successors, and a more informative record may justify its higher cost only when successful succession is sufficiently valuable. AI progress can therefore change not only how much development survives, but where it occurs and which developmental architecture the organization is willing to finance.

These choices aggregate into a market for reproduced judgment. When an AI improvement shifts organizational successor supply inward, successful successors become more valuable. That scarcity response attenuates the aggregate decline and can induce more assimilation or a switch toward more effective developmental designs, but

it does not reverse the inward shift in supply. In the diagnostic-exposure benchmark, it also reallocates successor production across organizations: exposed organizations contract, while less exposed organizations can expand as the scarcity value of reproduced judgment rises.

The implication is equally direct. AI use, human-task shares, review hours, and agreement on current actions can be identical across workflows that produce different future decision makers. What matters is whether recipients make attributable assessments before answers are revealed, whether subsequent feedback is tied to those assessments and to the maintained criterion, what diagnostic information their experience contains, what resources are devoted to assimilating it, and whether they can later perform independently. Verified AI information is developmentally valuable on the same footing as verified human information when it preserves the relevant distinctions. The broader implication is that scalable execution and the reproduction of judgment are distinct economic margins. AI can make a stock of existing judgment easier to deploy without making that judgment easier to reproduce in new people. It can therefore make expert decisions more abundant today while making independently formed decision makers scarcer tomorrow. Scalable AI can change not only how organizations make decisions, but where the economy reproduces the people capable of making them independently.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, vol. 4B, 1043–1171. Elsevier. https://doi.org/10.1016/S0169-7218(11)02410-5.

Acemoglu, Daron, Dingwen Kong, and Asuman Ozdaglar. 2026. *AI, Human Cognition and Knowledge Collapse.* NBER Working Paper 34910. National Bureau of Economic Research. https://doi.org/10.3386/w34910.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108 (6): 1488–1542. https://doi.org/10.1257/aer.20160696.

Afrouzi, Hassan, Andres Blanco, Andrés Drenik, and Erik Hurst. 2026. *Automation, Learning, and Career Dynamics.* NBER Working Paper 35157. National Bureau of Economic Research. https://doi.org/10.3386/w35157.

Agarwal, Nikhil, Alex Moehring, and Alexander Wolitzky. 2025. *Designing Human-AI Collaboration: A Sufficient-Statistic Approach.* NBER Working Paper 33949. National Bureau of Economic Research. https://doi.org/10.3386/w33949.

Agrawal, Ajay, Erik Brynjolfsson, and Anton Korinek, eds. 2026. *The Economics of Transformative AI.* 1st ed. University of Chicago Press.

Anderson, T. W. 1955. "The Integral of a Symmetric Unimodal Function over a Symmetric Convex Set and Some Probability Inequalities." *Proceedings of the American Mathematical Society* 6 (2): 170–176. https://doi.org/10.1090/S0002-9939-1955-0069229-1.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *The Review of Economic Studies* 29 (3): 155–173. https://doi.org/10.2307/2295952.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association* 23 (4): 1203–1271. https://doi.org/10.1093/jeea/jvaf023.

Bárány, Zsófia L., and Miklós Koren. 2026. *Broken Ladders: AI, Teamwork, and the Dynamics of Skill Formation in the Workplace.* Technical report. Working paper. Author's website. https://koren.mk/publications/broken-ladders.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. "Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics." *Proceedings of the National Academy of Sciences* 122 (26): e2422633122. https://doi.org/10.1073/pnas.2422633122.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York: National Bureau of Economic Research. https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-first-edition.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics* 140 (2): 889–942. https://doi.org/10.1093/qje/qjae044.

Caplin, Andrew. 2026. *Planning under Diagnostic Uncertainty: Question-Driven Learning in the Age of AI.* NBER Working Paper 35012. National Bureau of Economic Research. https://doi.org/10.3386/w35012.

Chatterji, Aaron, Daniel Rock, and Eduard Talamàs. 2025. "Transformative AI and Firms." In *The Economics of Transformative AI,* edited by Ajay K. Agrawal, Erik Brynjolfsson, and Anton Korinek. National Bureau of Economic Research Conference Report. Chicago, IL: University of Chicago Press. https://www.nber.org/books-and-chapters/economics-transformative-ai/transformative-ai-and-firms.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani. 2026. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality." *Organization Science* 37 (2): 403–423. https://doi.org/10.1287/orsc.2025.21838.

Fudenberg, Drew, and Luis Rayo. 2019. "Training and Effort Dynamics in Apprenticeship." *American Economic Review* 109 (11): 3780–3812. https://doi.org/10.1257/aer.20171939. https://www.aeaweb.org/articles?id=10.1257/aer.20171939.

Gans, Joshua. 2025. *The Microeconomics of Artificial Intelligence.* Cambridge, MA: MIT Press.

Gans, Joshua S. 2025. *A Model of the Babbage Firm.* NBER Working Paper 34145. National Bureau of Economic Research. https://doi.org/10.3386/w34145.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904. https://doi.org/10.1086/317671.

Garicano, Luis, and Luis Rayo. 2017. "Relational Knowledge Transfers." *American Economic Review* 107 (9): 2695–2730. https://www.aeaweb.org/articles?id=10.1257/aer.20160194.

———. 2025. *Training in the Age of AI: A Theory of Career Viability.* CEPR Discussion Paper 20634. Centre for Economic Policy Research. https://cepr.org/publications/dp20634.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *The Quarterly Journal of Economics* 121 (4): 1383–1435. https://doi.org/10.1093/qje/121.4.1383.

———. 2015. "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy." *Annual Review of Economics* 7:1–30. https://doi.org/10.1146/annurev-economics-080614-115748.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M. Phillips. 2024. "Production and Learning in Teams." *Econometrica* 92 (2): 467–504. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16748.

Ide, Enrique. 2026. *Automation, AI, and the Intergenerational Transmission of Knowledge.* ArXiv preprint arXiv:2507.16078, revised June 2026. https://doi.org/10.48550/arXiv.2507.16078.

Ide, Enrique, and Eduard Talamàs. 2025. "Artificial Intelligence in the Knowledge Economy." *Journal of Political Economy* 133 (12): 3762–3800. https://doi.org/10.1086/737233.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg. 2021. "Learning From Coworkers." *Econometrica* 89 (2): 647–676. https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16915.

Knight, Frank Hyneman. 1921. *Risk, uncertainty and profit.* Vol. 31. Houghton Mifflin.

Lamba, Rohit. 2026. *The Verification Hill.* Working Paper. SSRN. https://doi.org/10.2139/ssrn.6207638.

Petropoulos, Georgios. 2026. "Rungs Removed or Scaffolds Built? A Dynamic Theory of Autonomy, Review, and Learning in Early-Career Labor Markets." Working paper, available at SSRN, June 8, 2026. https://ssrn.com/abstract=6899423.

Polanyi, Michael. 1966. "The logic of tacit inference." *Philosophy* 41 (155): 1–18.

Ryle, Gilbert. 1949. *The Concept of Mind.* London: Hutchinson's University Library.

Wiles, Emma, Lisa Krayer, Mohamed Abbadi, Urvi Awasthi, Ryan Kennedy, Pamela Mishkin, Daniel Sack, and François Candelon. 2024. *GenAI as an Exoskeleton: Experimental Evidence on Knowledge Workers Using GenAI on New Skills.* Working Paper. SSRN. https://doi.org/10.2139/ssrn.4944588.

Xu, Fasheng, Jing Hou, Wei Chen, and Karen Xie. 2026. *Generative AI and Organizational Structure in the Knowledge Economy.* Technical report. arXiv preprint arXiv:2506.00532. https://doi.org/10.48550/arXiv.2506.00532.

Yin, Jiamin, Kee Yuan Ngiam, Sharon Swee-Lin Tan, and Hock Hai Teo. 2025. "Designing AI-Based Work Processes: How the Timing of AI Advice Affects Diagnostic Decision Making." *Management Science* 71 (11): 9361–9383. https://doi.org/10.1287/mnsc.2022.01454.

A Supplementary Model Statements

A.1 Two-direction diagnostic implementation

The two-direction benchmark can be implemented with four human diagnostic functions. There is one formed source and a mass of potential recipients. A live human-first case

supplies one target direction; a reserved counterfactual case supplies the other.

