When Everyone Writes Well: The Economics of Writing with AI: analytical supplement

Search, learning, selection, and field effects*

Johan Fourie†

This supplement states the assumptions behind the article's four predictions and field-level conjecture, and derives the results that require formal support. Writing generates candidate arguments, and the author's capability selects among them. Certain writing tasks also build or maintain that capability over time. An anchor alters the ranking rule, while signalling and field effects arise when the same technology is used across authors. The results are benchmark models linked by shared notation. They are not a single optimisation or equilibrium model, and they do not derive authors' delegation choices. Where a result depends on who delegates which task, the pattern of delegation is an assumption. The article's contribution is to apply these mechanisms to scholarly writing and to distinguish their empirical implications. The accompanying R script checks the displayed examples and comparative-static signs; those calculations corroborate the algebra without replacing the proofs below.

1. Common environment

An argument is a point x in an argument space $\mathcal{X}$, with quality $q(x)$. An *iteration* has two tasks. Variation produces a candidate argument at execution cost c_v, and selection checks it against the evidence at cost c_s. A total effort budget B therefore permits

$$n = \frac{B}{c_v + c_s}$$

checked candidates in the continuous relaxation. The integer problem uses $\lfloor B/(c_v + c_s) \rfloor$. Generating an unchecked alternative does not count as an iteration.

Assumption 1 (Recall)**.** *The author may retain the best draft found so far at no additional cost. A typewriter is consistent with recall because a completed draft can be kept. Its cost arises when the*

*This is the supplementary material for Fourie, Johan. 2026. "When Everyone Writes Well: The Economics of Writing with AI." Working Paper, Department of Economics, Stellenbosch University. I used Anthropic's Claude, OpenAI's Codex and Google's Gemini to map the literature, work through the accompanying derivation, classify the institutional affiliations in Section 8 of the paper by country, and draft and revise the manuscript; refine.ink, an AI review service, reviewed it. The claims, and the responsibility for them, are mine. Cite the paper, not this supplement.

†Department of Economics, Stellenbosch University. Email: johanf@sun.ac.za.

author revises that draft and must retype unchanged passages, which raises c_v.

The search benchmark holds the effort budget and candidate-quality distribution fixed under accurate selection. It also treats candidates as independent draws from F even when the author revises an existing draft; revising an incumbent affects c_v without altering F in this benchmark. Later sections relax these restrictions in turn. This separation is substantive because lowering execution cost need not improve the selected argument when composing alters the candidate distribution or when selection is distorted.

2. Prediction 1: cheaper search

Let $X_1, X_2, \ldots$ be independent draws from a non-degenerate distribution F with $\mathbb{E}\,|X_1| < \infty$. Under recall, the retained quality after n checked candidates is $Q_n = \max_{i \le n} X_i$.

Lemma 1 (Monotone, diminishing gains)**.** *For every integer $n \ge 1$,*

$$\Delta_n \equiv \mathbb{E}[Q_{n+1}] - \mathbb{E}[Q_n] = \int_{-\infty}^{\infty} F(x)^n \big(1 - F(x)\big)\, dx > 0,$$
$$\Delta_{n+1} - \Delta_n = -\int_{-\infty}^{\infty} F(x)^n \big(1 - F(x)\big)^2\, dx < 0.$$

Thus $\mathbb{E}[Q_n]$ is strictly increasing and strictly concave as an integer sequence.

Proof. The tail-integral representation is

$$\mathbb{E}[Q_n] = \int_0^{\infty} (1 - F(x)^n)\, dx - \int_{-\infty}^{0} F(x)^n\, dx.$$

Both integrals are finite because $\mathbb{E}\,|X_1| < \infty$. Differencing gives the first displayed integral, and differencing once more gives the second. On $x \ge 0$ the first integrand is bounded by $1 - F(x)$, and on $x < 0$ by $F(x)$; the corresponding bounds also cover the second integrand. For a non-degenerate distribution, choose x_0 with $F(x_0) \in (0, 1)$. Right-continuity gives an $\epsilon > 0$ on which $F(x) < 1$ to the right of x_0, while monotonicity gives $F(x) \ge F(x_0) > 0$. Thus $\{x : 0 < F(x) < 1\}$ has positive Lebesgue measure, so both signs are strict. □

For a differentiable relaxation, define for $t \ge 1$

$$\psi(t) = \int_0^{\infty} (1 - F(x)^t)\, dx - \int_{-\infty}^{0} F(x)^t\, dx.$$

