How many scientists? On Epistemic Diversity in Expanding-Frontier Discovery

Anonymous authors
Paper under double-blind review

Abstract

Scientific discovery is inherently sequential: each experiment updates beliefs and changes which hypotheses are worth testing next. As budget-limited discovery systems such as laboratory-in-the-loop agents are deployed at scale, this raises a basic organizational question: should discovery be driven by a single shared belief state, or distributed across several persistent searches? We formalize this question as an adaptive search process over an expanding hypothesis graph, and introduce `ISLAND` (Independent Search over Local Adaptive Neighborhoods for Discovery), which allocates a fixed experimental budget across persistent epistemic states with locally expanding frontiers. Multiple states can preserve distinct search directions and local coverage, but they fragment the budget each trajectory needs to resolve prerequisite hypotheses. Theoretically, we show that fragmentation limits what each isolated trajectory can reach, while heterogeneous commitments and prior diversity can offset this loss, yielding a budget-dependent crossover from concentrated to distributed search. Controlled simulations recover this regime structure, and on the SciGym systems-biology benchmark, distributed search likewise overtakes pooled search as the experimental budget grows. Applying these principles to gene regulatory network recovery on CausalBench, we find that persistent local search states improve over a matched single-state variant and existing baselines on K562. Together, these results identify the number and isolation of persistent search states as design variables for sequential scientific discovery.

1 Introduction

Scientific discovery proceeds through repeated cycles of hypothesis formation, experimentation, and revision; subsequently, the outcome of one experiment changes what is believed about a system and which questions are worth pursuing next (King et al., 2009). Recent advances in computational modelling, automated experimentation, and scientific agents increasingly support this process (Szymanski et al., 2023; Swanson et al., 2025; Ghareeb et al., 2026). Computational agents can generate hypotheses, search literature, run simulations, select experiments, and update their plans (Boiko et al., 2023; Lu et al., 2026; Gottweis et al., 2026). Together with increasingly automated laboratories, these systems make it possible to maintain and evaluate multiple candidate directions over successive rounds of experimentation (Frey et al., 2025; Burger et al., 2020).

Experimental capacity, however, remains limited (Jacobsson, 2026; Cornelio et al., 2026). Biological experiments consume samples, reagents, sequencing capacity, and experimental time, while computational laboratories may similarly impose finite simulation or evaluation budgets. For example, gene regulatory network discovery over hundreds of genes involves hundreds of thousands of possible directed interactions, while contemporary techniques such as CRISPR and Perturb-seq screens require perturbational resources that are limited per run (Dixit et al., 2016; Adamson et al., 2016; Replogle et al., 2022; Chevalley et al., 2025). Efficient discovery therefore requires deciding how a finite experimental budget should be allocated. In many such settings, experiments do more than update beliefs over a fixed list of hypotheses: they can make new follow-up hypotheses visible, meaningful, or actionable. A discovery system therefore searches an *expanding frontier*, where what can be tested later depends on what has been resolved earlier.

Active learning, Bayesian experimental design, Bayesian optimization, and bandit methods study how limited budgets should be allocated to informative queries or actions (Settles, 2009; Houlsby

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2011; Srinivas et al., 2010; Kalyanakrishnan et al., 2012). Batch acquisition methods can additionally encourage diversity among simultaneously selected queries (Kirsch et al., 2019), while distributed bandits consider multiple learners operating over shared action spaces (Landgren et al., 2021; Martínez-Rubio et al., 2019) (Extended related work in A). These approaches primarily study how to select informative actions given a current decision state or candidate set. Expanding-frontier discovery introduces an additional organizational question: when previous experiments determine which hypotheses become accessible next, should evidence and exposed hypotheses be pooled into one evolving search state, or maintained across several persistent states with separate histories?

We study how *epistemic diversity*, arising from different seeds, priors, and search histories, affects discovery when experiments expand the set of hypotheses available for later testing. A single pooled search state has the depth to resolve long chains of prerequisites, but a single search with a biased prior can commit the whole budget to a wrong direction. In contrast, several persistent states could hedge against a poor commitment, but each has less depth. We formalize this commitment-fragmentation tradeoff in adaptive search processes over an expanding hypothesis graph, introduce `ISLAND` (Independent Search over Local Adaptive Neighborhoods for Discovery), and show that it yields a budget-dependent crossover from concentrated to distributed search, which appears in controlled worlds and in the SciGym simulator (Duan et al., 2025). We apply this to a discovery problem in gene regulatory network recovery on CausalBench (Chevalley et al., 2025), and show that separate evidence-driven gene neighbourhoods improve observational network recovery over a matched single-state variant and compared baselines.

Contributions. Our core contributions are threefold:

1. **Conceptually**, we formulate expanding-frontier discovery as finite-budget adaptive search over an expanding hypothesis graph and identify a commitment-fragmentation tradeoff between trajectory depth and epistemic diversity.
2. **Theoretically**, we characterize how budget fragmentation, discovery-time distributions, and heterogeneous commitments determine when distributed search can outperform concentrated search.
3. **Empirically**, we demonstrate the resulting budget-dependent regime structure in controlled simulations and SciGym, and show that trajectory-dependent search improves gene regulatory network recovery on CausalBench.

2 `ISLAND`: Expanding-Frontier Discovery

We formalise the core `ISLAND` setting as finite-budget adaptive search over an expanding hypothesis graph, with the full model and conventions in App. B.1. Each persistent *epistemic state* maintains its own prior, observations, and locally exposed hypotheses; states never share updates, but draw on a common experimental budget (Fig. 1).

World and hypothesis graph. Let $\mathcal{V}$ be a set of n scientific variables and $\mathcal{H}$ a finite set of candidate hypotheses. Each hypothesis h concerns the variables in its *scope* $\mathcal{S}_h \subseteq \mathcal{V}$, $|\mathcal{S}_h| \geq 2$, and has an unknown truth value $Z_h \in \{0, 1\}$, so hypotheses are hyperedges of a scientific hypergraph over $\mathcal{V}$. The *hypothesis graph* $(\mathcal{H}, \mathcal{E})$ connects hypotheses whose scopes differ by adding, removing, or replacing one variable:

$$(h, g) \in \mathcal{E} \iff d_{\text{edit}}(\mathcal{S}_h, \mathcal{S}_g) = 1, \qquad d_{\text{edit}}(\mathcal{S}, \mathcal{T}) = \max\{|\mathcal{S} \setminus \mathcal{T}|, |\mathcal{T} \setminus \mathcal{S}|\}. \tag{1}$$

Write $\Gamma(h)$ for the neighbours of h, which determine which hypotheses can be exposed next.

Observations and beliefs. For analytical tractability, each hypothesis has a fixed but unknown latent effect $\mu_h \neq \tau$ and is true when it exceeds a known threshold τ, so $Z_h = \mathbf{1}\{\mu_h > \tau\}$. Each test of h yields an independent observation $Y \sim \mathcal{N}(\mu_h, \sigma^2)$ with known $\sigma^2 > 0$. State i holds a prior P_i under which $\mu_h \sim \mathcal{N}(\mu_{0,i}(h), \sigma^2/\kappa_{0,i}(h))$, with $\kappa_{0,i}(h) > 0$, independently across hypotheses, and updates it using only its own test-and-observation history $\mathcal{F}_{i,t}$, where local time t counts completed tests. With $m_{i,t}(h)$ and $v_{i,t}(h)$ the resulting conjugate posterior mean and variance (App. B.3) and Φ the standard normal CDF, the state's belief that h is true is

$$b_{i,t}(h) = \Pr_{P_i}(\mu_h > \tau \mid \mathcal{F}_{i,t}) = \Phi\left(\frac{m_{i,t}(h) - \tau}{\sqrt{v_{i,t}(h)}}\right). \tag{2}$$

Because the prior factorises, each test updates only the tested hypothesis's belief.