Table 1: Human functions in the two-direction diagnostic implementation

	Recipient function	**Source function**
Live human-first L	L^J: attributable pre-AI assessment	L^S: source-linked correction; current output protected
Counterfactual R	R^J: independent attempt on a non-identical context	R^S: source-linked correction supplying a missing target direction
Live AI-first	No required pre-answer human diagnostic attempt	Accountability may remain; no recipient-conditioned source correction is required

Proposition 13 (Four-function minimality)**.** *Within this correction-based two-dimensional benchmark, suppose live human-first production supplies only one contextual direction and each recipient begins with an unrestricted criterion. Removing any one of L^J, L^S, R^J, R^S, or making the live and counterfactual directions collinear, makes full-target diagnostic identification impossible for a nonempty open set of source and recipient criteria. With all four functions and linearly independent directions, exact diagnostic coverage is feasible and the recipient can enter the developmental pipeline.*

The organizational cycle is: the firm chooses live span and which relationships enter development; AI-first relationships receive an AI-originated action; developmental recipients commit to an assessment before AI revelation and receive a coupled source correction; residual senior capacity is allocated to rank-completing counterfactual diagnosis; verified records determine diagnostic coverage; assimilation resources then determine the probability of successful succession. If live human-first work itself supplies full target span, the reserved counterfactual domain is unnecessary.

A.2 Equivalent rank formulations

For recipient i, let X_i^H stack source-linked human context rows and let $H_i^H = X_i^{H\prime} X_i^H$.

Theorem 5 (Diagnostic rank)**.** *Let $M = I_d$, let the maintained context span be $\mathbb{R}^d$, and suppose source-linked diagnostic corrections are exact. The full source criterion is identified for recipient i if and only if* $\operatorname{rank}(H_i^H) = d$*. If* $\operatorname{rank}(H_i^H) < d$*, a distinct criterion generates the same source-linked equality record.*

Assumption 6 (Frictionless exact-acquisition benchmark)**.** *Conditional on receiving a verified diagnostic record that uniquely identifies the maintained criterion, the recipient can costlessly infer, retain, independently implement, and bear that criterion on subsequent target contexts.*

Corollary 2 (Exact reproduction under the frictionless benchmark)**.** *Under Assumption 6, recipient i exactly reproduces the full maintained criterion if and only if* $\operatorname{rank}(H_i^H) = d$*.*

This corollary is retained only as a polar case. The main text instead defines successful succession by a maintained independent-performance standard and endogenizes its probability through assimilation.

Proposition 14 (Target-component identification). *Let $M \in \mathbb{R}^{q\times d}$ be the fixed target map. The source-linked record $X_i^H\boldsymbol{\theta}$ identifies $M\boldsymbol{\theta}$ if and only if*

$$\ker X_i^H \subseteq \ker M,$$

equivalently $\operatorname{row}(M) \subseteq \operatorname{row}(X_i^H)$.

Proposition 15 (Coarse actions do not close a rank deficiency). *Under Assumption 1, let X_i^H be the source-linked equality matrix and let the recipient also observe any finite record of correct AI-first operative actions. If* $\operatorname{rank}(X_i^H) < d$, *the combined record is consistent with a nontrivial local perturbation of the full source criterion. Hence finite correct AI actions do not remove the full-rank requirement for exact full-target identification. If* $\operatorname{rank}(X_i^H) = d$, *the source-linked equalities identify the criterion regardless of the action record.*

A.3 Additional organizational regimes and capability channels

Proposition 16 (Developmental exit). *Let $\mathcal{J}(A)$ be defined by equation (9). If $g > \mathcal{J}(A)$, every optimum has $T^* = P^* = 0$. The optimal live span is the unique maximizer of*

$$[b+g]n - C(n) \quad \textit{over} \quad 0 \le n \le B/\tau.$$

If $g = \mathcal{J}(A)$, $T = 0$ is optimal, although additional developmental entry can be payoff-equivalent when it uses otherwise slack capacity.

Proposition 17 (Autonomy incidence). *Hold $(b, A, B, \tau, \rho, C, \zeta, c, \chi)$ fixed and vary g, without crossing a regime boundary. Let $\chi^* = \chi(p^*(A))$.*

(i) *In a strict under-span or balanced regime, $dn^*/dg = dT^*/dg = dP^*/dg = 0$.*

(ii) *At an interior over-span optimum,*

$$\frac{dn^*}{dg} = \frac{1+\tau/\rho}{C''(n^*)} > 0, \qquad \frac{dT^*}{dg} = -\frac{\tau}{\rho}\frac{dn^*}{dg} < 0, \qquad \frac{dP^*}{dg} = \chi^*\frac{dT^*}{dg} < 0. \tag{31}$$

(iii) *In the exhausted-development regime, $n^* = B/\tau$ and $T^* = P^* = 0$ locally.*

(iv) *In an interior developmental-exit regime, $C'(n^*) = b + g$, so $dn^*/dg = 1/C''(n^*) > 0$ and $T^* = P^* = 0$.*

With B fixed and $b(a), g(a), A(a), \tau(a), \rho(a)$ varying on an interior over-span interval while the assimilation technology is fixed, differentiating equation (16) gives

$$\Xi_A(a) = b'(a)+g'(a)\left(1+\frac{\tau}{\rho}\right)-\chi(p^*)A'(a)\frac{\tau}{\rho}-[\mathcal{J}(A)-g(a)]\left[\frac{\tau'}{\rho}-\frac{\tau\rho'}{\rho^2}\right], \qquad n' = \frac{\Xi_A}{C''(n)}. \tag{32}$$

The corresponding pipeline response follows from equation (17); successful succession additionally includes the ε_χ term in equation (18).

A.4 Robustness to source capacity and incomplete exposure

Suppose the organization can acquire formed-source capacity B at differentiable cost $F(B)$, with $F'(B) > 0$ and $F''(B) > 0$.

Proposition 18 (Endogenous source capacity)**.** *Suppose $\mathcal{J}(A) > g$ and the optimum is locally unique with $0 < T < n$, positive purchased capacity, and a binding capacity constraint. Holding $b, A, \tau, \rho, C, F, \zeta, c, \chi$ fixed while varying g, the first-order conditions imply*

$$C'(n) = b + g - [\mathcal{J}(A) - g]\frac{\tau}{\rho}, \qquad F'(B) = \frac{\mathcal{J}(A) - g}{\rho}.$$

Hence

$$\frac{dn}{dg} = \frac{1 + \tau/\rho}{C''(n)} > 0, \qquad \frac{dB}{dg} = -\frac{1}{\rho F''(B)} < 0,$$
$$\frac{dT}{dg} = \frac{1}{\rho}\left(\frac{dB}{dg} - \tau\frac{dn}{dg}\right) < 0, \qquad \frac{dP}{dg} = \chi(p^*)\frac{dT}{dg} < 0.$$

Incomplete human-first exposure may also have positive developmental value. Let $v(a) \geq 0$ be the value of a human-first live episode that is not completed into the developmental pipeline. The unreduced objective becomes

$$U_\alpha^v(n, h, T, p, a) = bn - C(n) + g(a)(n - h) + v(a)(h - T) + [A\chi(p) - \zeta c(p)]T. \tag{33}$$

Proposition 19 (Positive partial-development value)**.** *On any open capability interval with $g(a) > 0$ and $0 \leq v(a) < g(a)$ pointwise, every optimum of equation (33) has $h = T$. The reduced feasible set and objective are therefore equations (5) and (10); Proposition 4 and Theorem 1 are unchanged. At a capability where $v(a) = g(a)$, an optimum with $h = T$ exists; where $v(a) > g(a)$, the reduction need not hold.*

The earlier exogenous-success formulation is nested by fixing assimilation intensity and therefore fixing χ. In that case $\mathcal{J}(A) = A\chi - \zeta c$ is a constant developmental return, and the autonomy comparative static is unchanged. The main text strengthens this polar case by choosing assimilation endogenously and by allowing diagnostic geometry to determine the success technology.

To isolate miscoupling in counterfactual completion, let $q_c \in [0, 1]$ denote the fraction of candidate counterfactual completions whose required source response is correctly coupled to its intended recipient, conditional on the live diagnostic interaction already being correctly coupled. Effective counterfactual completion is then

$$q_c \frac{B - \tau n}{\rho}.$$

More candidate response volume does not compensate for lower coupling unless it raises the correctly coupled quantity. Appendix I gives the corresponding heterogeneous one-slot matching condition.

