Writing Is Not Thinking: analytical supplement

Search, learning, selection and field effects*

Johan Fourie†

This supplement states the assumptions behind the article's five predictions and derives the results that require formal support. It uses one allocation-and-selection framework. Writing produces candidate arguments, judgment selects among them, and some writing activities also change future judgment. Reference dependence changes the ranking rule; signalling and field effects arise when the same technology is used across authors. The accompanying R script checks the displayed examples and the signs used below. Those checks are corroborating calculations, not proofs.

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 meaningful candidate at execution cost c_v; 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 is not an iteration in this sense.

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 appears when the author branches from that draft and must reproduce text, which raises c_v.*

The search result fixes the effort budget, candidate-quality distribution and selection rule. It also treats candidates as independent draws from F even when the author branches from an incumbent; branching affects c_v but not the distribution in this benchmark. Later sections relax these restrictions one at a time. This separation is substantive: lowering execution cost need not improve the selected argument when composing changes the candidates or when selection is distorted.

*This is the supplementary material for Fourie, Johan. 2026. "Writing Is Not Thinking." Working Paper, Department of Economics, Stellenbosch University. I used Claude Opus 4.8, Claude Sonnet, Claude Fable 5 and OpenAI Codex for literature mapping, mathematical derivation, drafting and revision, and Claude and Codex for independent audits of the formal results. I verified the formal results analytically and numerically, and accept responsibility for the argument and any errors. Cite the article, not this supplement.

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

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; 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; finite absolute mean alone does not guarantee it.

Proposition 1 (Cheaper iteration and the distribution of gains)**.** *Suppose Assumption 1 holds, candidates have the common 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

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

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; 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

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

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 uses monotonicity, not differentiability or concavity. Common F, B and c_s are essential: the result does not rank writers who also differ in candidate quality, judgment or checking cost. □

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. The derivative in part (b) belongs only to the continuous relaxation.

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 judgment 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, opening 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 $F(z) < 1$, $dz/dc = -1/(1 - F(z)) < 0$: 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 better incumbent can shorten search rationally (Weitzman 1979). Recent search theory also shows that a larger menu does not produce greater default choice in the constant-cost, known-distribution benchmark; learning about the distribution, rising search costs or a search depth chosen in advance can do so (de Lara and Dean 2025). Section 4 adds a separate present-bias wedge.

3 Prediction 2: the manuscript and the writer

The static result holds judgment fixed. Delegation requires two coordinates. Let $d_L \in [0, 1]$ be the share of linguistic execution delegated to a machine and $d_R \in [0, 1]$ its share of forming the

first explicit representation of the problem. An answer-first workflow raises both. An author-first workflow can have high d_L and low d_R. Write $d = (d_L, d_R)$.

Current manuscript quality is

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

where J_t is current judgment and b is distortion in selection. Future judgment 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: those signs depend on workflow design. The functions are differentiable where marginal effects are taken; derivatives at the boundaries of the delegation intervals are one-sided. 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; the reduced form places any remaining selection error in b.

Let $\omega > 0$ be the weight placed on future judgment. 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 identity does not characterise an optimum. It says only that a current-quality gain must be compared with the value of the induced change in future judgment.

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 $n'(d_L) > 0$, the same realised d_L, and the same selection rule $b \equiv 0$ over the comparison.

(ii) G is submodular in checked candidates and judgment: $G_{nJ} \leq 0$.

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

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

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

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

(vi) The generation channel is isolated: 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 judgment 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), write $g_0 = g(0)$, $g_i = g(d_R^i)$ and define the 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 judgment 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)**.** The condition that the generation channel operates alone is restrictive. If an interface raises revision, verification or feedback more for the author who delegates more, the learning result can reverse. Likewise, if $G_{nJ} > 0$, candidates and judgment are complements and assistance can widen the current-quality gap. These reversals are implications of the model, not failures of it.

Remark 4 (Interaction with reference dependence)**.** The proposition sets $b = 0$ to isolate learning. If a generated first representation distorts the low type's selection more strongly, the current-quality effect of d_R depends additionally on $G_b b_J$ and cannot be signed from Assumption 2. Predictions 2 and 3 can therefore operate together, but their combined effect requires assumptions on the distortion as well as on learning.