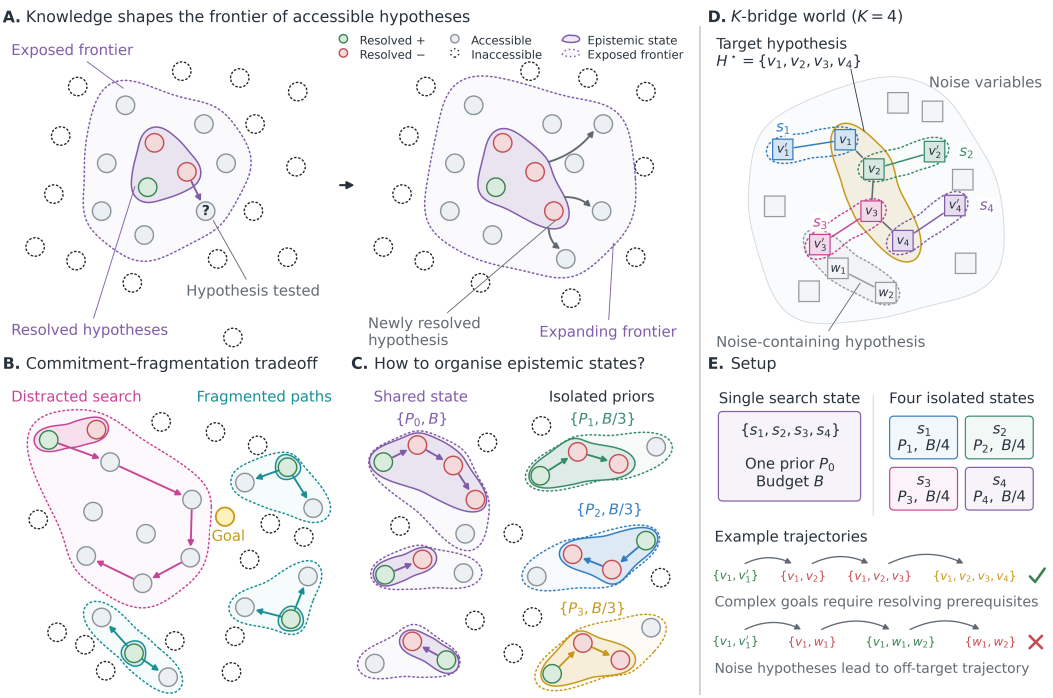


Figure 1: **Conceptual overview of ISLAND. (A)** Resolving hypotheses expands a state's locally exposed hypothesis set. **(B)** Under a fixed budget, distributed search trades trajectory depth against epistemic diversity. **(C)** We study when maintaining separate epistemic states improves discovery relative to one shared state, and how many states a given budget can support. **(D)** The K-bridge world used for analysis and controlled simulations; squares denote variables and enclosing hyperedges denote hypothesis scopes. **(E)** Example matched comparison between one state receiving all seeds and four isolated, prior-conditioned states.

Resolution and frontier expansion. Fix belief thresholds $0 < \eta_- < \frac{1}{2} < \eta_+ < 1$ and a minimum observation count $n_{\min} \geq 1$. Hypothesis h is *accepted by time* t, $h \in \mathcal{R}^+_{i,t}$, if at some $s \leq t$ it has received at least $n_{\min}$ observations and $b_{i,s}(h) \geq \eta_+$. Rejection, $h \in \mathcal{R}^-_{i,t}$, is defined analogously with $b_{i,s}(h) \leq \eta_-$. The *resolved* hypotheses are $\mathcal{R}_{i,t} = \mathcal{R}^+_{i,t} \cup \mathcal{R}^-_{i,t}$. These sets record historical threshold crossings rather than permanent decisions, and resolved hypotheses remain eligible for testing. State i begins from accepted *seed* hypotheses $\mathcal{R}_{i,0} = \mathcal{R}^+_{i,0}$, which represent prior knowledge and are exempt from the observation requirement. It can test only hypotheses in its cumulative *candidate set*, which grows whenever a hypothesis is first resolved, by acceptance or rejection (Fig. 1A):

$$\mathcal{C}_{i,0} = \mathcal{R}^+_{i,0} \cup \Gamma(\mathcal{R}^+_{i,0}), \qquad \mathcal{C}_{i,t+1} = \mathcal{C}_{i,t} \cup \Gamma(\mathcal{R}_{i,t+1} \setminus \mathcal{R}_{i,t}), \qquad \Gamma(\mathcal{A}) = \textstyle\bigcup_{h \in \mathcal{A}} \Gamma(h). \tag{3}$$

After t tests, state i selects $a_{i,t} \in \mathcal{C}_{i,t}$ using a history-adapted acquisition rule $\pi(\cdot \mid P_i, \mathcal{F}_{i,t}, \mathcal{C}_{i,t})$: uniform selection, softmax over posterior entropy, Thompson sampling, or an LUCB-style rule that prioritises hypotheses whose confidence intervals contain τ (App. B.4). The rule is held fixed within each comparison and does not depend on the remaining budget.

Experimental budget and objective. Following Bayesian experimental design (Rainforth et al., 2024), the budget B counts hypothesis tests, not the computation or memory required to maintain states. A system of M states splits it into integer local budgets $B_1, \ldots, B_M \geq 1$ that differ by at most one and sum to B, so $1 \leq M \leq B$. Seeds come from a fixed set $\mathcal{R}^+_{\text{known}}$ of known true hypotheses, which is excluded from evaluation, and each new true discovery counts once if any state has accepted it:

$$D_B = \sum_{h \in \mathcal{H} \setminus \mathcal{R}^+_{\text{known}}} Z_h \, \mathbf{1}\Big\{h \in \textstyle\bigcup_{i=1}^{M} \mathcal{R}^+_{i,B_i}\Big\}, \qquad J(B, M) = \mathbb{E}[D_B]. \tag{4}$$

Expectations are over observation, acquisition and prior randomness for a fixed world.

ISLAND: organising epistemic states. Priors come from a family $P(\omega)$ with $\omega \sim \nu$, where ω may specify a favoured search direction; draws are *shared* when $\omega_1 = \cdots = \omega_M$ and *diverse* when they are independent. The baseline is a single pooled state that receives all seeds, one prior $P_0 = P(\omega_0)$ with $\omega_0 \sim \nu$, and the whole budget (Fig. 1E), that is, the $M = 1$ member of the family. For a specified acquisition rule, prior family, and seed-allocation schedule, we seek $M^\star(B) \in \arg\max_{1 \le M \le B} J(B, M)$. An unrestricted controller could emulate every isolated state, and isolated execution is equivalent in law to running the same initialisations sequentially with private state (Lemma B.1). Thus, ISLAND is characterized by persistent, separated epistemic states rather than by the mechanics of parallel execution (Fig. 1C).

3 Theoretical Analysis

We introduce the K-bridge world (Fig. 1D), a hypothesis graph that isolates the competing effects of organising discovery across multiple epistemic states. Reaching a distal target requires resolving a chain of prerequisite hypotheses, and each resolution exposes further candidates that compete for tests. We derive the graph's constraints on any trajectory (Sec. 3.1), express budget splitting through discovery-time laws (Sec. 3.2), and identify mechanisms that offset fragmentation (Sec. 3.3); proofs are in App. B.

The K-bridge world. Let $K \ge 3$ and $X \ge 0$. The variables are bridge variables $v_1, \dots, v_K$, seed partners $v'_1, \dots, v'_K$ and noise variables $w_1, \dots, w_X$, with $n = 2K + X$, and every scope of size 2 to K is a hypothesis. The seeds $s_k = \{v_k, v'_k\}$ form $\mathcal{R}^+_{\text{known}}$, and the only other true hypothesis is the target $H^\star = \{v_1, \dots, v_K\}$, so $J(B, M)$ is the probability that some state accepts $H^\star$. Each edit changes a scope's *bridge count*, its number of bridge variables, by at most one, so $H^\star$ is exposed only after at least $K - 2$ intermediate hypotheses have been resolved, for example along $\{v_1, v'_1\} \to \{v_1, v_2\} \to \cdots \to H^\star$ (Lemma B.4). All intermediates are false, so correct search advances by rejecting them, although an erroneous acceptance also exposes their successors.

Pooled and isolated states. The pooled state (subscript 0) receives all K seeds, prior P_0 and budget B, whereas M isolated states split the seeds as evenly as possible (one each, round-robin, when $M > K$) and draw $\omega_1, \dots, \omega_M$ i.i.d. from ν (App. B.1). We write $J_{\text{single}}(B) = J(B, 1)$ and $J_{\text{iso}}(B, M)$ for the probabilities that the pooled state, or at least one isolated state, accepts $H^\star$.

3.1 Graph constraints: depth and frontier competition

Besides a minimum depth, the graph creates competition for tests. Every state starts with at least the $3n - 5$ candidates exposed by one seed, and a scope with $K - 1$ bridge variables has at least $K(n - K + 1)$ neighbours, of which only $H^\star$ increases the bridge count. Hence each uniform draw selects any given candidate with probability at most $1/(3n - 5)$, and selects $H^\star$ with probability at most $\bar{p} = 1/[K(n - K + 1) + 1]$ (Cor. B.6). Consider $m = n_{\min}$, the minimum number of observations before a hypothesis can be resolved.