A.5 Human-only basis as a polar case

When $\mathcal{A}_i = \{0\}$, let $\mathfrak{B}$ be the set of all q-element subsets of candidate directions that span $\mathbb{R}^q$, and define

$$K(a) = \min_{B \in \mathfrak{B}} \sum_{e \in B} g_e(a). \tag{34}$$

Theorem 6 (Human diagnostic basis)**.** *Under the common senior-capacity normalization, suppose all $g_e(a) > 0$. Conditional on exact diagnostic coverage, frictionless developmental conversion, and zero verified AI diagnostic span, the current-surplus-maximizing workflow preserves human-first provenance on a minimum-cost basis $B^*(a)$ and assigns every diagnostically redundant candidate episode to AI-first execution. Conversely, every exact-coverage workflow contains a target-spanning subset whose autonomy cost is at least $K(a)$.*

B Formal Foundations

B.1 Action agreement and diagnostic equality

Lemma 2 (Finite correct actions need not identify the criterion)**.** *Under Assumption 1, for any finite contexts $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$, there exists $\varepsilon > 0$ such that every $\widetilde{\boldsymbol{\theta}}$ with $\|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| < \varepsilon$ generates the same operative actions as $\boldsymbol{\theta}$ on all m contexts.*

Proof. For each e, choose $\epsilon_e > 0$ from Assumption 1. Let $K = \max_e \|\boldsymbol{x}_e\|$. If $K = 0$, the result is immediate. Otherwise choose $\varepsilon < \min_e \epsilon_e / K$. Then

$$|(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})' \boldsymbol{x}_e| \le \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \, \|\boldsymbol{x}_e\| < \epsilon_e$$

for every e, so all action cells are unchanged. □

Proof of Proposition 1. Lemma 2 gives an open neighborhood of criteria producing the same finite action record. Because M has positive row rank, that neighborhood contains a perturbation h with $Mh \neq 0$. Thus the same correct actions are compatible with a different maintained criterion. □

Proof of Proposition 2. Every verified target-linked measurement observed by recipient i has the form $\phi' \boldsymbol{z}$ for a known direction $\boldsymbol{z}$. If $\mathcal{D}_i = \mathbb{R}^q$, these measurements determine every component of ϕ. If $\mathcal{D}_i \neq \mathbb{R}^q$, choose nonzero $h \in \mathcal{D}_i^{\perp}$; then ϕ and $\phi + h$ generate exactly the same verified diagnostic record, so the target is not identified. With $\mathcal{A}_i = \{0\}$ and $M = I_d$, this is $\operatorname{rank}(X_i^H) = d$. With human measurements only and general M, the equivalent source-coordinate condition is Proposition 14. □

Proof of Proposition 12. Take $q = 2$, no verified AI diagnostic span, and the same live cases, AI assignments, final source-consistent actions, scheduled review time, number of human-first diagnostic episodes, and developmental-resource budget in two workflows. For a positive mass of otherwise identical recipients, let workflow A use two verified human directions $\boldsymbol{z}_1 = (1, 0)'$ and $\boldsymbol{z}_2 = (0, 1)'$, while workflow B replaces the second direction by another episode with $\widetilde{\boldsymbol{z}}_2 = (1, 0)'$. The counts of AI actions, human involvement, and review can be identical because only the context assigned to the second diagnostic episode differs. Give the two recorded episodes the same strictly positive durable encoding precision in both workflows. In workflow A the durable information matrix is positive definite, so Proposition 6 gives a strictly positive success probability. In workflow B the durable information matrix is singular, so the success probability is zero. The named conventional observables are unchanged while successful-successor output differs. □

To prove Proposition 15, suppose $\operatorname{rank}(X_i^H) < d$ and choose nonzero $h \in \ker X_i^H$. By Lemma 2, for sufficiently small nonzero scalar c, $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + ch$ leaves every observed

AI action unchanged. It also leaves every source-linked equality unchanged because $X_i^H h = 0$. Thus the combined record is locally non-identifying. Full rank makes the equality system unique. The same argument applies in target coordinates for the basis results.

B.2 Recipient-level coupling

A correction generated from recipient i's attempt is $m_i(\boldsymbol{x}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_i^0)'\boldsymbol{x}$. Delivery to recipient j instead of i gives

$$\boldsymbol{\theta}'\boldsymbol{x} + (\boldsymbol{\theta}_j^0 - \boldsymbol{\theta}_i^0)'\boldsymbol{x}.$$

Thus recipient-level assignment matters even when aggregate attempt and correction distributions are unchanged.

C Proofs for the Baseline and Organizational Design

Proof of Proposition 13. Without L^J, the correction technology produces no recipient-specific live restriction. Without L^S, the live attempt is not anchored to the source criterion. Without R^J, the correction technology produces no recipient-specific restriction in the missing direction. Without R^S, the counterfactual attempt is not anchored to the source criterion. In each case the two-dimensional source-linked equality record has rank at most one. Collinear L and R directions have the same defect. Any accompanying finite AI-first action record does not repair this rank loss: by Lemma 2, a sufficiently small perturbation in a missing equality direction leaves those operative actions unchanged. With all four functions and independent directions, the equality design matrix has rank two, so the maintained target is exactly covered and the recipient can enter the developmental pipeline. □

Proof of Lemma 1. Fix a feasible allocation (n, h, T, p) with $h > T$, and reduce h to T. Diagnostic coverage, assimilation, and successful-successor output are unchanged. Current surplus rises by $g(a)(h - T)$, while senior-capacity use weakly falls by $(\tau_H - \tau_A)(h - T)$. Hence any optimum can be replaced, holding (n, T, p) fixed, by an optimum with $h = T$. If $g(a) > 0$, the payoff increase is strict whenever $h > T$, so every optimum satisfies $h = T$. □

Proof of Proposition 19. Fix (n, T, p). Relative to $h = T$, choosing $h > T$ changes the objective in equation (33) by

$$(v(a) - g(a))(h - T)$$

and changes senior-capacity use by

$$(\tau_H - \tau_A)(h - T) \geq 0.$$

If $v(a) < g(a)$, every $h > T$ is therefore strictly worse and weakly more capacity intensive, so every optimum has $h = T$. Optimizing p then gives $\mathcal{J}(A)$ and reduces the problem to equations (5) and (10). If $v(a) = g(a)$, replacing any feasible $h > T$ by $h = T$ leaves payoff unchanged and weakly relaxes capacity, so an optimum with $h = T$ exists. For $v(a) > g(a)$, incomplete human-first exposure can have positive net value relative to AI-first provenance, so the reduction is not generally valid. □

Proof of Proposition 18. At an active pipeline, assimilation is chosen first through equation (7); holding A and the assimilation technology fixed while varying g therefore holds p^* and $\mathcal{J}(A)$ fixed. Because $F'(B) > 0$, the capacity constraint binds at an interior optimum. Let λ be its multiplier. The first-order conditions for n, T, and B are

$$b + g - C'(n) = \lambda\tau, \qquad \mathcal{J}(A) - g = \lambda\rho, \qquad F'(B) = \lambda.$$

Eliminating λ gives the two conditions stated in the proposition. Differentiating them with respect to g yields

$$C''(n)n_g = 1 + \frac{\tau}{\rho}, \qquad F''(B)B_g = -\frac{1}{\rho}.$$

The binding capacity constraint gives $T = (B - \tau n)/\rho$, so $T_g = (B_g - \tau n_g)/\rho < 0$. Because $\chi(p^*)$ is fixed, $P_g = \chi(p^*)T_g < 0$. □

Proof of Proposition 3. After Lemma 1, total senior-capacity use is

$$\tau_A(n - T) + (\tau_H + \rho_R)T = \tau n + \rho T.$$

Every developmental recipient receives a human-first live episode, so $T \leq n$. These conditions are necessary. They are sufficient by assigning a human-first live episode and one compatible counterfactual episode to each of any mass T satisfying them. Maximizing T at fixed n gives equation (11). The two segments intersect at $B/(\tau + \rho)$. Successful successor output then follows from $P = \chi(p^*)T$. □

Proof of Proposition 4. For $W = \mathcal{J}(A) - g > 0$, maximizing T at fixed n is optimal. On the complete-development branch $T = n$,

$$U'(n) = b + \mathcal{J}(A) - C'(n).$$

On the over-span branch $T = (B - \tau n)/\rho$,

$$U'(n) = R_A - C'(n).$$

Both derivatives strictly decrease. The right derivative at $\bar{n}$ is lower than the left derivative by $W(1 + \tau/\rho) > 0$. The four cases follow from the one-sided derivatives at $\bar{n}$ and the derivative at B/τ. Strict concavity in n gives uniqueness. The optimized assimilation choice is unique by Assumption 5, and $P = \chi(p^*)T$. □

Proof of Proposition 16. At any feasible (n, T), the direct coefficient on T after optimizing assimilation is $\mathcal{J}(A) - g$. If $g > \mathcal{J}(A)$, lowering T weakly relaxes capacity and strictly raises payoff, so every optimum has $T = P = 0$. The remaining problem is strictly concave in n. At equality, lowering T cannot reduce payoff, so $T = 0$ is optimal. □

Proof of Proposition 17. On the under-span branch, $C'(n^*) = b + \mathcal{J}(A)$, which does not contain g. A strict balanced optimum remains at $\bar{n}$, also independent of g. On the interior over-span branch,

$$C'(n^*) = b + g - [\mathcal{J}(A) - g]\frac{\tau}{\rho}.$$

Differentiation gives equation (31); differentiating $T = (B - \tau n)/\rho$ gives the pipeline sign, and $P = \chi(p^*)T$ gives the successor sign. The exhausted and exit cases follow from their binding conditions and first-order equations. □

For reference, on an interior over-span branch the position-preserving derivatives holding g, A, and the assimilation technology fixed are

$$\frac{\partial n^*}{\partial B} = 0, \qquad \frac{\partial n^*}{\partial \rho} = \frac{[\mathcal{J}(A) - g]\tau}{\rho^2 C''(n^*)}, \qquad \frac{\partial n^*}{\partial \tau} = -\frac{\mathcal{J}(A) - g}{\rho C''(n^*)}.$$

They reproduce the earlier incidence logic with the net developmental-pipeline value $\mathcal{J}(A) - g$.