This agrees with $\mathbb{E}[Q_n]$ at integer $t = n$. It is finite for $t \ge 1$ and twice differentiable for $t > 1$, with

$$\psi'(t) = -\int F(x)^t \ln F(x)\, dx > 0, \qquad \psi''(t) = -\int F(x)^t \big(\ln F(x)\big)^2\, dx < 0.$$

To justify differentiation, fix $t_0 > 1$ and choose $1 < a < t_0$. In a neighbourhood on which $t \ge a$, $F^t |\ln F|^k$ is bounded by $C_{a,k} F$ on $x < 0$ and by $C_{a,k}(1 - F)$ on $x \ge 0$, for $k = 1, 2$. These bounds

are integrable. No finite derivative at $t = 1$ is asserted, because a finite absolute mean alone does not guarantee one.

Proposition 1 (Cheaper iteration and the distribution of gains)**.** *Under Assumption 1, suppose candidates share the distribution F and selection identifies the best realised quality.*

(a) Expected retained quality is increasing and concave in the integer number of checked candidates.

(b) In the continuous relaxation, if $B/(c_v + c_s) > 1$, then

$$\frac{\partial}{\partial c_v}\psi\left(\frac{B}{c_v + c_s}\right) = -\psi'\left(\frac{B}{c_v + c_s}\right)\frac{B}{(c_v + c_s)^2} < 0.$$

The same sign holds for c_s.

(c) Let writers differ only in an execution-cost schedule $c_v^0(\theta)$, with $c_v^{0\prime}(\theta) < 0$. A technology offers the common execution cost $\bar{c}$, so the author uses $c_v^1(\theta) = \min\{\bar{c}, c_v^0(\theta)\}$. Among affected writers $c_v^0(\theta) > \bar{c}$, both the gain in checked candidates and the gain in expected quality are strictly decreasing in θ. Writers already at or below $\bar{c}$ are unaffected. With integer revision counts the ordering is weak because of rounding. Every writer compared can afford at least one checked candidate before adoption.

Proof. Part (a) is Lemma 1, and part (b) follows from the chain rule on the stated domain. For part (c), every affected writer has the common endpoint $\bar{n} = B/(\bar{c} + c_s)$ and the initial count $n_0(\theta) = B/(c_v^0(\theta) + c_s)$. Hence

$$\Delta n(\theta) = B\left[\frac{1}{\bar{c} + c_s} - \frac{1}{c_v^0(\theta) + c_s}\right], \qquad \Delta n'(\theta) = \frac{Bc_v^{0\prime}(\theta)}{(c_v^0(\theta) + c_s)^2} < 0.$$

The initial count $n_0(\theta)$ is strictly increasing in θ. Since ψ is strictly increasing on $t \geq 1$ and the endpoint $\bar{n}$ is common, $\psi(\bar{n}) - \psi(n_0(\theta))$ is strictly decreasing in θ. This ordering requires only monotonicity, without differentiability or concavity. Holding F, B, and c_s common under accurate judgement is essential, because the ranking need not hold when any of these primitives also varies across writers. □

In the integer problem, $\mathbb{E}[Q_{\lfloor B/(c_v+c_s)\rfloor}]$ is weakly decreasing in either cost and changes strictly only when a cost change crosses an integer threshold.

Remark 1 (The selection bottleneck)**.** As execution cost approaches its attainable minimum, the number of checked candidates approaches $B/(\bar{c} + c_s)$, and B/c_s if $\bar{c} = 0$. Cheap generation does not produce an unlimited number of checked revisions while judgement remains costly.

Remark 2 (Optimal stopping and the default)**.** The fixed-budget result is distinct from sequential search. In the standard search-with-recall problem, taking another independent draw costs c and the reservation value z solves $c = \int_z^\infty (1 - F(x))\,dx$. Where F is continuous at z and $F(z) < 1$, $dz/dc = -1/(1 - F(z)) < 0$, so lower search cost raises the reservation value and makes immediate

stopping less likely. Holding the cost fixed, a free incumbent x_0 ends search immediately when $x_0 \geq z$, so a stronger initial draft can shorten search rationally (Weitzman, 1979). Recent search theory also shows that a larger menu does not increase reliance on a default in the constant-cost, known-distribution benchmark, unless features such as learning about the distribution or rising search costs are added (de Lara and Dean, 2025). Section 4 adds a separate present-bias wedge.