Proposition 3.1 (Fragmentation in the bridge world)**.** *(i) Any state that accepts $H^\star$ has performed at least $(K-1)m$ tests. Hence $J_{\text{iso}}(B, M) = 0$ when $\lceil B/M \rceil < (K-1)m$, whereas $J_{\text{single}}(B) > 0$ for $B \ge (K-1)m$ under uniform selection. (ii) Under uniform selection, a state with local budget t accepts $H^\star$ with probability at most $F^+(t) = \binom{t}{(K-1)m}(3n-5)^{-(K-2)(m-1)}\,\bar{p}^{\,m}$, so $J_{\text{iso}}(B, M) \le M\,F^+(\lceil B/M \rceil)$. Both parts also hold under the final-posterior criterion of Sec. 4.*

Part (ii) follows the prerequisite chain that exposes $H^\star$. The first test of each prerequisite is unconstrained, but its remaining $m - 1$ tests must revisit that hypothesis among at least $3n - 5$ candidates. All m target tests must select $H^\star$ among at least $1/\bar{p}$ candidates. Thus, for $m \ge 2$, the competition penalty compounds with depth. When M divides B, the bound is at most $M^{1-(K-1)m}$ times a factor independent of M, an envelope on fragmented search rather than a claim that success is monotone in M. Excessive splitting can thus make the prerequisite depth unattainable and leaves each state few tests for the competition at every stage, before any benefit from heterogeneous initialisations is considered. App. B.4 extends these bounds to other acquisition policies.

3.2 BUDGET ALLOCATION THROUGH DISCOVERY TIMES

We now separate these graph constraints from the general effect of dividing a budget among independent trajectories. A state's *discovery time* T is the local time at which it first accepts $H^\star$. With local budget t, the state succeeds with probability $F(t) = \Pr(T \le t)$ and fails with probability $S(t) = 1 - F(t)$, and the *cumulative hazard* $\Psi(t) = -\log S(t)$ adds across independent states.

Proposition 3.2 (Allocation identity)**.** *Conditional on the world, with initialisations and trajectory randomness independent across states,* $J_{\mathrm{single}}(B) = F_0(B)$ *and* $J_{\mathrm{iso}}(B, M) = 1 - \prod_{i=1}^{M} S_i(B_i)$. *For symmetric states (identical discovery-time laws) and* M *dividing* B*, splitting is preferred iff* $M\Psi_{\mathrm{iso}}(B/M) > \Psi_0(B)$.

For symmetric states, splitting helps exactly when several short trajectories together accumulate more discovery hazard than one long trajectory. Each discovery-time law depends on the acquisition rule, prior, seeds and competition among exposed candidates, so pooled and isolated states can have different laws even under the same rule; the pooled state, for example, starts with the neighbourhoods of all K seeds exposed. The allocation identity also identifies when concentrating the budget is preferable. If pooled and isolated states share the same discovery-time law and the probability of discovery on the next test conditional on no discovery so far is non-decreasing, splitting cannot improve success (Prop. B.18). This increasing-failure-rate (IFR) condition captures trajectories for which continued search becomes more promising. Splitting can nevertheless help when the common law is non-IFR, as can occur with mixtures of commitments of different quality, or when isolation changes the discovery-time law. More generally, for a common law independent of M, the best split among M dividing B maximises $\Psi(t)/t$ with $t = B/M$, a fixed-budget analogue of the restart criterion of Luby et al. (1993) (Prop. B.25).

3.3 EPISTEMIC DIVERSITY AND THE BUDGET-DEPENDENT CROSSOVER

We first consider heterogeneous initial commitments, which can make continued investment in an unsuccessful trajectory less attractive than starting another. The seeds and prior of a state determine which directions it pursues, and some initialisations lead towards the target while others direct search into noise (Fig. 1E, bottom). To isolate this mechanism, consider a completion-time surrogate in which each initialisation, including the pooled one, is independently favourable with probability $\alpha \in (0, 1)$. A favourable state advances independently with probability $p \in (0, 1)$ per test and discovers the target after $Q \in \mathbb{N}$ advances, whereas an unfavourable state never discovers it. Write $F_{Q,p}(t) = \Pr\{\mathrm{Bin}(t, p) \ge Q\}$.

Proposition 3.3 (Commitment versus fragmentation)**.** *In this surrogate,* $J(B, M) = 1 - \prod_{i=1}^{M}[1 - \alpha F_{Q,p}(B_i)]$. *(i) If* $Q \le B$ *and* $\lceil B/M \rceil < Q$*, then* $J(B, M) = 0 < J(B, 1)$. *(ii) For every fixed* $M \ge 2$*, as* $B \to \infty$*,* $J(B, M) \uparrow 1 - (1 - \alpha)^M > \alpha \ge J(B, 1)$. *Consequently, every fixed* $M \ge 2$ *eventually overtakes* $M = 1$*, and* $M^\star(B) \to \infty$.

For $Q \ge 2$, small budgets favour the pooled state because every isolated state falls short of the Q advances (part (i)), whereas at large budgets a single state's success saturates at α and independent commitments hedge against an unfavourable draw. The optimal state count therefore diverges with the budget, with $M^\star(B) = \Theta(B)$ for fixed (α, Q, p) (Prop. B.26); cf. restarts and portfolios (Luby et al., 1993; Gomes et al., 1998; Huberman et al., 1997). The failure rate in finding the target within the finite bridge world converges to 0 with B in our exploring policies (Prop. B.27), but a strict advantage at a fixed budget persists when poor commitments advance sufficiently slowly instead of failing (Prop. B.22). The second departure, a difference between laws, yields the same crossover in a stagewise surrogate whenever isolated trajectories progress faster than the pooled trajectory through their slowest stage (Prop. B.20).

Prior diversity. To isolate diversity, fix $M \ge 2$ states with the same seeds, acquisition rule and local budget t, and let $S(\omega, t)$ be the probability, conditional on the world, that a state with prior $P(\omega)$ has not discovered $H^\star$ after t tests. Observation and acquisition randomness is independent across states, including states that share a prior draw.

Proposition 3.4 (Diversity gain)**.** *With* $S = S(\omega, t)$*, shared draws give* $J_M^{\mathrm{sh}} = 1 - \mathbb{E}_\nu[S^M]$ *and diverse draws give* $J_M^{\mathrm{div}} = 1 - (\mathbb{E}_\nu S)^M$*. By Jensen's inequality,* $J_M^{\mathrm{div}} \ge J_M^{\mathrm{sh}}$*, strictly unless* S *is almost surely constant. For* $M = 2$*, the gain is* $\mathrm{Var}_\nu S$.

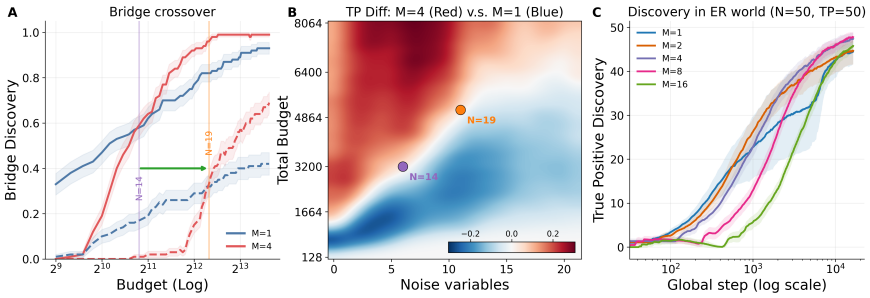


Figure 2: **Budget-dependent regime structure.** (**A**) Bridge recovery under softmax acquisition; vertical lines mark the first observed crossover on the tested budget grid. (**B**) Crossover regimes under LUCB-style acquisition as budget and the number of noise variables vary. (**C**) Multi-target recovery in a sparse random-factor world across different numbers of search states.

Let J_1 be the success probability of a single state with the same seeds, prior $P(\omega)$ with $\omega \sim \nu$, and budget Mt. Then $J_M^{\text{div}} - J_1 = (J_M^{\text{sh}} - J_1) + (J_M^{\text{div}} - J_M^{\text{sh}})$. The first term measures splitting under a shared prior draw and is non-positive when the discovery-time law given that draw is IFR (Cor. B.23). The second is the non-negative diversity gain of Prop. 3.4. This attribution assumes matched seeds, and with different seeds the diversity ordering needs further conditions (App. B.6.1).

4 Controlled Discovery Experiments

Next, we test the commitment–fragmentation tradeoff from Section 3 in controlled discovery problems with known ground truth. We first examine how the preferred search organization changes with budget and problem complexity, then study the mechanisms underlying these shifts, including frontier exposure and epistemic diversity.