D Proof of the Autonomy-Rebound Theorem

Proof of Theorem 1. Equation (7) contains A, χ, ζ, c but not g, so a pure autonomy change has $p^{*\prime}(a) = 0$. On the interior over-span branch,

$$C'(n) = b + g - [\mathcal{J}(A) - g]\frac{\tau}{\rho}.$$

With only g varying,

$$C''(n)n' = g'\left(1 + \frac{\tau}{\rho}\right),$$

which proves equation (12). The frontier gives $T' = -(\tau/\rho)n'$, and equation (6) gives $P' = \chi(p^*)T'$.

Current operating surplus net of developmental-resource costs is

$$Y = (b + g)n - C(n) - [g + \zeta c(p^*)]T.$$

Because p^* is fixed under the experiment,

$$Y' = g'(n - T) + [b + g - C'(n)]n' - [g + \zeta c(p^*)]T'.$$

The first-order condition gives $b + g - C' = [A\chi(p^*) - \zeta c(p^*) - g]\tau/\rho$, while $T' = -(\tau/\rho)n'$. Therefore

$$Y' = g'(n - T) + A\chi(p^*)\frac{\tau}{\rho}n' > 0.$$

For nonemptiness, take $B = 1$, $\tau = 0.1$, $\rho = 0.2$, $b = 1$, $g = 0.1$, $C(n) = 0.1n^2$, $A = 1$, $\zeta = 1$, $\chi(p) = 1 - e^{-p}$, and $c(p) = p^2/2$. The assimilation FOC is $e^{-p} = p$, with unique solution $p^* \approx 0.567$, giving $\mathcal{J}(A) \approx 0.272$ and $W \approx 0.172 > 0$. Then $\bar{n} = 10/3$, $R_A \approx 1.014$, the unique optimum satisfies $n^* \approx 5.07$ and $T^* \approx 2.47$, strictly inside the over-span region, with $P^* = \chi(p^*)T^* > 0$. The strict inequalities persist under sufficiently small perturbations. No step uses $\alpha < 1$. □

Differentiating equation (16) with the displayed channels active gives equation (32). Log differentiation of $T = (B - \tau n)/\rho$ gives equation (17); adding $d \log \chi$ gives equation (18).

To express the pure-autonomy span response in the model's structural primitives, define

$$\varepsilon_C(n) = \frac{nC''(n)}{C'(n)} > 0.$$

At an interior over-span optimum with $g > 0$ and $C'(n) = R_A$,

$$\frac{d \log n}{d \log g} = \frac{(1 + \tau/\rho)g}{\varepsilon_C(n)R_A}, \qquad -\frac{d \log T}{d \log g} = -\frac{d \log P}{d \log g} = \frac{\ell}{1 - \ell}\,\frac{(1 + \tau/\rho)g}{\varepsilon_C(n)R_A},$$

where the equality for P uses the fact that $\chi(p^*)$ is fixed under pure autonomy.

For reference, define the local *developmental-rebound index*

$$\mathcal{R}(a) = \frac{\ell}{1-\ell}(\varepsilon_n - \varepsilon_L) - \varepsilon_R.$$

On an interior over-span branch, equation (17) is $d\log T/d\log a = -\mathcal{R}(a)$, while successful succession satisfies

$$\frac{d\log P}{d\log a} = \varepsilon_\chi - \mathcal{R}(a).$$

E Proofs for Diagnostic Rank and Precision

Proof of Theorem 5. Because $H_i^H = X_i^{H\prime} X_i^H$, $\operatorname{rank}(H_i^H) = \operatorname{rank}(X_i^H)$. Full rank makes the equality system $X_i^H \boldsymbol{\theta}$ uniquely solve for $\boldsymbol{\theta}$. Under rank deficiency choose nonzero $h \in \ker X_i^H$ and set $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} + h$; the source-linked record is unchanged while the full criterion differs. □

Corollary 2 follows from Theorem 5 and Assumption 6.

Proof of Proposition 14. The record identifies $M\boldsymbol{\theta}$ exactly when

$$X_i^H \widetilde{\boldsymbol{\theta}} = X_i^H \boldsymbol{\theta} \Longrightarrow M\widetilde{\boldsymbol{\theta}} = M\boldsymbol{\theta}.$$

Writing $h = \widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$, this is $X_i^H h = 0 \Rightarrow Mh = 0$, equivalent to $\ker X_i^H \subseteq \ker M$. Taking orthogonal complements gives the row-space condition. □

E.1 Approximate diagnostic identification

Suppose a source-linked record is

$$y_i = X_i^H \boldsymbol{\theta} + \boldsymbol{\eta}_i, \qquad \mathbb{E}[\boldsymbol{\eta}_i \mid X_i^H] = 0, \qquad \operatorname{Var}(\boldsymbol{\eta}_i \mid X_i^H) = \sigma^2 I.$$

When H_i^H is nonsingular, ordinary least squares yields $\widehat{\boldsymbol{\theta}}_i$.

Proposition 20 (Worst-direction diagnostic error)**.** *If H_i^H is nonsingular,*

$$\sup_{\|\boldsymbol{x}\|=1} \mathbb{E}\left[((\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta})'\boldsymbol{x})^2 \mid X_i^H\right] = \frac{\sigma^2}{\lambda_{\min}(H_i^H)}.$$

If H_i^H is singular, at least one full-target direction is unidentified.

Proof. Ordinary least squares gives

$$\operatorname{Var}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta} \mid X_i^H) = \sigma^2 (H_i^H)^{-1}.$$

The largest Rayleigh quotient over unit directions is $\sigma^2/\lambda_{\min}(H_i^H)$. Under singularity, a null-space component is not identified. □

Rank determines exact identification possibility; the smallest normalized eigenvalue measures weakest-direction precision in the diagnostic record itself. Successful succession additionally depends on durable assimilation.

F Proofs for the Human Diagnostic Basis

Proof of Proposition 5. Let $\pi_i : \mathbb{R}^q \to \mathbb{R}^q/\mathcal{A}_i$ be the quotient map. The verified AI measurements identify every component in $\mathcal{A}_i$. Adding human-first episodes H identifies the full target exactly when

$$\mathcal{A}_i + \operatorname{span}\{z_e : e \in H\} = \mathbb{R}^q,$$

equivalently when the quotient vectors $\{\pi_i(z_e) : e \in H\}$ span $\mathbb{R}^q/\mathcal{A}_i$. Every spanning set in the quotient contains a basis of size $q - r_i^A$. Under positive autonomy costs and the common senior-capacity normalization, deleting a human-first episode redundant in the quotient strictly raises current surplus and weakly relaxes source capacity without changing identification. Hence an optimal exact-identification workflow uses a quotient basis and chooses the one with minimum total foregone autonomy value, K_i^H. Conversely, every exact-identification workflow contains such a quotient basis and therefore incurs human autonomy cost at least K_i^H. If $\mathcal{A}_i = \{0\}$, the quotient is $\mathbb{R}^q$ and Theorem 6 is recovered. If $\mathcal{A}_i = \mathbb{R}^q$, the quotient is zero-dimensional and the empty human set identifies the target. For fixed $\mathcal{A}_i$, the quotient-space bases form the same weighted linear-matroid problem used in Theorem 2, so its exchange argument applies verbatim. □

Proof of Corollary 1. Let H^0 be a minimum-cost residual basis under $\mathcal{A}_i^0$. Since $\mathcal{A}_i^0 \subseteq \mathcal{A}_i^1$,

$$\mathcal{A}_i^1 + \operatorname{span}\{z_e : e \in H^0\} = \mathbb{R}^q.$$

Hence H^0 contains a subset H^1 whose images form a basis of $\mathbb{R}^q/\mathcal{A}_i^1$, with $|H^1| = q - \dim \mathcal{A}_i^1$. Therefore

$$K_i^H(\mathcal{A}_i^1) \le \sum_{e \in H^1} g_{ie} \le \sum_{e \in H^0} g_{ie} = K_i^H(\mathcal{A}_i^0).$$

If the inclusion is strict, $\dim \mathcal{A}_i^1 > \dim \mathcal{A}_i^0$, so H^1 is a proper subset of H^0; positive autonomy costs make the second inequality strict. A one-dimensional increase in verified AI span reduces $q - \dim \mathcal{A}_i$ by exactly one. □