3. Prediction 2: the manuscript and the writer

The static benchmark holds judgement fixed. Delegation has two coordinates. Let $d_L \in [0,1]$ denote the share of linguistic execution delegated to a machine and $d_R \in [0,1]$ its share in forming the first explicit representation of the problem. An answer-first workflow raises both, whereas an author-first workflow can combine high d_L with low d_R. Write $d = (d_L, d_R)$.

Current manuscript quality is

$$Q_t = G(n(d_L), J_t, b(d_R, J_t)), \qquad G_n(n, J, 0) > 0, \quad G_J > 0, \quad G_b < 0,$$

where J_t is current judgement and b is distortion in selection. The strict sign $G_n(n, J, 0) > 0$ is imposed only in the unbiased benchmark $b = 0$ on domains where additional candidates have positive probability of improving on any retained draft (Proposition 1; under Gaussian evaluation noise, Proposition 3(b); and Remark 5). Under systematic ranking distortion the sign of G_n is unrestricted and can turn negative (Proposition 3(c)).

J_t is research-specific human capital, a reduced-form capability to evaluate candidate arguments and to formulate arguments from evidence in a field. "Judgement" refers to its evaluative component. Candidate value q is taken as given, so deciding what makes a question worth asking lies outside the model. How J enters G is therefore a substantive modelling choice, and Remark 5 shows that it determines the sign of G_{nJ}. Future judgement is

$$J_{t+1} = (1-\delta)J_t + L(g(d_R), r(d), v(d), f(d); J_t).$$

The inputs are the author's own generation g, active revision r, verification v, and feedback f. Learning may depend on the current stock J_t. No sign is imposed on the response of g to d_L, or on the responses of r, v, and f, because those signs depend on how the workflow is organised. The functions are differentiable where marginal effects are taken, with one-sided derivatives at the boundaries of the delegation intervals. Writing $n = n(d_L)$ assumes that representation delegation does not itself change the number of checked candidates. Verification affects current quality through the count and accuracy of selection, and the reduced form places any remaining selection error in b.

Let $\omega > 0$ be the weight placed on future judgement. With $W_t = Q_t + \omega J_{t+1}$, the marginal

effect of either delegation coordinate $k \in \{L, R\}$ is

$$\frac{dW_t}{dd_k} = \frac{dQ_t}{dd_k} + \omega \frac{dJ_{t+1}}{dd_k}.$$

This derivative identifies a marginal trade-off without characterising an optimum; any gain in current quality must be weighed against the value of the induced change in future judgement.

Take two author types $i \in \{H, L\}$ with $J_H > J_L > 0$.

Assumption 2 (Gap comparison)**.**

(i) Both types face the same G, L, $n(\cdot)$, and $g(\cdot)$, with $n'(d_L) > 0$. Over the comparison, both have the same realised d_L and the same unbiased selection rule $b \equiv 0$.

(ii) G is submodular in checked candidates and judgement, $G_{nJ} \leq 0$ (the formulation case of Remark 5).

(iii) L is submodular in own generation and judgement, $L_{gJ} \leq 0$.

(iv) Representation delegation is weakly decreasing in judgement, $d_R^L \geq d_R^H$.

(v) Own generation is weakly decreasing in representation delegation and weakly productive:

$$g'(d_R) \leq 0, \qquad L_g \geq 0.$$

(vi) To isolate the generation channel, r, v, and f are equal across types and constant in d_R on $[0, d_R^L]$.

Proposition 2 (Assisted convergence and counterfactual divergence)**.** *Under Assumption 2:*

(a) the assisted current-quality gap $Q_t^H - Q_t^L$ is weakly decreasing in d_L; and

(b) relative to the counterfactual $d_R^H = d_R^L = 0$, representation delegation weakly enlarges the next-period judgement gap $J_{t+1}^H - J_{t+1}^L$.

Neither statement compares the absolute next-period gap with the initial gap $J_H - J_L$.

Proof. For part (a),

$$\Gamma(d_L) = G(n(d_L), J_H, 0) - G(n(d_L), J_L, 0),$$

so

$$\Gamma'(d_L) = n'(d_L)\big[G_n(n(d_L), J_H, 0) - G_n(n(d_L), J_L, 0)\big] \leq 0.$$

The sign follows from $n' > 0$ and submodularity.