4 Prediction 3: reference points and defaults

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

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

Current reference-point theory derives stronger reference effects under greater value uncertainty in its own environment (Dean et al. 2026). The writing application represents that result as $\alpha_u > 0$. The sign $\alpha_J < 0$—weaker judgment gives the generated draft more weight—is this article's conjecture. Neither sign establishes that the first draft is harmful: an informative reference can improve evaluation.

The next result separates random error from systematic canonical pull. Parts (a) and (b) share a Gaussian generator. Part (c) is an existence result in a different, bounded environment.

Proposition 3 (Selection under truth, noise and canonical pull)**.** *Let $a_m = \mathbb{E}[\max_{i\le m} Z_i]$ for independent standard normal Z_i.*

(a) If $q_i \sim N(\mu, \sigma_q^2)$ and the author selects $\arg\max_i q_i$, then $\mathbb{E}[q_{\hat{\imath}}] = \mu + \sigma_q a_m$, strictly increasing in m when $\sigma_q > 0$.

(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, then

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

For finite noise this remains increasing in m, but 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 wins ties. 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 $\mathbb{E}[q_i \mid S_1, \ldots, S_m] = \mathbb{E}[q_i \mid S_i]$. Therefore

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

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 $\Delta_c \geq 0.7$, $f_1 = 0.46$, $f_2 \simeq 0.5347$ and the maximum occurs at $m = 2$. Additive independent Gaussian noise cannot produce this decline; it only adds to σ_ε^2 in part (b). That is a statement about the Gaussian environment, not a necessity theorem for every form of noise. Part (c) is a discrete illustration of the reduced form above: Δ_c can be read as a similarity premium αs_c assigned to the canonical type. The dominance bound on Δ_c is a sufficient condition, not a claim that smaller reference-point effects generally reverse the value of search.

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, the common long-run discount factor is normalised to one, and 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 exactly the region in 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 region.

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 gives the three regions: 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. □

Reference dependence and present bias are different mechanisms. The first changes the ranking of candidates already examined; the second can reduce search below the patient benchmark. A larger candidate menu alone produces neither result in the standard known-distribution search model.

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 changing 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 explains why covariance attenuation and improved screening of second-language writers can coexist. They concern different changes in signal production. It also shows why a negative pooled covariance is not a consequence of cheaper English alone; part (c) gives necessary, not sufficient, conditions for inversion. Part (a) is a mixing comparison that holds group-specific moments fixed; it requires marginal adopters to share the assisted group's stated moments. Covariance attenuation does not by itself imply a lower correlation, a less informative posterior or worse screening, because the variance of S and the reader's updating can change.

6 Prediction 5: field effects

Private iteration and field-level aggregation are different 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 wins 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 exactly when* $d\ln M(a)/da > d\ln K(a)/da$*.*

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 $z_j - y = (z_j - \bar{z}) + (\bar{z} - y)$, square and average; 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, not convergence in argument space. Part (b) is an exact identity only for the displayed aggregation rule and squared loss. It does not establish that every form of scholarly diversity raises welfare. A shared selector can nevertheless create correlated losses that private accuracy comparisons miss, 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 not be conflated with either correlated selection or screening congestion. Part (c) is a capacity identity: a falling screened share does not by itself establish a fall in screening accuracy or welfare.

7 Scope of the formal results

The five propositions do not form a general welfare theorem for AI-assisted writing. Proposition 1 fixes candidate production and accurate selection. Proposition 2 isolates a generation-only learning channel. Proposition 3(c) is an existence result in a deliberately bounded environment. Proposition 4 states conditions on signal content rather than deriving an equilibrium response to cheaper prose. Proposition 5 uses explicit aggregation and capacity rules.

These restrictions are the formal content of the article's boundary for use. Assistance can improve a manuscript without improving judgment; it can also improve both when it adds revision, verification or feedback. Linguistic assistance need not distort selection, and a common generator need not homogenise arguments when authors retain independent evaluation. The results identify the conditions under which each prediction follows and the conditions under which it can reverse.

References

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

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.

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.

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.