4.1 Budget-dependent search regimes

Setup. We implement the Gaussian observation model and expanding hypothesis graph of Section 2. At fixed total budget B, we compare one pooled state with $M \in \{2, 4, 8, 16\}$ isolated states using softmax, an LUCB-style rule (adapting Kalyanakrishnan et al. (2012)), and Thompson sampling. The pooled state receives all known seeds and one prior; isolated states receive distributed seeds, independent priors from the same marginal distribution, and equal budget splits. Bridge experiments count recovery when at least one state's final posterior satisfies the acceptance criterion after the minimum observation count. Multi-target experiments count tested hypotheses whose observation-weighted aggregate posterior mean exceeds the truth threshold. These metrics differ from the cumulative objective in Eq. 4, so the crossovers below test the mechanism empirically rather than follow from the theory (App. B.8). Full implementation details are provided in App. E.

K-bridge world. We first simulate the K-bridge world with $K{=}4$ (Fig. 1D). At low budgets, pooled search recovers $H^\star$ more often, consistent with the fragmentation mechanism explored in Sec. 3.1. As the budget increases, $M{=}4$ overtakes $M{=}1$ (Fig. 2A), showing a crossover consistent with the mechanism illustrated by Prop. 3.3. Increasing the number of noise variables shifts this crossover towards larger budgets as the hypothesis universe is enlarged (Fig. 2B). The regime structure also depends on the acquisition rule. e.g. under Thompson sampling, the advantage appears only in an intermediate budget band (App. E.3).

Multi-target worlds. We next consider sparse random-factor worlds with $n{=}50$ variables, 16 known seeds, and 50 true hypotheses to discover. Recovery now involves multiple targets rather than a single bridge hypothesis. The preferred state count shifts across the tested budget range: $M{=}1$ leads at low budgets, $M{=}2$ at intermediate budgets, and $M{=}8$ at larger budgets before recovery converges (see Fig. 2C). Although this does not identify a global $M^\star(B)$, it shows that the budget-dependent regime structure extends beyond the single-target bridge construction. Additional worlds and acquisition policies are reported in Appendix E.6.

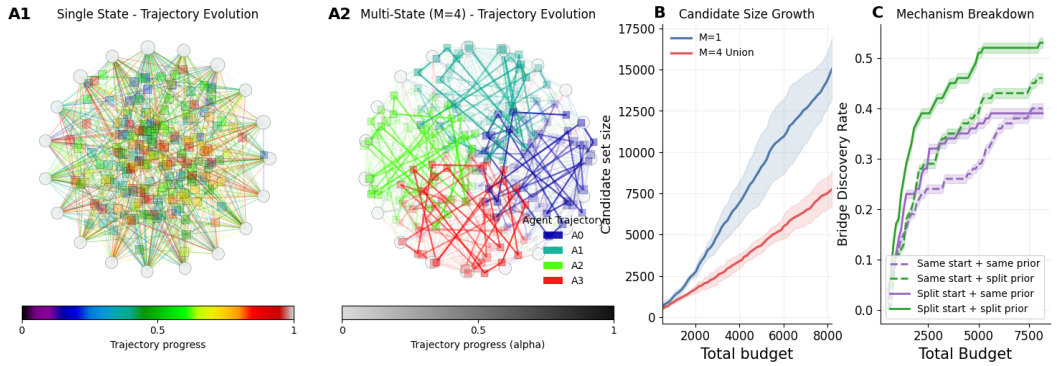


Figure 3: **Search diagnostics and initialisation ablations.** (**A**) Representative bridge-world search trajectories under matched total budget; positions are illustrative rather than a graph-distance embedding. (**B**) Growth of the exposed candidate set under pooled and distributed search. (**C**) Effects of seed and prior diversity on bridge recovery. Details are provided in Appendices E.4 and E.5.

4.2 Mechanisms of Search Organisation

Frontier exposure and trajectory specialisation. Search organization changes both which hypotheses become exposed and how queries are allocated. Representative trajectories show pooled search spreading across a broader region, while isolated states remain more locally specialized (see Fig. 3A). The pooled candidate set also grows faster than the union of the isolated sets in this comparison (see Fig. 3B), so each isolated state selects from a smaller local candidate set. Candidate-set size alone does not determine selection difficulty, which also depends on posterior scores and the acquisition rule; additional diagnostics are reported in Appendix E.4.

Seed and prior diversity. To isolate initialization effects, we cross shared versus distributed seeds with shared versus independent prior draws while holding state count, budget, acquisition rule, and prior strength fixed (see Fig. 3C). With identical seeds and local budgets, the shared-versus-independent prior comparison measures the diversity gain of Prop. 3.4. At $B{=}4096$, prior diversity increases mean bridge recovery by 0.10 with shared seeds and 0.11 with distributed seeds; distributing seeds yields gains of 0.09 and 0.10, respectively (see App. E.5). The distributed-search advantage is therefore not explained solely by different starting locations: diverse priors improve recovery even when all states begin from the same seeds.

5 Search Organisation in Scientific Discovery Settings

Having established these regimes in controlled settings, we turn to scientific discovery tasks. We use two benchmarks that probe complementary aspects of search organisation: SciGym tests how pooled and distributed search behave as the experimental budget increases, while CausalBench examines whether trajectory-dependent frontier expansion can improve discovery within a large, fixed hypothesis space. Unlike the Gaussian hypothesis-graph model above, both settings define search states and information frontiers through the structure of the underlying scientific task.

5.1 Budget-Dependent Experimental Search

SciGym (Duan et al., 2025) is a systems-biology simulator in which agents design sequential experiments to recover hidden reaction networks. We retain this experimentation loop but replace the native free-text interface with structural search, enabling controlled comparison of search organisation across budgets.

Setup. Each epistemic state maintains a beam of candidate reaction networks, selects experiments using its current candidates, and expands the beam through one-edit structural neighbourhoods. We compare one pooled beam initialised with four candidates ($M{=}1$) against four isolated beams initialised with one candidate each ($M{=}4$), matching aggregate initial information, total beam capacity, experimental budget, and output capacity. Our representation applies to 90 of 137 systems, which

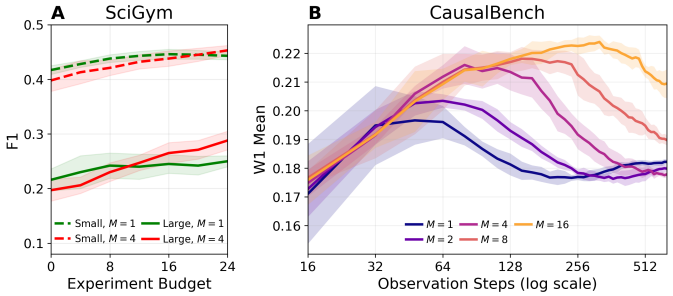


Figure 4: **Budget-dependent recovery in scientific discovery tasks.** (**A**) Reaction-network recovery under pooled and distributed search across experimental budgets in SciGym. (**B**) Gene regulatory network recovery across different numbers of persistent search states and update budgets in CausalBench.

Method	**W_1 ↑**	**BioPrec ↑**
GRNBoost	0.164	0.276
DCDI-DSF	0.164	0.252
DCDI-G	0.170	0.296
ISLAND (Ours)		
M=1	0.182	0.424
M=16	**0.209**	**0.550**

Table 1: **K562 gene regulatory network recovery.** Methods are evaluated on the top 1,000 predictions; W_1 denotes mean Wasserstein score and BioPrec biological precision. Full results table in Appendix D.3.

we stratify by the number of actionable species. We evaluate directed reactant–product recovery using F_1 over ten seeds and budgets up to B=24. Full details are provided in Appendix C.

Budget-dependent recovery. Pooled search performs better at small budgets, while isolated searches continue improving and overtake at larger budgets (Fig. 4A). At B=24, the mean advantage of distributed search reaches +3.73 F_1 points in larger systems and +1.03 points in smaller systems. The later crossover in smaller systems is consistent with concentrated search remaining sufficient over a wider budget range, although task structure and acquisition policy may also contribute. Because trajectory access, acquisition, and aggregation vary jointly, this experiment evaluates the overall search organisation rather than isolation alone.

LLM agents. In SciGym's native LLM-agent loop at matched interaction budget, a same-context pair does not improve on a single agent, whereas a different-context pair does, mirroring the decomposition after Prop. 3.4. However, this is mostly exploratory in nature, as the context change also alters the available information, so this does not strictly isolate prior diversity (App. C.3).

5.2 Trajectory-dependent frontier expansion for gene network recovery

We evaluate persistent, trajectory-dependent search in gene regulatory network (GRN) recovery using CausalBench (Chevalley et al., 2025). Here, the global hypothesis space is fixed, while each state explores a candidate frontier shaped by its own search history. CausalBench comprises two sparse, large-scale datasets, K562 and RPE1.