For part (b), let $g_0 = g(0)$ and $g_i = g(d_R^i)$, with learning loss

$$\Lambda_i = L(g_0, \cdot; J_i) - L(g_i, \cdot; J_i) = \int_{g_i}^{g_0} L_g(s, \cdot; J_i)\, ds \geq 0,$$

where the dots denote the common fixed values of r, v, and f. Assumptions (iv) and (v) give $g_L \leq g_H \leq g_0$. Submodularity gives $L_g(s, \cdot; J_L) \geq L_g(s, \cdot; J_H)$. Therefore

$$\Lambda_L \geq \int_{g_H}^{g_0} L_g(s, \cdot; J_L)\, ds \geq \int_{g_H}^{g_0} L_g(s, \cdot; J_H)\, ds = \Lambda_H.$$

The change in the judgement gap relative to no delegation is $\Lambda_L - \Lambda_H \geq 0$. □

Part (a) is strict if $G_{nJ} < 0$ on a set of positive measure between J_L and J_H. Part (b) is strict if either the marginal product of generation is strictly higher for the low type on a non-trivial common interval, or $g_L < g_H$ and generation is strictly productive on part of the additional interval. Strict delegation ordering alone is insufficient when g is locally flat.

Remark 3 (Scope of the gap result)**.** Assumptions (ii), (iv), and (vi) are behavioural restrictions, not results derived from endogenous delegation choices. Holding the other learning inputs (r, v, f) constant is a strong restriction; if an interface raises them more for the author who delegates more, the learning comparison can reverse. Likewise, if $G_{nJ} > 0$, candidates and judgement are complements and assistance widens the current-quality gap. Remark 5 shows that this holds when judgement reduces evaluation noise, so part (a) holds in the formulation case and reverses in the evaluation case. Which case better describes scholarly writing is an empirical question.

Remark 4 (Interaction with anchoring)**.** The proposition sets $b = 0$ to isolate learning. Once b varies, the current-quality effect of d_R for type i includes $G_b\, b_{d_R}(d_R^i, J_i)$. If a generated first representation distorts the low type's selection more strongly, comparing this term across types requires restrictions on b_{d_R} and its variation with J, which Assumption 2 does not supply. Predictions 2 and 3 can therefore operate together, though their combined effect depends on assumptions about both ranking distortion and learning.

4. Prediction 3: anchoring and defaults

Let x_0 be the first generated draft and u_t denote value uncertainty. Write $s(x_i, x_0)$ for the similarity of candidate x_i to x_0. A reduced-form perceived score is

$$\widetilde{q}_i = q_i + \alpha(J_t, u_t) s(x_i, x_0) + \varepsilon_i, \qquad \alpha_J < 0, \quad \alpha_u > 0.$$

The first draft acts as an anchor because candidates are scored partly by their similarity to it. This is an anchoring or fixation mechanism, distinct from reference dependence in which outcomes are valued as gains and losses relative to a benchmark. The sign $\alpha_u > 0$ reflects the finding that anchoring effects are larger when the range of plausible values is wider (Lee and Morewedge, 2022), and models in which a reference point provides comparative information under value uncertainty imply the same sign (Dean et al., 2026). The sign $\alpha_J < 0$, meaning that weaker judgement places greater weight on the generated draft, is supported by evidence from adjacent settings that domain

knowledge reduces anchoring (Smith, Windschitl and Bruchmann, 2013), although experts remain susceptible (Englich, Mussweiler and Strack, 2006). Neither sign has been tested in scholarly writing, and neither implies that an initial draft is harmful, since an informative anchor can make evaluation more accurate.

The next result separates random evaluation error from a systematic bias towards a canonical account. Parts (a) and (b) share a Gaussian candidate generator, whereas part (c) is an existence result in a separate bounded environment.

Proposition 3 (Selection under alternative ranking rules)**.** *Let $a_m = \mathbb{E}[\max_{i\le m} Z_i]$ for independent standard normal Z_i. In parts (a) and (b), $q_i \sim N(\mu, \sigma_q^2)$ with $\sigma_q > 0$.*

(a) If the author selects $\arg\max_i q_i$, then $\mathbb{E}[q_{\hat{\imath}}] = \mu + \sigma_q a_m$, which is strictly increasing in m.