Proof of Proposition 6. If the durably encoded directions do not span $\mathbb{R}^q$, then equation (20) is singular, so $\chi_i = 0$ by definition. Conversely, if a subset of record directions spans $\mathbb{R}^q$ and each direction in that subset receives strictly positive precision, then for every nonzero v,

$$v' G_i(H, p, a) v = \sum_j p_{ij} (v' z_{ij})^2 > 0,$$

so $G_i > 0$. The Gaussian criterion error then has a continuous density with full support on $\mathbb{R}^q$, while the ellipsoid $\{u : u'\Omega_i u \le \overline{L}_i\}$ has positive volume because $\Omega_i > 0$ and $\overline{L}_i > 0$. Hence $\chi_i(H, p, a) > 0$. Conditional on (n_i, T_i, H) and an interior precision choice, the terms depending on p_{ij} in the Lagrangian are $[A\chi_i(H, p, a) - \zeta_i c_i(H, p, a)]T_i$. For $T_i > 0$, differentiating gives equation (23). □

Proof of Proposition 7. For part (i), write $\Sigma_k = G_k^{-1}$. If $G_1 \ge G_0 > 0$, inversion reverses the Loewner order, so $\Sigma_1 \le \Sigma_0$. The success set

$$C_i = \{u \in \mathbb{R}^q : u'\Omega_i u \le \overline{L}_i\}$$

is convex, centrally symmetric, bounded, and has nonempty interior because $\Omega_i > 0$ and $\overline{L}_i > 0$. Anderson's Gaussian inequality implies that reducing the covariance of a centered Gaussian weakly increases the probability assigned to every such set (Anderson 1955). Hence $\chi_1 \geq \chi_0$. If $G_1 > G_0$, then $\Sigma_1 < \Sigma_0$; the strict form of the same comparison gives $\chi_1 > \chi_0$ for this nontrivial ellipsoid.

For part (ii), equal senior requirements imply that the capacity term is identical under the two designs, so their within-pipeline payoff difference is

$$A(\chi_1 - \chi_0) - \Delta_i^c.$$

Conditional on a positive developmental pipeline, design 1 is therefore preferred exactly when this expression is nonnegative. Since $\chi_1 - \chi_0 > 0$, rearranging gives equation (24); strict inequalities on either side give the corresponding strict ranking. □

Proof of Proposition 8. The scalar fixed-design first-order condition is

$$A\chi'(p^*) = \zeta c'(p^*).$$

Differentiating with respect to A gives

$$\chi'(p^*) + [A\chi''(p^*) - \zeta c''(p^*)]\frac{dp^*}{dA} = 0.$$

Therefore

$$\frac{dp^*}{dA} = \frac{\chi'(p^*)}{\zeta c''(p^*) - A\chi''(p^*)} > 0$$

under Assumption 5. Since $A = \alpha V$, $dp^*/dV = \alpha\, dp^*/dA > 0$ whenever $\alpha > 0$. Strict monotonicity of χ then implies $d\chi(p^*)/dV > 0$. □

Proof of Theorem 6. Any exact-coverage workflow must contain target contexts spanning $\mathbb{R}^q$. Every spanning set contains a linearly independent q-element subset, hence a basis $B \in \mathfrak{B}$. Under the common senior-capacity normalization, every basis uses the same diagnostic capacity. Because all autonomy costs are positive, deleting a diagnostically redundant human-first episode strictly raises current surplus and weakly relaxes source capacity without changing exact identification. Therefore an optimal exact-coverage workflow uses a basis. Among bases it chooses one with minimum total forgone autonomy value, which is $K(a)$. Conversely, every exact-coverage workflow contains a basis and therefore incurs autonomy cost at least $K(a)$. □

Proof of Theorem 2. Fix the verified AI span and hence the residual-basis family $\mathfrak{B}$. Let $c(B) = \sum_{j\in B} g_j$. If $e \notin B^*$, increasing only g_e leaves $c(B^*)$ unchanged and weakly raises the cost of every basis containing e; hence B^* remains uniquely optimal.

Now let $e \in B^*$. Every basis containing e rises in cost by exactly t. Since B^* is the unique cheapest such basis at $t = 0$, the cheapest cost among bases containing e is $K(0) + t$. If an excluding basis exists, its cheapest cost is $K(0) + \Delta_e$ and is independent of t. The minimum of these two affine quantities proves all three finite-threshold statements. If no excluding basis exists, every feasible basis contains e, so all basis costs rise by t and $K(t) = K(0) + t$. □

F.1 Endogenous success selection

Let $\boldsymbol{\delta} = \boldsymbol{\theta}_i^0 - \boldsymbol{\theta}$, $\|\boldsymbol{\delta}\| = 1$, and suppose candidate contexts are uniform on the unit sphere. A success-based router selects only contexts satisfying $|\boldsymbol{\delta}'\boldsymbol{x}| \le \kappa_s$. Let

$$Q_{\kappa_s} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}' \mid |\boldsymbol{\delta}'\boldsymbol{x}| \le \kappa_s].$$

Proposition 21 (Success selection destroys diagnostic span)**.** *Selected cases satisfy* $|(\boldsymbol{\theta}_i^0 - \boldsymbol{\theta})'\boldsymbol{x}| \le \kappa_s$. *As* $\kappa_s \downarrow 0$,

$$\boldsymbol{\delta}' Q_{\kappa_s} \boldsymbol{\delta} \to 0, \qquad \lambda_{\min}(Q_{\kappa_s}) \to 0.$$

In the limit, selected support has rank at most $d - 1$.

Proof. The first inequality is the selection rule and

$$0 \le \mathbb{E}[(\boldsymbol{\delta}'\boldsymbol{x})^2 \mid |\boldsymbol{\delta}'\boldsymbol{x}| \le \kappa_s] \le \kappa_s^2.$$

The Rayleigh quotient gives $\lambda_{\min}(Q_{\kappa_s}) \le \boldsymbol{\delta}' Q_{\kappa_s} \boldsymbol{\delta}$. At zero tolerance the selected support lies in $\boldsymbol{\delta}^\perp$. □

G Aggregate Equilibrium: Proofs and Derivations

G.1 Organizational supply and the aggregate fixed point

For organization i, write the complete internal choice as x_i, successful successor output as $P_i(x_i, a)$, and all payoff components other than successor value as $F_i(x_i, a)$. Let x_1 and x_2 be optimal at $V_1 < V_2$. Optimality gives

$$F_i(x_1, a) + \alpha V_1 P_i(x_1, a) \ge F_i(x_2, a) + \alpha V_1 P_i(x_2, a)$$

and

$$F_i(x_2, a) + \alpha V_2 P_i(x_2, a) \ge F_i(x_1, a) + \alpha V_2 P_i(x_1, a).$$

Adding yields

$$\alpha(V_2 - V_1)[P_i(V_2, a) - P_i(V_1, a)] \ge 0.$$

For $\alpha > 0$, every optimal successful-successor choice is therefore nondecreasing in V. The argument allows the organization to change developmental entry, assimilation intensity, and diagnostic design between the two values.

Let $\overline{P} < \infty$ be a common bound on feasible aggregate successful succession. Under the continuity condition in Proposition 9,

$$\mathcal{T}(P) = \mathcal{S}(\Phi'(P), a)$$

is a continuous map from $[0, \overline{P}]$ into itself, so a fixed point exists. For uniqueness, suppose $P_1 < P_2$ were fixed points. Strict concavity of Φ implies $\Phi'(P_1) > \Phi'(P_2)$, while monotone organizational supply implies

$$\mathcal{T}(P_1) = \mathcal{S}(\Phi'(P_1), a) \ge \mathcal{S}(\Phi'(P_2), a) = \mathcal{T}(P_2).$$

Fixed-point equality would require $P_1 = \mathcal{T}(P_1) \ge \mathcal{T}(P_2) = P_2$, a contradiction. This proves Proposition 9.

The continuity requirement is aggregate rather than firm-by-firm. Individual policies may jump when a diagnostic design switches or developmental entry begins or ends. An atomless distribution of the relevant organizational thresholds is one sufficient regularity condition.

G.2 Scarcity feedback

For the finite result, let $\mathcal{S}^0$ and $\mathcal{S}^1$ satisfy $\mathcal{S}^1(V) \leq \mathcal{S}^0(V)$ for every relevant V. Suppose, contrary to the theorem, that $P^{\text{agg},1} > P^{\text{agg},0}$. Then

$$V^1 = \Phi'(P^{\text{agg},1}) < \Phi'(P^{\text{agg},0}) = V^0.$$

Monotonicity of $\mathcal{S}^1$ in V implies

$$P^{\text{agg},1} = \mathcal{S}^1(V^1) \leq \mathcal{S}^1(V^0) \leq \mathcal{S}^0(V^0) = P^{\text{agg},0},$$

a contradiction. Hence aggregate successful succession weakly falls and V weakly rises. If $\mathcal{S}^1(V^0) < \mathcal{S}^0(V^0)$, equality in aggregate succession would imply equality in V and contradict the strict inward shift, so both inequalities are strict.