Expanding local gene subspaces. Starting from a small initial seed set, each state estimates conditional dependencies within its current candidate genes and randomly sampled outside scouts. Genes with strong cross-boundary evidence are promoted into the candidate set, making both gene exposure and conditioning sets trajectory-dependent. States remain separate during search, with edge scores combined only at evaluation. At a fixed total budget, M therefore also controls search locality: a single state expands towards the full gene set, whereas multiple states maintain smaller, evidence-driven neighbourhoods. On K562 (622 genes), we divide 32 seed genes across $M \in \{1, 2, 4, 8, 16\}$ states, allocate up to 640 updates equally, and evaluate the top 1,000 edges using mean Wasserstein score (W_1) and biological precision (BioPrec). As each update bootstraps the same observational matrix, the update budget here measures sampling budget rather than new experiments; this setting tests the locality mechanism rather than experimental allocation. Implementation details in App. D.2.

Observation-only recovery. The preferred organisation again depends on budget: M=1 leads at the smallest budgets and M=16 at the largest (Fig. 4B). At 640 updates, M=16 reaches W_1 0.209 and BioPrec 0.550, compared with 0.182 and 0.424 for the single-state variant. However, the gain is not monotone in M, at 640 updates, M=2 and M=4 fall below M=1 (Table 7). Across 20 paired runs, the gains are +0.027 in W_1 (95% CI $[0.024, 0.029]$) and +0.126 in BioPrec ($[0.112, 0.140]$), with both metrics improving in every run, and exceeding the reported GRNBoost and DCDI baselines on K562. The multi-state endpoint advantage replicates on RPE1 (App. D.4).

Search locality and frontier expansion. Exposing all 622 genes to a single state from initialization gives a similar endpoint (W_1 0.182; BioPrec 0.426), suggesting that restricted initial access alone does not explain the multi-state advantage. One possible mechanism is that persistent local trajectories preserve useful conditioning subspaces, whereas continued frontier growth moves each state towards a global estimator. The budget trajectories in Fig. 4B are consistent with this interpretation, with larger M delaying the loss of locality. Random subspace sampling with 16 genes per step yields a smaller gain from increasing M (W_1 $0.171 \to 0.175$). These results support persistent, evidence-driven local search as a promising discovery organisation (App. D.5).

Intervention selection. We also extend `ISLAND` to choosing which genes to perturb. Starting from 160 total observational updates, `ISLAND` selects 80 unique source genes in five batches of 16. All methods receive the same intervention budget, and are evaluated through a common selected-source proxy constructed from perturbation responses, comparing the evidence recovered by their selected sources rather than their native network outputs. `ISLAND` achieves higher mean endpoint W_1 and BioPrec than other baselines. Empirically, within `ISLAND`, $M{=}4$ initially trails $M{=}1$ in W_1 but overtakes it at later batches, suggesting modest benefit in distributed discovery (App. D.6).

6 Discussion

Our results suggest that in expanding-frontier discovery, the organisation of belief states is part of the search problem rather than an implementation detail. A pooled state can devote the full budget to resolving prerequisites but remains sensitive to a poor initialization, whereas isolated states hedge against such commitments while fragmenting the budget. Across theory, controlled simulations, and two biological benchmarks, this tradeoff often favours concentrated search at small budgets and distributed search at larger ones, although the ordering depends on the acquisition rule.

How many scientists? If each scientist is a persistent epistemic state, our analysis gives a conditional answer: concentration is necessary when individual trajectories cannot clear their prerequisites, while additional states can help when heterogeneous commitments or improved local search efficiency offset fragmentation (Prop. 3.1, 3.3, B.20). Whether splitting can help at all depends on the discovery-time law: when isolation leaves that law unchanged, an increasing hazard makes pooling optimal (Prop. B.18), whereas a mixture of well- and poorly-directed commitments creates the regime in which restarts and portfolios pay off (Luby et al., 1993; Gomes et al., 1998; Huberman et al., 1997). When splitting helps and the local law does not depend on M, the natural design variable is the budget per state: the best local budget $t^\star$ maximises $\Psi(t)/t$ (Prop. B.25), so $M^\star(B) \approx B/t^\star$, growing in proportion to the budget in our surrogate (Prop. B.26). The crossover is then an explicit function of quantities a practitioner can reason about: prerequisite depth, the advance rate of a well-directed search, and the fraction of well-directed starting commitments (Props. 3.3 and B.20). The same tradeoff between many short and few long searches governs how island-model evolutionary algorithms size their subpopulations (Whitley et al., 1999; Cantú-Paz, 2000), and how successive halving and Hyperband divide a fixed budget (Jamieson & Talwalkar, 2016; Li et al., 2018).

Limitations. Our theory identifies what determines $M^\star$ but does not estimate it for a new problem: depth, advance rates, and the fraction of favourable commitments are known in constructed worlds but must be inferred elsewhere, and the plug-in rule of App. B.7 needs pilot estimates of discovery hazards, which are hard to obtain when discoveries cannot be verified directly. In addition, our evaluations are retrospective; SciGym varies several components of search organisation jointly, while CausalBench uses a method-induced frontier. Testing distributed search under real scientific discovery systems will also include confounds such as costs, delays, and physical limits.

Implications and future directions. Choosing M online, by starting many states and retiring unpromising ones, could avoid the pilot hazard estimates our plug-in rule requires (App. B.7). Partial isolation is another direction: our states never communicate, yet diversity in scientific communities helps most while transient, and densely connected groups converge quickly at the cost of long-run performance (Zollman, 2010; Lazer & Friedman, 2007). As autonomous systems take on more of the experimental loop, how many lines of inquiry to sustain becomes an explicit design choice, shaped by factors such as prerequisite depth and diversity in search processes in addition to budget.

REFERENCES

Britt Adamson, Thomas M. Norman, Marco Jost, Min Y. Cho, James K. Nuñez, Yuwen Chen, Jacqueline E. Villalta, Luke A. Gilbert, Max A. Horlbeck, Marco Y. Hein, Ryan A. Pak, Andrew N. Gray, Carol A. Gross, Atray Dixit, Oren Parnas, Aviv Regev, and Jonathan S. Weissman. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell*, 167(7):1867–1882.e21, 2016. doi: 10.1016/j.cell.2016.11.048.

Daniil A. Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578, 2023. doi: 10.1038/s41586-023-06792-0.

Philippe Brouillard, Sébastien Lachapelle, Alexandre Lacoste, Simon Lacoste-Julien, and Alexandre Drouin. Differentiable causal discovery from interventional data. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 21865–21877, 2020.

Benjamin Burger, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M. Alston, Buyi Li, Rob Clowes, Nicola Rankin, Brandon Harris, Reiner Sebastian Sprick, and Andrew I. Cooper. A mobile robotic chemist. *Nature*, 583(7815):237–241, 2020. doi: 10.1038/s41586-020-2442-2.

Erick Cantú-Paz. *Efficient and Accurate Parallel Genetic Algorithms*. Kluwer Academic Publishers, 2000.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995. doi: 10.1214/ss/1177009939.

Mathieu Chevalley, Yusuf H. Roohani, Arash Mehrjou, Jure Leskovec, and Patrick Schwab. A large-scale benchmark for network inference from single-cell perturbation data. *Communications Biology*, 8(1), March 2025. ISSN 2399-3642. doi: 10.1038/s42003-025-07764-y. URL http://dx.doi.org/10.1038/s42003-025-07764-y.

Cristina Cornelio, Takuya Ito, Ryan Cory-Wright, Sanjeeb Dash, and Lior Horesh. The need for verification in artificial intelligence-driven scientific discovery. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 384(2317), April 2026. ISSN 1471-2962. doi: 10.1098/rsta.2024.0591. URL http://dx.doi.org/10.1098/rsta.2024.0591.

Thomas Desautels, Andreas Krause, and Joel W. Burdick. Parallelizing exploration-exploitation tradeoffs in Gaussian process bandit optimization. *Journal of Machine Learning Research*, 15 (119):4053–4103, 2014.

Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P. Fulco, Livnat Jerby-Arnon, Nemanja D. Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, Britt Adamson, Thomas M. Norman, Eric S. Lander, Jonathan S. Weissman, Nir Friedman, and Aviv Regev. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*, 167(7):1853–1866.e17, 2016. doi: 10.1016/j.cell.2016.11.038.

Haonan Duan, Stephen Zhewen Lu, Caitlin Fiona Harrigan, Nishkrit Desai, Jiarui Lu, Michał Koziarski, Leonardo Cotta, and Chris J. Maddison. Measuring scientific capabilities of language models with a systems biology dry lab, 2025. URL https://arxiv.org/abs/2507.02083.