(b) If the author instead selects on $\widetilde{q}_i = q_i + \varepsilon_i$, where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is independent with finite $\sigma_\varepsilon^2 \ge 0$, then

$$\mathbb{E}[q_{\hat{\imath}}] = \mu + \frac{\sigma_q}{\sqrt{\sigma_q^2 + \sigma_\varepsilon^2}}\,\sigma_q a_m.$$

This remains strictly increasing in m, though it is strictly attenuated for $m \ge 2$ when $\sigma_\varepsilon^2 > 0$.

(c) In a separate environment, each candidate is independently canonical with probability $\pi \in (0,1)$ and has true quality q_c. Otherwise its true quality is drawn from a non-degenerate bounded distribution H with $\mathbb{E}[H] > q_c$. Originals are scored at their true quality; a canonical candidate receives proxy score $q_c + \Delta_c \ge \sup \operatorname{supp}(H)$ and is selected when tied. Then

$$f_m \equiv \mathbb{E}[q_{\hat{\imath}} \mid m] = (1-\pi)^m M_m + \big(1-(1-\pi)^m\big) q_c, \qquad M_m = \mathbb{E}[\max_{i\le m} H_i].$$

The sequence f_m is unimodal up to a possible two-point plateau and is eventually strictly decreasing. It converges to q_c, whereas truth-based selection from the mixture converges to $\operatorname{ess\,sup} H > q_c$.

Proof. Part (a) is the Gaussian order statistic. For part (b), put $S_i = q_i + \varepsilon_i$ and $\tau^2 = \sigma_q^2 + \sigma_\varepsilon^2$. Gaussian projection gives

$$\mathbb{E}[q_i \mid S_i] = \mu + \frac{\sigma_q^2}{\tau^2}(S_i - \mu).$$

The index $\hat{\imath}$ is measurable with respect to the full signal vector. Independence across candidates gives $\mathbb{E}[q_i \mid S_1, \dots, S_m] = \mathbb{E}[q_i \mid S_i]$. Therefore

$$\mathbb{E}[q_{\hat{\imath}} \mid S_1, \dots, S_m] = \sum_i \mathbf{1}\{\hat{\imath} = i\}\, \mathbb{E}[q_i \mid S_1, \dots, S_m] = \mu + \frac{\sigma_q^2}{\tau^2}(S_{\hat{\imath}} - \mu).$$

Taking expectations and using $\mathbb{E}[S_{\hat{\imath}}] = \mu + \tau a_m$ gives the result.

For part (c), if no canonical candidate appears, an original with expected quality M_m is kept; otherwise a canonical candidate is kept. This proves the displayed formula. Write $r = 1 - \pi$ and

$f_m = q_c + t_m$, where $t_m = r^m(M_m - q_c) > 0$. Then

$$\frac{t_{m+1}}{t_m} = r\left(1 + \frac{M_{m+1} - M_m}{M_m - q_c}\right).$$

Lemma 1, applied to H, makes the numerator in the fraction strictly decreasing and the denominator strictly increasing. The ratio is therefore strictly decreasing and converges to $r < 1$. The sequence rises while the ratio exceeds one, can be flat for one step if the ratio equals one, and falls thereafter. Bounded convergence gives $f_m \to q_c$. Under truth-based selection, the candidate-quality distribution is the mixture of an atom at q_c and H. Its maximum converges almost surely to the mixture's essential supremum, which is $\operatorname{ess\,sup} H > q_c$; bounded convergence then gives the stated expectation limit. □

For example, with $H = U[0,1]$, $\pi = 0.2$, $q_c = 0.3$, and $\Delta_c \geq 0.7$, expected quality rises from $f_1 = 0.46$ to its maximum $f_2 \simeq 0.5347$ at $m = 2$. Additive independent Gaussian noise cannot produce this decline because it only increases σ_ε^2 in part (b); that property belongs specifically to the Gaussian benchmark and does not cover every possible distribution of evaluation error. Part (c) provides a discrete illustration of the reduced form above, with Δ_c representing a similarity premium αs_c assigned to the canonical type, and the dominance bound on Δ_c is a sufficient condition that does not imply reversal under smaller anchoring biases.

Remark 5 (Two readings of judgement, two signs of G_{nJ})**.** The sign of the interaction between checked candidates and judgement depends on which capability J represents. Both cases below assume unbiased ranking and hold checking costs and the candidate-quality distribution fixed apart from the stated role of J.