At a differentiable equilibrium,

$$P^{\text{agg}} = \mathcal{S}(V, a), \qquad V = \Phi'(P^{\text{agg}}).$$

Differentiating gives

$$P_a^{\text{agg}} = \mathcal{S}_a + \mathcal{S}_V \Phi''(P^{\text{agg}}) P_a^{\text{agg}},$$

so

$$P_a^{\text{agg}} = \frac{\mathcal{S}_a}{1 - \mathcal{S}_V \Phi''(P^{\text{agg}})}.$$

Because $\mathcal{S}_V \geq 0$ and $\Phi'' < 0$, the denominator is weakly larger than one. If $\mathcal{S}_a < 0$, then $V_a = \Phi'' P_a^{\text{agg}} > 0$. This proves Theorem 3.

For the elasticity representation, define

$$\eta = -\frac{P^{\text{agg}} \Phi''(P^{\text{agg}})}{V} > 0, \qquad \varepsilon_{\mathcal{S},V} = \frac{V \mathcal{S}_V}{P^{\text{agg}}} \geq 0.$$

At equilibrium,

$$-\mathcal{S}_V \Phi''(P^{\text{agg}}) = \eta \varepsilon_{\mathcal{S},V}.$$

Multiplying the level derivative by a/P^{agg} gives equation (28).

G.3 Residual-basis benchmark: reallocation and local incidence

In the frictionless residual-basis benchmark, organization i's problem depends on successor value and basis cost through

$$z_i = \alpha V - K_i.$$

The same revealed-preference argument used above implies that optimal $P_i = T_i$ is nondecreasing in z_i. Between two equilibria,

$$z_i^1 - z_i^0 = \alpha \Delta V - \Delta K_i.$$

Hence

$$\Delta K_i > \alpha \Delta V \Longrightarrow P_i^1 \leq P_i^0, \qquad \Delta K_i < \alpha \Delta V \Longrightarrow P_i^1 \geq P_i^0,$$

with strict inequalities wherever the policy is strictly responsive. This proves part (i) of Theorem 4 once Theorem 3 supplies $\Delta V > 0$.

For part (ii), before any basis switch a common increase t in direction e's autonomy value gives $\Delta K_i = t$ for exposed organizations and $\Delta K_i = 0$ for unexposed organizations. Suppose $\alpha \Delta V \geq t$. Then every exposed organization's z_i weakly rises and every unexposed organization's z_i strictly rises. Revealed preference would imply weakly higher successor supply for every organization, contradicting the strict aggregate contraction. Therefore $0 < \alpha \Delta V < t$, and the exposed and unexposed signs follow.

The following local sufficient-statistic result is retained only for this frictionless benchmark.

Proposition 22 (Local diagnostic incidence)**.** *Work away from basis ties. Let E_e and U_e denote organizations whose locally unique minimum residual basis does and does not contain direction e, and let $\mathcal{S}_E$ and $\mathcal{S}_U$ be their aggregate successful-successor supplies. Suppose these group supplies are differentiable at equilibrium and $\mathcal{S}_V > 0$. Define*

$$\phi_e = \frac{\mathcal{S}_{E,V}}{\mathcal{S}_V} \in [0, 1] \tag{35}$$

and

$$\psi = \frac{\eta \varepsilon_{\mathcal{S},V}}{1 + \eta \varepsilon_{\mathcal{S},V}} \in [0, 1). \tag{36}$$

For a normalized marginal increase a_e in the autonomy value of direction e,

$$\frac{dP^{\text{agg}}}{da_e} = -\frac{\mathcal{S}_V}{\alpha} \frac{\phi_e}{1 + \eta \varepsilon_{\mathcal{S},V}}. \tag{37}$$

Moreover,

$$\frac{dV}{da_e} = \frac{\psi \phi_e}{\alpha}, \qquad \frac{dP_E}{da_e} = \frac{\mathcal{S}_{E,V}}{\alpha}(\psi \phi_e - 1), \qquad \frac{dP_U}{da_e} = \frac{\mathcal{S}_{U,V}}{\alpha} \psi \phi_e.$$

If $\phi_e > 0$, exposed aggregate succession falls whenever $\mathcal{S}_{E,V} > 0$, while unexposed aggregate succession rises whenever $\mathcal{S}_{U,V} > 0$. If in addition $P_U > 0$, the unexposed share of aggregate succession rises.

Proof. On one side of a basis-switch threshold, a normalized marginal increase in direction e's autonomy value changes an exposed organization's basis cost one-for-one by Theorem 2; it does not change an unexposed organization's basis cost. Therefore

$$\mathcal{S}_{E,a_e} = -\frac{1}{\alpha} \mathcal{S}_{E,V}, \qquad \mathcal{S}_{U,a_e} = 0, \qquad \mathcal{S}_V = \mathcal{S}_{E,V} + \mathcal{S}_{U,V}. \tag{38}$$

Thus $\mathcal{S}_{a_e} = -(\mathcal{S}_V / \alpha)\phi_e$. Substitution into equation (27) gives equation (37), and

$$V_{a_e} = \Phi''(P^{\text{agg}}) P^{\text{agg}}_{a_e} = \frac{\psi \phi_e}{\alpha}.$$

Applying this price response to the two group supplies gives the displayed group derivatives. Since $0 \leq \psi < 1$ and $0 \leq \phi_e \leq 1$, the stated signs follow. □

Corollary 3 (Basis-switch reversal)**.** *Suppose direction e is initially in organization i's unique minimum-cost residual basis but is replaceable. Before the exchange threshold in Theorem 2, a responsive organization contracts succession after a marginal improvement in e. Immediately after it switches to a basis excluding e, further improvement leaves its basis cost locally unchanged. If other responsive organizations remain exposed so the scarcity value continues to rise, organization i weakly expands succession, strictly so when its policy is locally strictly increasing in V. A diagnostic bottleneck cannot generate this reversal.*

Proof. Immediately before a finite exchange threshold the improved direction belongs to organization i's unique minimum basis, so further improvement raises K_i. Immediately after the switch, Theorem 2 implies that further improvement leaves K_i locally unchanged. If other responsive organizations remain exposed, aggregate supply continues to shift inward and the scarcity value rises. The organization's net index $\alpha V - K_i$ then increases, so its successor supply weakly rises; strict local responsiveness gives the strict sign. A bottleneck remains in every basis, so this insulation never occurs. □

G.4 Constrained efficiency under full appropriation

When $\alpha = 1$, market payments are transfers. The planner chooses each organization's full internal vector x_i to solve

$$\max_{\{x_i\}} \left\{ \int_0^1 F_i(x_i, a)\, di + \Phi(P^{\text{agg}}) \right\}, \qquad P^{\text{agg}} = \int_0^1 P_i(x_i, a)\, di,$$

subject to each organization's technological constraints. The marginal social value of organization i's successful successor output is $\Phi'(P^{\text{agg}}) = V$. Hence, conditional on maintained targets and organizational technologies, the planner's organization-by-organization problem coincides with equation (25) evaluated at the competitive deployment price. Under the maintained concavity and existence conditions, the competitive continuation equilibrium decentralizes the constrained-efficient allocation. For $\alpha < 1$, organizations respond to only αV, creating the incomplete-appropriation wedge of Proposition 10; equilibrium scarcity pricing does not remove that wedge.

H Proofs for Welfare and AI Timing

Proof of Proposition 10. Fix V, capability, and the scalar benchmark architecture, and write $A = \alpha V$. The organization's complete choice can be written as $x = (n, T, p)$ with payoff $F(x) + AP(x)$, where $P(x) = \chi(p)T$. Revealed preference therefore implies that optimal successful succession is nondecreasing in A; hence $\alpha < 1$ yields weakly lower P than $\alpha = 1$.