J. D. Esary, F. Proschan, and D. W. Walkup. Association of random variables, with applications. *The Annals of Mathematical Statistics*, 38(5):1466–1474, 1967.

Yoav Freund, H. Sebastian Seung, Eli Shamir, and Naftali Tishby. Selective sampling using the query by committee algorithm. *Machine Learning*, 28(2–3):133–168, 1997. doi: 10.1023/A:1007330508534.

Nathan C. Frey, Isidro Hötzel, Samuel D. Stanton, Ryan Kelly, Robert G. Alberstein, Emily K. Makowski, Karolis Martinkus, Daniel Berenberg, Jack Bevers, Tyler Bryson, Pamela Chan, Yongmei Chen, Alicja Czubaty, Tamica D'Souza, Henri Dwyer, Anna Dziewulska, James W. Fairman, Allen Goodman, Jennifer Hofmann, Henry Isaacson, Aya Ismail, Samantha James, Taylor Joren, Simon Kelow, James R. Kiefer, Matthieu Kirchmeyer, Joseph Kleinhenz, James T. Koerber, Julien Lafrance-Vanasse, Andrew Leaver-Fay, Jae Hyeon Lee, Edith Lee, Donald Lee, Wei-Ching Liang, Joshua Yao-Yu Lin, Sidney Lisanza, Andreas Loukas, Jan Ludwiczak, Sai Pooja Mahajan, Omar Mahmood, Homa Mohammadi-Peyhani, Santrupti Nerli, Ji Won Park, Jaewoo Park, Stephen Ra, Sarah Robinson, Saeed Saremi, Franziska Seeger, Imee Sinha, Anna M. Sokol, Natasa Tagasovska, Hao To, Edward Wagstaff, Amy Wang, Andrew M. Watkins, Blair Wilson, Shuang Wu, Karina Zadorozhny, John Marioni, Aviv Regev, Yan Wu, Kyunghyun Cho, Richard Bonneau, and Vladimir Gligorijević. Lab-in-the-loop therapeutic antibody design with deep learning. February 2025. doi: 10.1101/2025.02.19.639050. URL http://dx.doi.org/10.1101/2025.02.19.639050.

Ali E. Ghareeb, Benjamin Chang, Ludovico Mitchener, Angela Yiu, Caralyn J. Szostkiewicz, Dmytro Shved, Gavin J. Gyimesi, Jon M. Laurent, Samantha M. Wright, Muhammed T. Razzak, Andrew D. White, Silvia C. Finnemann, Michaela M. Hinks, and Samuel G. Rodriques. A multi-agent system for automating scientific discovery. *Nature*, 655(8122):497–505, May 2026. ISSN 1476-4687. doi: 10.1038/s41586-026-10652-y. URL http://dx.doi.org/10.1038/s41586-026-10652-y.

Carla P. Gomes, Bart Selman, and Henry Kautz. Boosting combinatorial search through randomization. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI)*, pp. 431–437, 1998.

Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Petar Sirkovic, Artiom Myaskovsky, Grzegorz Glowaty, Felix Weissenberger, Alessio Orlandi, Dan Popovici, Anil Palepu, Keran Rong, Ryutaro Tanno, Khaled Saab, Fan Zhang, Jacob Blum, Andrew Carroll, Kavita Kulkarni, Nenad Tomašev, Dina Zverinski, Ivor Rendulic, Elahe Vedadi, Florian Hasler, Luka Rimanic, Marina Boia, Ivan Budiselic, Ben Feinstein, Mathias Bellaiche, Tom Sheffer, Jan Freyberg, Jeremy Ratcliff, Ottavia Bertolli, Katherine Chou, Avinatan Hassidim, Burak Gokturk, Amin Vahdat, Yuan Guan, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R. D. Costa, José R. Penadés, Gary Peltz, Yossi Matias, James Manyika, Demis Hassabis, Yunhan Xu, Pushmeet Kohli, Annalisa Pawlosky, Alan Karthikesalingam, and Vivek Natarajan. Accelerating scientific discovery with co-scientist. *Nature*, 655(8122):487–496, May 2026. ISSN 1476-4687. doi: 10.1038/s41586-026-10644-y. URL http://dx.doi.org/10.1038/s41586-026-10644-y.

Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4):926–939, 2014. doi: 10.1016/j.ijar.2013.11.007.

S. G. Hoggar. Chromatic polynomials and logarithmic concavity. *Journal of Combinatorial Theory, Series B*, 16(3):248–254, 1974.

Lu Hong and Scott E. Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004. doi: 10.1073/pnas.0403723101.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning, 2011. URL https://arxiv.org/abs/1112.5745.

Kexin Huang, Romain Lopez, Jan-Christian Hütter, Takamasa Kudo, Antonio Rios, and Aviv Regev. Sequential optimal experimental design of perturbation screens guided by multi-modal priors. In *Research in Computational Molecular Biology (RECOMB)*. Springer, 2024. doi: 10.1007/978-1-0716-3989-4_2.

Bernardo A. Huberman, Rajan M. Lukose, and Tad Hogg. An economics approach to hard computational problems. *Science*, 275(5296):51–54, 1997.

T. Jesper Jacobsson. AI-generated hypotheses and the emergence of autonomous scientific discovery. *ACS Materials Letters*, 8(6):1457–1464, May 2026. ISSN 2639-4979. doi: 10.1021/

acsmaterialslett.6c00224. URL http://dx.doi.org/10.1021/acsmaterialslett.6c00224.

Kevin Jamieson and Ameet Talwalkar. Non-stochastic best arm identification and hyperparameter optimization. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 51 of *PMLR*, pp. 240–248, 2016.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, pp. 655–662, 2012.

Kirthevasan Kandasamy, Akshay Krishnamurthy, Jeff Schneider, and Barnabas Poczos. Parallelised bayesian optimisation via thompson sampling. In Amos Storkey and Fernando Perez-Cruz (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 133–142. PMLR, 09–11 Apr 2018. URL https://proceedings.mlr.press/v84/kandasamy18a.html.

Ross D. King, Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Whelan, and Amanda Clare. The automation of science. *Science*, 324(5923):85–89, 2009. doi: 10.1126/science.1165620.

Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. BatchBALD: Efficient and diverse batch acquisition for deep Bayesian active learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Philip Kitcher. The division of cognitive labor. *The Journal of Philosophy*, 87(1):5–22, 1990. doi: 10.2307/2026796.

Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision making in multi-agent multi-armed bandits. *Automatica*, 125:109445, 2021. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2020.109445. URL https://www.sciencedirect.com/science/article/pii/S0005109820306439.

David Lazer and Allan Friedman. The network structure of exploration and exploitation. *Administrative Science Quarterly*, 52(4):667–694, 2007.

Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185):1–52, 2018.

Chris Lu, Cong Lu, Robert Tjarko Lange, Yutaro Yamada, Shengran Hu, Jakob Foerster, David Ha, and Jeff Clune. Towards end-to-end automation of ai research. *Nature*, 651(8107):914–919, March 2026. ISSN 1476-4687. doi: 10.1038/s41586-026-10265-5. URL http://dx.doi.org/10.1038/s41586-026-10265-5.

Michael Luby, Alistair Sinclair, and David Zuckerman. Optimal speedup of Las Vegas algorithms. *Information Processing Letters*, 47(4):173–180, 1993.

Clare Lyle, Arash Mehrjou, Pascal Notin, Andrew Jesson, Stefan Bauer, Yarin Gal, and Patrick Schwab. DiscoBAX: Discovery of optimal intervention sets in genomic experiment design. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23170–23189. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/lyle23a.html.

David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Arash Mehrjou, Ashkan Soleymani, Andrew Jesson, Pascal Notin, Yarin Gal, Stefan Bauer, and Patrick Schwab. GeneDisco: A benchmark for experimental design in drug discovery. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=-w2oomO6qgc.

Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics (Oxford, England)*, 35(12):2159–2161, 2019. doi: 10.1093/bioinformatics/bty916.

Ian Osband, John Aslanides, and Albin Cassirer. Randomized prior functions for deep reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, pp. 8626–8638, 2018.

Tom Rainforth, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1), February 2024. ISSN 0883-4237. doi: 10.1214/23-sts915. URL http://dx.doi.org/10.1214/23-STS915.

Joseph M. Replogle, Reuben A. Saunders, Angela N. Pogson, Jeffrey A. Hussmann, Alexander Lenail, Alina Guna, Lauren Mascibroda, Eric J. Wagner, Karen Adelman, Gila Lithwick-Yanai, Nika Iremadze, Florian Oberstrass, Doron Lipson, Jessica L. Bonnar, Marco Jost, Thomas M. Norman, and Jonathan S. Weissman. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*, 185(14):2559–2575.e28, 2022. doi: 10.1016/j.cell.2022.05.013.

Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. A tutorial on thompson sampling. *Foundations and Trends in Machine Learning*, 11(1):1–96, 2018. doi: 10.1561/2200000070.

Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

H. Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory (COLT)*, pp. 287–294. ACM, 1992. doi: 10.1145/130385.130417.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 1015–1022, 2010.

Kyle Swanson, Wesley Wu, Nash L. Bulaong, John E. Pak, and James Zou. The virtual lab of ai agents designs new sars-cov-2 nanobodies. *Nature*, 646(8085):716–723, July 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09442-9. URL http://dx.doi.org/10.1038/s41586-025-09442-9.

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson, Yan Zeng, and Gerbrand Ceder. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature*, 624(7990):86–91, November 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06734-w. URL http://dx.doi.org/10.1038/s41586-023-06734-w.

Panagiotis Tigas, Yashas Annadani, Andrew Jesson, Bernhard Schölkopf, Yarin Gal, and Stefan Bauer. Interventions, where and how? Experimental design for causal models at scale. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 24130–24143, 2022.

Simon Tong and Daphne Koller. Active learning for structure in Bayesian networks. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 863–869, 2001.

Michael Weisberg and Ryan Muldoon. Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2):225–252, 2009. doi: 10.1086/644786.

Darrell Whitley, Soraya Rana, and Robert B. Heckendorn. The island model genetic algorithm: On separability, population size and convergence. *Journal of Computing and Information Technology*, 7(1):33–47, 1999.

David Williams. *Probability with Martingales*. Cambridge University Press, 1991.

Kevin J. S. Zollman. The communication structure of epistemic communities. *Philosophy of Science*, 74(5):574–587, 2007. doi: 10.1086/525605.

Kevin J. S. Zollman. The epistemic benefit of transient diversity. *Erkenntnis. An International Journal of Scientific Philosophy*, 72(1):17–35, 2010. doi: 10.1007/s10670-009-9194-6.

SUPPLEMENTARY INFORMATION

Contents

A Related Work **17**

B Formal Analysis and Proofs **19**

B.1 Formal Setup, Probability Space, and Restart Equivalence 19

B.2 Bridge Universe, Access Depth, and Exposure 22

B.3 Gaussian Beliefs and Resolution 23

B.4 Policy-Facing Competition 24

B.5 Completion-Time Structure 27

B.6 Proofs for Section 3 30

B.7 Allocation, Restarts, and Convergence 33

B.8 Final-Posterior Recovery and Its Relation to the Theory 34

C Additional Experimental Details: SciGym **35**

C.1 Structural-search protocol 35

C.2 Budget-dependent structural recovery 35

C.3 Native LLM-agent extension 36

D Additional Experimental Details: CausalBench **36**

D.1 Datasets and evaluation 36

D.2 Observation-only search protocol 37

D.3 K562 observation-only results 37

D.4 RPE1 replication 38

D.5 Search-space controls 38

D.6 Intervention selection and network evaluation 39

E Synthetic Simulations: Details and Extended Results **41**

E.1 Simulator and experimental protocol 41

E.2 Evaluation and statistical reporting 44

E.3 Bridge experiments and policy robustness 46
E.4 Trajectory and candidate-set diagnostics 46
E.5 Seed allocation and prior diversity 46
E.6 Larger sparse worlds 47

F AI use statement **48**

A Related Work

Bayesian active learning, experimental design, and batch selection. Active learning and Bayesian experimental design provide principled frameworks for selecting informative queries or experiments under limited budgets (Settles, 2009; Chaloner & Verdinelli, 1995; Rainforth et al., 2024). Information-theoretic approaches such as BALD select observations that are expected to reduce uncertainty about model parameters or predictions (Houlsby et al., 2011), while query-by-committee methods use disagreement among candidate models to guide acquisition (Seung et al., 1992; Freund et al., 1997). Bayesian optimisation similarly balances information gathering and exploitation when evaluations are costly (Srinivas et al., 2010). Batch and parallel methods extend these ideas to several simultaneous evaluations, explicitly encouraging non-redundant acquisitions or accounting for pending experiments (Kirsch et al., 2019; Desautels et al., 2014; Kandasamy et al., 2018). These methods are highly relevant to scientific discovery, but generally optimize experiment selection from a common posterior over a fixed or externally specified candidate space. Our work instead asks how the discovery process itself should be organised across persistent belief states. In expanding-frontier discovery, different histories can expose different future hypotheses, so diversity can alter the evolution of the search space rather than only the composition of the next acquisition batch.

Causal discovery and biological experimental design. Our biological experiments connect to active causal discovery and experimental design for gene regulatory networks. Early work studied intervention selection for learning Bayesian-network structure (Tong & Koller, 2001; Hauser & Bühlmann, 2014), while more recent Bayesian approaches support experimental design in larger and nonlinear structural causal models (Tigas et al., 2022). In functional genomics, methods such as IterPert use sequential experimental design to allocate limited perturbation-screen budgets (Huang et al., 2024), and DiscoBAX explicitly seeks diverse, high-value intervention sets (Lyle et al., 2023). GeneDisco provides a benchmark for active experimental design over genetic perturbations (Mehrjou et al., 2022), while CausalBench evaluates gene-network inference using large-scale single-cell perturbation data (Chevalley et al., 2025). These approaches primarily address which intervention or experiment should be selected given a model and candidate universe. Our contribution is complementary: we study how adaptive search should be distributed across one or several persistent epistemic states.

Multi-agent and distributed exploration. Distributed bandits and cooperative exploration study how multiple learners allocate actions, exchange observations, and reduce regret or sample complexity (Landgren et al., 2021; Martínez-Rubio et al., 2019). Related work on parallel Bayesian optimisation distributes costly evaluations across multiple workers (Desautels et al., 2014; Kandasamy et al., 2018), while randomized priors and posterior sampling provide mechanisms for maintaining temporally coherent exploratory behaviour (Osband et al., 2018; Russo et al., 2018). These settings typically retain a common action space and focus on coordination or efficient execution. Our setting instead makes the accessible hypothesis set trajectory dependent: resolving one hypothesis can expose new hypotheses for subsequent testing.

Epistemic diversity and the division of cognitive labour. A closely related literature in philosophy of science studies how scientific communities should distribute effort across competing research programmes. Formal models of the division of cognitive labour argue that maintaining heterogeneous approaches can improve collective discovery (Kitcher, 1990; Hong & Page, 2004), while epistemic-landscape models study how populations of scientists explore different regions of an initially uncertain research space (Weisberg & Muldoon, 2009). Communication itself can create a tradeoff between rapid convergence and collective reliability (Zollman, 2007), and transient diversity can preserve exploration before eventual consensus (Zollman, 2010). Our work shares the premise that persistent epistemic diversity can be useful, but places it in a finite-budget sequential decision problem with explicit beliefs, experimental observations, and trajectory-dependent hypothesis exposure. This allows us to study how the benefit of diversity interacts quantitatively with the loss of search depth caused by splitting a fixed experimental budget.

Restart and portfolio search. Our allocation analysis is also related to restart strategies and algorithm portfolios. Randomized search procedures can benefit from multiple independent runs when

Table 2: Comparison with related work. Many related areas study limited-budget decision making or diversity in search, but typically do not jointly model persistent epistemic states and trajectory-dependent expansion of the hypothesis frontier.

Literature	Limited budget	Expanding frontier	Design objective
Bayesian active learning / BOED / batch design	✓	×	Select informative experiments from a common candidate space
Causal discovery / biological experimental design	✓	×	Select interventions to identify causal structure or biological targets
Multi-agent bandits / distributed exploration	✓	×	Coordinate exploration to reduce regret or sample complexity
Epistemic diversity / cognitive labour	∼	∼	Distribute scientific effort across competing approaches
Restart / portfolio search	✓	×	Allocate computation across independent search trajectories
Autonomous scientific discovery / AI scientists	∼	∼	Automate hypothesis generation, experimentation, and scientific reasoning
This work	✓	✓	Choose how many **persistent search states** to maintain under a **commitment–fragmentation tradeoff**

✓: explicitly modelled; ×: typically not explicit; ∼: sometimes present or implicit.

completion times are heterogeneous or heavy tailed, producing optimal restart schedules and portfolio effects (Luby et al., 1993; Gomes et al., 1998; Huberman et al., 1997). This connection appears explicitly in our discovery-time analysis: several shorter trajectories can outperform one long trajectory when their aggregate discovery hazard is larger. Expanding-frontier discovery adds an additional mechanism, because different persistent states can maintain distinct priors, exposed hypothesis sets, and scientific commitments. Island models in evolutionary computation similarly partition a population into subpopulations with limited migration, and their analysis of subpopulation number, size, and migration rate (Whitley et al., 1999; Cantú-Paz, 2000) parallels our question of how many isolated states a budget supports; our setting differs in that each state's accessible hypothesis set is itself trajectory-dependent.