Evaluation. Let candidates be drawn as in part (b), and let judgement lower evaluation noise, with finite $\sigma_\varepsilon^2 = v(J) > 0$ and $v'(J) < 0$. Then

$$G(m,J) = \mu + \frac{\sigma_q^2\, a_m}{\sqrt{\sigma_q^2 + v(J)}}, \qquad G(m+1,J) - G(m,J) = \frac{\sigma_q^2\,(a_{m+1} - a_m)}{\sqrt{\sigma_q^2 + v(J)}}.$$

Because a_m is strictly increasing, the increment is strictly increasing in J. In the continuous relaxation on $m > 1$,

$$G_{mJ} = -\frac{\sigma_q^2\, a'(m)\, v'(J)}{2\big[\sigma_q^2 + v(J)\big]^{3/2}} > 0.$$

An additional candidate is worth more to a more accurate judge, so search and evaluation are complements and a common expansion in m widens the assisted current-quality gap.

Formulation. Let judgement instead raise the quality $\mu(J)$ of the author's own account, with $\mu'(J) > 0$, and let the m machine candidates be independent draws from a non-degenerate distribution F that does not depend on J and has a finite upper-tail expectation. Selection is accurate

and the author keeps the best of her own account and the machine candidates, so

$$G(m, J) = \mathbb{E}\big[\max\{\mu(J), Y_1, \ldots, Y_m\}\big], \qquad G(m+1, J) - G(m, J) = \int_{\mu(J)}^{\infty} F(x)^m \big(1 - F(x)\big)\, dx.$$

The increment is non-negative and weakly decreasing in J, because a higher J raises the lower limit of integration. It is strictly positive whenever $F(\mu(J)) < 1$, so that machine draws have positive probability of exceeding $\mu(J)$, and vanishes when $F(\mu(J)) = 1$. Where F is continuous at $\mu(J)$, its derivative in J is $-\mu'(J)F(\mu(J))^m\big(1 - F(\mu(J))\big)$, which is strictly negative on the interior domain $0 < F(\mu(J)) < 1$ (where both $G_m > 0$ and $G_J > 0$ also hold strictly); at an atom of F or at the boundaries, only the weak statement holds. Additional machine candidates are worth at least as much to the author whose own account is weaker, so search and formulation are substitutes, matching the submodularity condition in Assumption 2(ii).

Each case therefore rests on its own assumptions, and neither follows from the other. Where both capabilities matter, the sign of G_{mJ} depends on which binds.

Lemma 2 (The present-bias stopping wedge)**.** *Suppose one further attempt costs $e > 0$ now and yields expected benefit $\Delta q_m > 0$ one period later. Utility is linear in quality and the common long-run discount factor is normalised to one. An indifferent author stops. A patient author continues when $\Delta q_m > e$. A quasi-hyperbolic author who weights the delayed benefit by $\beta \in (0, 1)$ continues when $\beta \Delta q_m > e$. Thus*

$$e < \Delta q_m \leq \frac{e}{\beta}$$

is the exact interval on which the patient author continues and the present-biased author stops at that decision. If Δq_m is weakly decreasing in successive attempts, the present-biased author stops no later than the patient author and strictly earlier whenever a reachable increment lies in this interval.

Proof. The patient net benefit is $\Delta q_m - e$ and the present-biased current self's net benefit is $\beta \Delta q_m - e$. Comparing each with zero partitions Δq_m into three intervals. Both stop when $\Delta q_m \leq e$, only the patient author continues on the displayed interval, and both continue when $\Delta q_m > e/\beta$. When increments decline, the continuation set under the higher threshold e/β is a subset of the continuation set under e, which proves the ordering of stopping times. □

Remark 6 (Satisficing)**.** A satisficing author stops at the first option whose perceived quality reaches an aspiration level A (Simon, 1955; Caplin, Dean and Martin, 2011). A generated draft with $\widetilde{q}_0 \geq A$ ends search before the author makes any attempt of her own. Without the tool, and for an author who must produce a manuscript and holds an aspiration above the blank-page outcome, the zero-effort option never meets the aspiration. Satisficing differs from the wedge in Lemma 2 because it specifies an aspiration-based stopping rule, whereas present bias explains why the author stops even though the patient benchmark would continue. Neither result implies that present bias lowers A, which would require a separate model.

Anchoring and default acceptance operate at different margins, since anchoring alters how examined candidates are ranked whereas satisficing and present bias reduce how many candidates are examined.