For the interior comparative statics, equation (8) gives $p_A^* > 0$. On an interior over-span branch,

$$C'(n) = b + g - [\mathcal{J}(A) - g]\frac{\tau}{\rho},$$

so the envelope result $\mathcal{J}'(A) = \chi(p^*)$ gives

$$C''(n)n_A = -\chi(p^*)\frac{\tau}{\rho} < 0.$$

Thus lower A raises n, while $T = (B - \tau n)/\rho$ falls. Since both T and $\chi(p^*)$ increase with A, successful succession is strictly lower at the private $\alpha < 1$ optimum. On the under-span branch, $T = n$ and

$$C'(n) = b + \mathcal{J}(A),$$

so $n_A = \chi(p^*)/C''(n) > 0$; lower A lowers $n = T$, and it also lowers p^* and χ. This proves the stated signs. □

Proof of Proposition 11. At a fixed common allocation and developmental technology, moving from human-first to AI-first timing changes private payoff by $G - \alpha V \Delta P$ and planner payoff by $G - V\Delta P$. Opposite strict rankings are exactly equation (30). □

I Heterogeneous Diagnostic Matching and Nonlinear Criteria

Let I be a finite set of live human-first recipients and J a finite set of available counterfactual slots. In this benchmark each recipient requires exactly one additional target-relevant diagnostic direction, and each slot can be assigned to at most one recipient. An edge (i, j) means that slot j, if correctly coupled to recipient i, individually supplies the missing direction and completes that recipient's diagnostic record.

Proposition 23 (One-slot diagnostic matching frontier)**.** *The maximum number of completed records equals the cardinality of a maximum matching. Every human-first live recipient can be completed if and only if*

$$|N(S)| \geq |S| \qquad \text{for every } S \subseteq I.$$

Proof. Under the one-missing-slot benchmark, feasible one-to-one diagnostic assignments are graph matchings; Hall's theorem gives the covering condition. □

For nonlinear local criteria, let the source index be $s(\boldsymbol{x}; \vartheta)$, continuously differentiable in $\vartheta \in \mathbb{R}^d$, and let $G_i(\vartheta)$ stack source-linked episode gradients.

Proposition 24 (Local nonlinear diagnostic rank)**.** *If $G_i(\vartheta)$ has full column rank, the source parameter is locally diagnostically identified. If rank is deficient, a nonzero local parameter direction leaves all first-order source-linked episode predictions unchanged.*

Proof. Use a nonsingular sub-Jacobian and the inverse-function theorem. Under rank deficiency choose nonzero h with $G_i(\vartheta)h = 0$. □

J Feature Normalization and Spectral Coverage

Rank is invariant under invertible recodings. Ordinary eigenvalues are invariant under orthogonal transformations and scale equivariant. Let $Q > 0$ be the second-moment matrix of the maintained deployment domain and define

$$\underline{\lambda}_Q(H) = \inf_{u \neq 0} \frac{u'Hu}{u'Qu}.$$

Lemma 3 (Basis-invariant spectral coverage)**.** *Under an invertible recoding $\widetilde{\boldsymbol{x}} = A\boldsymbol{x}$, $\widetilde{H} = AHA'$, $\widetilde{Q} = AQA'$, rank is preserved, and*

$$\underline{\lambda}_{\widetilde{Q}}(\widetilde{H}) = \underline{\lambda}_Q(H).$$

Proof. Rank is preserved by invertible multiplication. In the generalized Rayleigh quotient substitute $v = A'u$. □

K Forward-Looking Human Lineage

The static continuation value compresses several future benefits. This extension separates a successor's deployment value from the value of becoming a future source. It studies a fully autonomous limit. As in Assumption 2, AI action origination does not require a formed-human source; the additional limiting restriction here sets residual human accountability requirements for a fixed live cohort of mass $N > 0$ to zero.

Because residual accountability is zero in this limit, serving the cohort AI-first uses no formed-human source capacity. To keep the lineage argument one-dimensional, this appendix also takes the frictionless developmental limit in which every covered pipeline succeeds, so developmental entry and successful succession coincide: $T_t = P_t$.

Let S_t be inherited formed-source capacity. If the cohort is served entirely AI-first, current operating value is

$$Y_A(a) = [b + g(a)]N - C(N). \tag{39}$$

A successful developmental recipient yields private deployment value $D = \alpha V_D$, excluding her future value as a source. Preserving one such recipient forgoes autonomy value $g(a)$, uses ρ units of current source capacity, and adds $\kappa > 0$ units to next period's source stock:

$$S_{t+1} = (1 - \delta)S_t + \kappa P_t, \qquad 0 \le P_t \le \min\{N, S_t/\rho\}. \tag{40}$$

The forward-looking organization solves

$$W(S; a) = Y_A(a) + \max_{0 \le P \le \min\{N, S/\rho\}} \left\{ [D - g(a)]P + \beta W((1 - \delta)S + \kappa P; a) \right\}. \tag{41}$$

Theorem 7 (Human-lineage exit)**.** *Let $\beta \in [0, 1)$, $\delta \in (0, 1)$, and choose a compact invariant state interval.*

(i) *If $g(a) > D$, the unique optimal policy is $P_t = 0$ at every positive state. Thus $S_t = (1 - \delta)^t S_0 \to 0$, while the live cohort remains served and current AI-mediated operating value remains $Y_A(a)$.*

(ii) *If $g(a) < D$, every optimal policy chooses*

$$P_t = \min\{N, S_t/\rho\}.$$

If $\kappa > \delta\rho$, then $S = 0$ is absorbing and every $S_0 > 0$ converges to

$$S^h = \frac{\kappa N}{\delta}.$$

(iii) *If $g(a) = D$, $P_t = 0$ is optimal but need not be unique.*

If $g'(a) > 0$, current AI-mediated operating value satisfies $Y_A'(a) = g'(a)N > 0$. Thus a capability threshold $a^\dagger$ satisfying $g(a^\dagger) = D$ can switch the optimal lineage from regeneration to extinction while current operating value continues to rise.

Choose $\bar{S} \ge \kappa N/\delta$, so $[0, \bar{S}]$ is invariant under equation (40). The feasible correspondence for P is nonempty, compact valued, and continuous. The Bellman operator in equation (41) is a contraction with modulus β, yielding a unique bounded continuous value function.

Proof of Theorem 7. Because $Y_A(a)$ is independent of S and P, any feasible policy $\{P_t\}$ has discounted value

$$\frac{Y_A(a)}{1 - \beta} + [D - g(a)] \sum_{t=0}^{\infty} \beta^t P_t.$$

The state affects only which nonnegative sequences $\{P_t\}$ are feasible.

If $g > D$, every positive P_t strictly lowers the displayed sum. The unique optimum is therefore $P_t = 0$ at every date, giving $S_t = (1 - \delta)^t S_0 \to 0$.

If $g < D$, increasing any feasible P_t raises current payoff and weakly expands every subsequent feasible set because it raises S_{t+1}. Hence every optimum uses the maximal feasible successful-succession mass at every date:

$$P_t = \min\{N, S_t/\rho\}.$$

The induced transition is

$$S_{t+1} = \begin{cases} (1 - \delta + \kappa/\rho)S_t, & 0 \le S_t < \rho N, \\ (1 - \delta)S_t + \kappa N, & S_t \ge \rho N. \end{cases}$$

If $\kappa > \delta\rho$, the first-branch slope exceeds one, so every positive initial stock reaches the second branch in finite time. The second branch has slope $1 - \delta \in (0, 1)$ and unique fixed point $S^h = \kappa N/\delta > \rho N$. Every positive path therefore converges to S^h. At $g = D$, the reproduction term vanishes and $P = 0$ is optimal but not unique. Finally, differentiating equation (39) gives $Y'_A = g'N > 0$. □

The dynamic result applies to a strong but transparent limit in which AI action origination is source-independent and residual human accountability requirements for current production are zero. In that limit, scalable non-transmitting action origination can make source-stock regeneration stop without a contemporaneous production collapse. Positive human oversight requirements for AI-first current production would reconnect source depletion to current output.

L Notation and Summary

Symbol	Meaning
$\boldsymbol{\theta}$	Local diagnostic representation of formed source judgment; $M\boldsymbol{\theta} = \phi$ is the maintained target
$s_{\boldsymbol{\theta}}(\boldsymbol{x})$	Source diagnostic index $\boldsymbol{\theta}'\boldsymbol{x}$
$\Gamma_{\boldsymbol{x}}$	Map from diagnostic index to operative action
$t_i(\boldsymbol{x})$	Attributable pre-AI recipient diagnostic index
$m_i(\boldsymbol{x})$	Source-linked local diagnostic correction
X_i^H, H_i^H	Source-linked human-first design and diagnostic information matrices
L^J, L^S, R^J, R^S	Four human diagnostic functions in the two-direction implementation
n	Total live span
h	Human-first live mass; baseline reduction has $h = T$
T	Developmental-pipeline mass: recipients receiving the benchmark coverage process
p, p^*	Assimilation intensity and its optimum in the scalar benchmark
$\chi(p)$	Probability that a developmental recipient meets the maintained independent-performance standard
$P = \chi(p)T$	Successful-successor mass
$\overline{L}_i, \Omega_i$	Maintained criterion-loss standard and future-context second-moment matrix governing independent-performance loss
$c(p), \zeta$	Assimilation resource cost and its scale parameter

Symbol	Meaning
$\mathcal{J}(A)$	Optimized developmental value $\max_p\{A\chi(p) - \zeta c(p)\}$ before the autonomy opportunity cost
B	Focal formed-source capacity; endogenous in Proposition 18
$F(B)$	Convex cost of acquiring formed-source capacity
τ_A, τ_H, ρ_R	AI-first live, human-first live, and counterfactual senior requirements
τ	Reduced AI-first accountability requirement τ_A
ρ	Incremental developmental requirement $\tau_H - \tau_A + \rho_R$
b	Current surplus common to live modes
$g(a)$	Autonomy value of AI-first rather than human-first execution
$A = \alpha V$	Privately valued continuation benefit per successful successor
$W = \mathcal{J}(A) - g$	Net direct value of a developmental pipeline relative to AI-first execution
$\bar{n}$	Complete-development capacity $B/(\tau + \rho)$
$v(a)$	Value of incomplete human-first exposure in Proposition 19
$\mathcal{A}_i(a), r_i^A(a)$	Verified AI diagnostic span for recipient i and its rank
$K_i^H(a)$	Minimum residual human diagnostic-basis cost after crediting $\mathcal{A}_i(a)$
$K(a)$	Minimum human diagnostic-basis cost in the no-verified-AI polar case
$\mathcal{E}_i^A(a), H$	Verified AI diagnostic items and chosen human diagnostic episodes in the general developmental design
$G_i(H, p, a)$	Durable human information matrix assembled from encoded human and AI diagnostic items
$\chi_i(H, p, a)$	Probability that the general diagnostic design produces a successful successor
$K_i^C(H, a)$	Current autonomy opportunity cost of the chosen human diagnostic design
$R_i(H)$	Additional senior-capacity requirement of the chosen diagnostic design
$c_i(H, p, a)$	Assimilation-resource cost of the complete diagnostic record
$g_i^L(a)$	Organization i's live autonomous-execution return in aggregate equilibrium
$P_i(V, a)$	Organization i's optimized successful-successor supply
P^{agg}	Aggregate successful succession supplied to the deployment sector
$\Phi(P^{\text{agg}}), V$	Deployment surplus and equilibrium scarcity value, with $V = \Phi'(P^{\text{agg}})$
$\mathcal{S}(V, a)$	Aggregate organizational successful-successor supply
$\mathcal{S}_E, \mathcal{S}_U$	Benchmark aggregate successor supplies of organizations exposed and unexposed to a targeted diagnostic direction
ϕ_e	Benchmark share of marginal successor supply carried by organizations diagnostically exposed to direction e
$\eta, \varepsilon_{\mathcal{S},V}$	Deployment-demand and organizational-supply elasticities in aggregate equilibrium
ψ	Benchmark scarcity-incidence coefficient $\eta\varepsilon_{\mathcal{S},V}/(1 + \eta\varepsilon_{\mathcal{S},V})$
Δ_e	Cost gap to the cheapest residual diagnostic basis excluding episode e
ℓ	Base live senior-load share $\tau n/B$ on the reduced frontier
$\mathcal{R}$	Local developmental-rebound index
S	Formed-human source stock in the lineage extension
N	Fixed live cohort in the fully autonomous lineage limit
D	Private deployment value of a successor excluding lineage value

Symbol	Meaning
κ	Source-capacity contribution of a successful successor

The scalar baseline chooses assimilation first through equation (7), then maximizes equation (10) over live span and developmental entry. Proposition 4 gives the four positive-development regimes, Proposition 16 the developmental-exit case, and Theorem 1 the pure autonomy response. Proposition 18 shows the sign survives endogenous purchase of source capacity. Proposition 2, Proposition 5, Corollary 1, Propositions 6 and 7, and Theorem 2 establish exact diagnostic coverage, the residual human basis after verified AI information, endogenous developmental success, developmental dominance and upgrading, and the frictionless placement benchmark. Proposition 8 gives the intensive assimilation response to successor scarcity. Proposition 9 and Theorems 3 and 4 close the market for successful successors and characterize scarcity feedback and benchmark geometric reallocation. Theorem 7 records the forward-looking source-stock threshold in the frictionless developmental limit.

Map of the Theory

Figure 3 summarizes how the main economic objects fit together. The first two blocks isolate the informational problem. Diagnostic coverage and assimilation determine the technology for producing successful successors; organizations choose how much of that technology to operate and which diagnostic designs to preserve; and equilibrium scarcity feeds back into the value of those choices. Measurement sits outside this causal chain because its role is to recover these otherwise hidden margins from observed workflows.

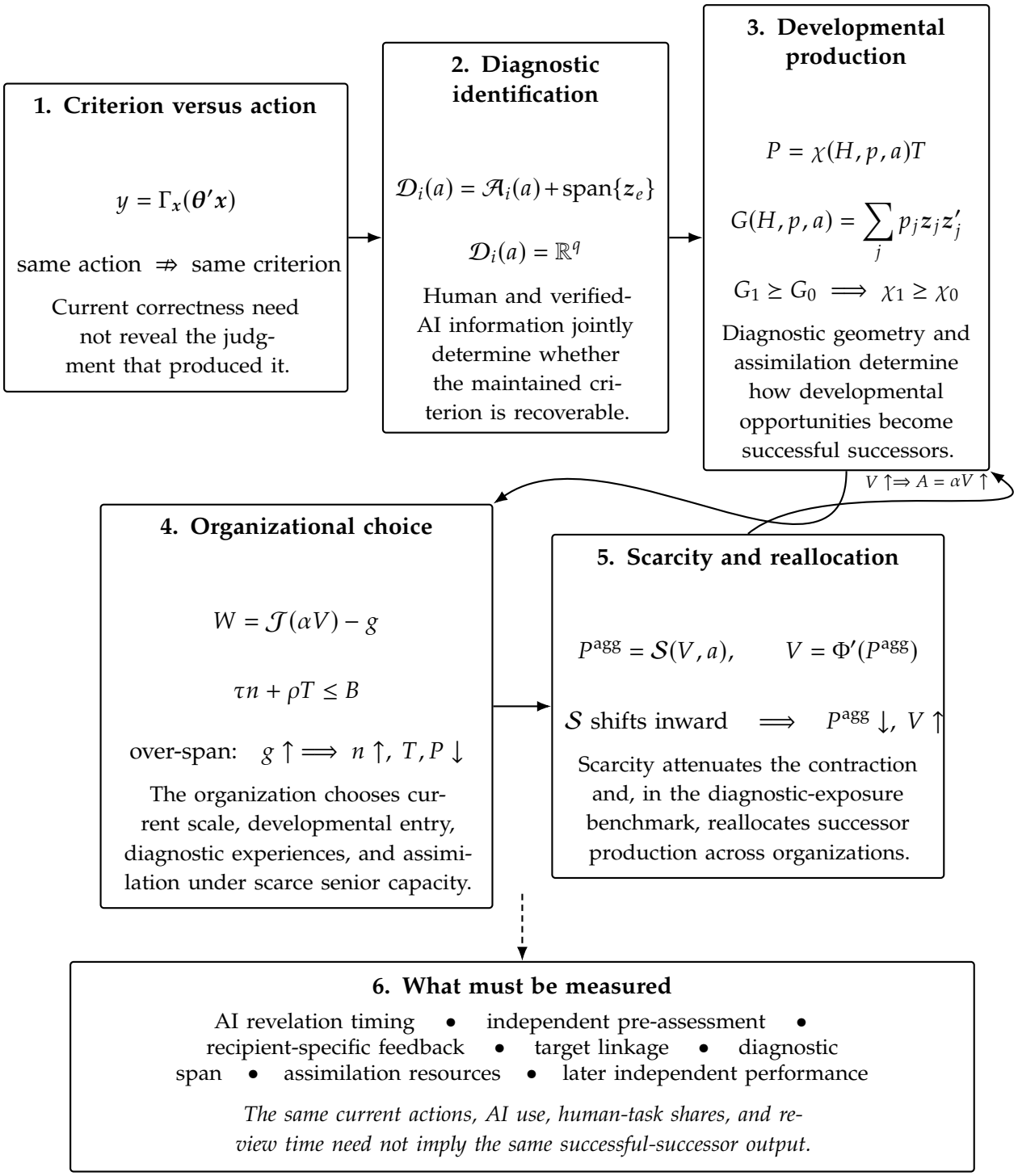


Figure 3: Map of the theory. Current work produces an operative action and can also produce information about the criterion behind that action. Diagnostic identification determines whether the maintained criterion is available to be learned; diagnostic geometry and assimilation determine how developmental opportunities become independent capability; organizational choice determines how much of that developmental technology is operated; and the resulting supply of successful successors determines the scarcity value of reproduced judgment. That scarcity value feeds back into developmental investment. Measurement must therefore recover the informational and developmental structure of the workflow rather than infer succession from current output or human involvement alone.