5. Prediction 4: prose as a signal

Let Y denote research quality and S an observed prose score such as fluency. Let $A \in \{0, 1\}$ denote use of assistance, with $p = \Pr(A = 1)$, and define

$$\mu_a^Y = \mathbb{E}[Y \mid A = a], \qquad \mu_a^S = \mathbb{E}[S \mid A = a], \qquad \kappa_a = \mathrm{Cov}(Y, S \mid A = a).$$

This reduced form does not replace a signalling equilibrium; it records what any account of pooled attenuation or inversion must establish. In costly-signalling models, a technology can change κ_a by altering how signal costs vary with type (Spence, 1973; Gans, 2024; Galdin and Silbert, 2025; Cowgill, Hernández-Lagos and Wright, 2026).

Proposition 4 (Attenuation and the requirements for inversion)**.** *The pooled association satisfies*

$$\mathrm{Cov}(Y, S) = (1 - p)\kappa_0 + p\kappa_1 + \mathrm{Cov}(\mu_A^Y, \mu_A^S).$$

For binary A, the last term is $p(1 - p)(\mu_1^Y - \mu_0^Y)(\mu_1^S - \mu_0^S)$. Consequently:

(a) if mean research quality is equal across adoption groups, so $\mu_0^Y = \mu_1^Y$, and $0 \le \kappa_1 \le \kappa_0$, then increasing the share of assisted prose weakly attenuates the pooled covariance and cannot invert it;

(b) if assistance merely removes an independent language disturbance from prose, the covariance with research quality need not fall and the correlation can rise because the variance of irrelevant noise falls; and

(c) a negative pooled relation requires a negative within-group covariance or a sufficiently negative between-group term. Selection into adoption by language cost that is independent of research quality supplies neither condition by itself.

Proof. The displayed equation is the law of total covariance. Under mean independence the between-group term is zero, so part (a) is a weighted average of two non-negative covariances and has derivative $\kappa_1 - \kappa_0 \le 0$ with respect to p when the conditional distributions are held fixed. For part (b), write unassisted prose as $S_0 = Y + \eta + \ell$ and assisted prose as $S_1 = Y + \eta$, where Y, η, and the language disturbance ℓ are mutually independent apart from their displayed addition, and η and ℓ have mean zero. Both covariances equal $\mathrm{Var}(Y)$, while $\mathrm{Var}(S_1) = \mathrm{Var}(S_0) - \mathrm{Var}(\ell)$, so the correlation rises when $\mathrm{Var}(\ell) > 0$. Part (c) is the contrapositive of the decomposition when the within-group terms are non-negative. □

The proposition shows why covariance attenuation across the pool and improved screening of second-language writers can occur together, since they reflect different changes in how prose signals are produced. It also shows why cheaper English alone does not turn the pooled covariance negative; part (c) states necessary conditions for inversion without making them sufficient. Part (a) is a compositional comparison that holds group-specific moments fixed and therefore requires marginal adopters to share the assisted group's stated moments. Finally, covariance attenuation does not by itself imply a lower correlation or less accurate screening, because both the variance of S and the reader's updating rule can change.

6. Fifth claim: field effects

Private iteration and field-level aggregation are distinct problems. Let $x_j^\star$ be researcher j's idiosyncratic optimum in a scalar projection of argument space. Let $x_j(m)$ be the argument kept after m candidates in an exchangeable population of researchers. For a welfare illustration, let $z_j = h(x_j)$ be a scalar prediction for a bounded continuous map h, and let y be the realised target.

Proposition 5 (Diversity and screening capacity)**.**

(a) Under own-truth selection, suppose $x_j(m) = x_j^\star + e_j(m)$*, with* $\mathbb{E}[e_j(m)] = 0$*,* $\mathrm{Cov}(x_j^\star, e_j(m)) = 0$*, and* $\mathrm{Var}(e_j(m)) \to 0$*. Then* $\mathrm{Var}(x_j(m)) \to \mathrm{Var}(x_j^\star)$*, which need not be zero. Under a common canonical proxy, suppose the same canonical point* x_c *for every researcher is generated independently with probability* $\pi > 0$ *on each draw and is selected whenever present. On a bounded argument space,* $x_j(m) \to x_c$ *in mean square and cross-sectional argument variance converges to zero.*

(b) For the equal-weight pooled prediction $\bar{z} = N^{-1}\sum_j z_j$*,*

$$(\bar{z} - y)^2 = \frac{1}{N}\sum_j (z_j - y)^2 - \frac{1}{N}\sum_j (z_j - \bar{z})^2.$$

Holding average individual squared error fixed, a fall in prediction diversity raises the pooled squared error by the same amount.

(c) Let the positive differentiable functions $M(a)$ *and* $K(a)$ *be the number of submitted manuscripts and the number that available attention can screen at technology level* a*. On an interval where* $M(a) > K(a)$*, the screened share is* $s(a) = K(a)/M(a)$*. It falls if and only if* $d\ln M(a)/da > d\ln K(a)/da$*.*

Proof. For own-truth selection, $\mathrm{Var}(x_j(m)) = \mathrm{Var}(x_j^\star) + \mathrm{Var}(e_j(m))$. Under the common proxy, $\Pr(x_j(m) \neq x_c) = (1-\pi)^m$. Boundedness then gives mean-square convergence and hence variance convergence. For part (b), write $z_j - y = (z_j - \bar{z}) + (\bar{z} - y)$, then square and average across j; the cross term is zero. Part (c) follows by log-differentiating $s = K/M$. □

Part (a) assumes the link between more search and smaller spatial error; Proposition 1 alone proves improvement in selected quality, without implying convergence in argument space. Part (b) is an exact identity only for the displayed aggregation rule and squared loss, and does not establish that every form of scholarly diversity raises welfare. A shared selector can nevertheless produce correlated losses that comparisons of individual accuracy omit, as in models of algorithmic monoculture (Kleinberg and Raghavan, 2021). The separate learning externality in Acemoglu, Kong and Ozdaglar (2026) concerns the stock of general knowledge and should be kept distinct from both correlated selection and screening congestion. Part (c) is a capacity identity, so a falling screened share does not by itself establish lower screening accuracy or lower welfare.

7. Scope of the formal results

The five propositions do not form a general welfare theorem for AI-assisted writing, and the welfare derivative in Section 3 identifies a marginal trade-off without ranking complete workflows or proving that the article's recommended sequence of work is optimal. Each proposition isolates one mechanism under explicit restrictions stated in its section. Assistance can improve a manuscript without building judgement, or support both when paired with active human inputs; linguistic delegation need not distort selection, and a shared generator need not homogenise arguments when authors evaluate candidates independently.

References

Acemoglu, Daron, Dingwen Kong, and Asuman Ozdaglar. 2026. "AI, Human Cognition and Knowledge Collapse." National Bureau of Economic Research Working Paper 34910.

Caplin, Andrew, Mark Dean, and Daniel Martin. 2011. "Search and Satisficing." *American Economic Review*, 101(7): 2899–2922.

Cowgill, Bo, Pablo Hernández-Lagos, and Nataliya Langburd Wright. 2026. "Does AI Cheapen Talk? Theory and Evidence from Global Entrepreneurship and Hiring." *Management Science*. Articles in Advance.

Dean, Mark, Benjamin Enke, Thomas Graeber, and Pietro Ortoleva. 2026. "Reference Points as Information." Working paper, March 2026.

de Lara, Lucas, and Mark Dean. 2025. "Rational Choice Overload." Working paper, July 2, 2025.

Englich, Birte, Thomas Mussweiler, and Fritz Strack. 2006. "Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making." *Personality and Social Psychology Bulletin*, 32(2): 188–200.

Galdin, Anaïs, and Jesse Silbert. 2025. "Making Talk Cheap: Generative AI and Labor Market Signaling." Working paper. arXiv:2511.08785.

Gans, Joshua S. 2024. "How Will Generative AI Impact Communication?" *Economics Letters*, 242: 111872.

Kleinberg, Jon, and Manish Raghavan. 2021. "Algorithmic Monoculture and Social Welfare." *Proceedings of the National Academy of Sciences*, 118(22): e2018340118.

Lee, Chang-Yuan, and Carey K. Morewedge. 2022. "Noise Increases Anchoring Effects." *Psychological Science*, 33(1): 60–75.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*, 69(1): 99–118.

Smith, Andrew R., Paul D. Windschitl, and Kathryn Bruchmann. 2013. "Knowledge Matters: Anchoring Effects Are Moderated by Knowledge Level." *European Journal of Social Psychology*, 43(1): 97–108.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics*, 87(3): 355–374.

Weitzman, Martin L. 1979. "Optimal Search for the Best Alternative." *Econometrica*, 47(3): 641–654.