Autonomous scientific discovery and scientific agents. Automated and AI-assisted scientific discovery has progressed from early robot-scientist systems (King et al., 2009) to autonomous chemistry and materials laboratories (Burger et al., 2020; Szymanski et al., 2023; Boiko et al., 2023). More recent systems use language models and multi-agent architectures for hypothesis generation, experiment planning, scientific reasoning, and iterative discovery (Lu et al., 2026; Gottweis et al., 2026; Duan et al., 2025; Swanson et al., 2025; Ghareeb et al., 2026). Laboratory-in-the-loop systems further demonstrate increasingly tight integration between learned models and physical experimentation (Frey et al., 2025). These works establish the growing feasibility of maintaining multiple computational search processes over successive scientific experiments. Our contribution addresses a complementary organisational question: under a fixed experimental budget, when should such systems pool their evidence into one evolving search state, and when should they preserve several isolated states?

Distinction. Table 2 summarizes the main distinctions between our setting and related approaches. Across these literatures, diversity enters through acquisition, ensembles, distributed execution, heterogeneous research programmes, or independent restarts. We instead study the number and isolation of persistent belief states in a setting where previous discoveries determine which hypotheses become available next. Concentrating the budget supports deeper resolution along a trajectory, whereas distributing it preserves distinct commitments and frontier coverage. Our main contribution is to formalise this commitment–fragmentation tradeoff, characterise its budget-dependent con-

sequences, and test the resulting regime structure across controlled worlds and scientific discovery benchmarks.

B Formal Analysis and Proofs

This appendix states the assumptions behind Sections 2 and 3 and proves the results used there. Table 3 summarises the notation. Section B.1 gives the complete formal model, including the conventions that Section 2 only summarises, and fixes the probability space and comparison conventions. Sections B.2–B.4 analyse the Gaussian bridge search and its acquisition rules. Section B.5 introduces separate completion-time surrogates. The remaining sections establish allocation and diversity results and distinguish cumulative acceptance from final-posterior recovery.

B.1 Formal Setup, Probability Space, and Restart Equivalence

Hypothesis universe and graph. The hypothesis set $\mathcal{H}$ is finite, and every scope satisfies $|\mathcal{S}_h| \geq 2$. Finiteness bounds the scope sizes; in the K-bridge world the maximum arity is K. By equation 1, $d_{\mathrm{edit}}(\mathcal{S}, \mathcal{T}) = 1$ iff $(|\mathcal{S} \setminus \mathcal{T}|, |\mathcal{T} \setminus \mathcal{S}|) \in \{(1,0),(0,1),(1,1)\}$, that is, iff $\mathcal{T}$ is obtained from $\mathcal{S}$ by one deletion, one addition, or one replacement. The relation is symmetric, so $(\mathcal{H}, \mathcal{E})$ is an undirected graph. For $\mathcal{A} \subseteq \mathcal{H}$, $\Gamma(\mathcal{A}) = \bigcup_{h \in \mathcal{A}} \Gamma(h)$, with $\Gamma(\varnothing) = \varnothing$.

World and initialisation. A world $\mathcal{W}$ fixes the finite hypothesis graph, latent means $(\mu_h)_{h \in \mathcal{H}}$ with $\mu_h \neq \tau$, and truths $Z_h = \mathbf{1}\{\mu_h > \tau\}$. An initialisation $\mathcal{I}$ fixes each state's seed set $\mathcal{R}^+_{i,0}$ and prior parameter ω_i. Under the prior $P_i = P(\omega_i)$, the latent means are independent across hypotheses with $\mu_h \sim \mathcal{N}(\mu_{0,i}(h), \sigma^2/\kappa_{0,i}(h))$, where $\mu_{0,i}$ and $\kappa_{0,i}$ are determined by ω_i. Assume $\sigma^2 > 0$ and $\kappa_{0,i}(h) > 0$. Unless an average over initialisations is stated explicitly, probabilities and expectations below are conditional on $(\mathcal{W}, \mathcal{I})$. The posterior probabilities $\Pr_{P_i}$ are subjective probabilities under the state's prior, whereas repeated-experiment probabilities condition on the fixed world.

Histories and acquisition. Let $a_{i,s}$ be the hypothesis selected by state i after s completed tests and $Y_{i,s+1}$ the resulting observation. The history after t tests is

$$\mathcal{F}_{i,t} = \big(\mathcal{R}^+_{i,0},\, \omega_i,\, a_{i,0}, Y_{i,1}, \ldots, a_{i,t-1}, Y_{i,t}\big),$$

and we use the same symbol for the σ-field it generates. The count $n_{i,t}(h) = \sum_{s<t} \mathbf{1}\{a_{i,s} = h\}$, the posterior quantities of Section B.3, the resolved sets, and the candidate set $\mathcal{C}_{i,t}$ are all functions of $\mathcal{F}_{i,t}$. An acquisition rule maps $(P_i, \mathcal{F}_{i,t}, \mathcal{C}_{i,t})$ to a distribution on $\mathcal{C}_{i,t}$. The rule and its parameters are held fixed within each architectural comparison and do not depend on the total budget B or the local budget B_i. Consequently, the law of a state's first t steps is the same for every local budget $B_i \geq t$. This property is used in Proposition 3.2, where a local success probability $F_i(t)$ must not depend on B_i, and in Corollary B.23, where a single state with budget Mt and an isolated state with budget t must follow the same law up to time t.

Stack construction. For each state i and hypothesis h, let $(\xi_{i,h,k})_{k \geq 1}$ be i.i.d. $\mathcal{N}(\mu_h, \sigma^2)$. Let $(U_{i,t})_{t \geq 0}$ be i.i.d. uniform on $[0,1]$, with all stacks and randomisation variables mutually independent. After t completed tests, the rule selects $a_{i,t}$ using $\mathcal{F}_{i,t}$ and $U_{i,t}$, for example by inverting the cumulative selection probabilities over a fixed ordering of $\mathcal{C}_{i,t}$. The k-th test of h returns $\xi_{i,h,k}$. This construction realises the observation model for every history-adapted acquisition rule. Conditional on $(\mathcal{W}, \mathcal{I})$, distinct states have independent trajectories because they use disjoint independent sources. The observations of a hypothesis are successive entries of its i.i.d. Gaussian stack. Although the number revealed may depend on previous outcomes, the strong law applies along any trajectory on which infinitely many entries are revealed.

Resolution and retained access. With thresholds $0 < \eta_- < \frac{1}{2} < \eta_+ < 1$ and $n_{\min} \geq 1$, the cumulative accepted and rejected sets are

$$\mathcal{R}^+_{i,t} = \mathcal{R}^+_{i,0} \cup \big\{h : \exists s \leq t,\ n_{i,s}(h) \geq n_{\min},\ b_{i,s}(h) \geq \eta_+\big\},$$
$$\mathcal{R}^-_{i,t} = \big\{h : \exists s \leq t,\ n_{i,s}(h) \geq n_{\min},\ b_{i,s}(h) \leq \eta_-\big\},$$

Table 3: Notation used in the formulation and analysis.

Symbol	Meaning
Hypotheses and observations	
$\mathcal{V}$, n	Scientific variables and their number.
$\mathcal{H}$, $\mathcal{S}_h$	Finite hypothesis universe and scope of h.
Z_h, μ_h, τ, σ^2	Truth value, latent mean, truth threshold, and observation variance.
$\mathcal{E}$, $\Gamma(h)$, d_{edit}	Hypothesis graph, neighbourhood of h, and one-edit distance.
Epistemic states and beliefs	
$\mathcal{W}$, $\mathcal{I}$	Fixed world and collection of state initialisations.
$P_i = P(\omega_i)$, ν, P_0	State prior, prior-parameter distribution, and pooled-state prior.
$\mu_{0,i}(h)$, $\kappa_{0,i}(h)$	Prior mean and positive concentration.
$n_{i,t}(h)$, $\bar{Y}_{i,t}(h)$	Observation count and sample mean.
$m_{i,t}(h)$, $v_{i,t}(h)$, $b_{i,t}(h)$	Posterior mean, variance, and truth probability.
$\eta_{\pm}$, $n_{\min}$	Acceptance and rejection thresholds, and minimum observation count.
$\mathcal{R}^+_{i,t}$, $\mathcal{R}^-_{i,t}$, $\mathcal{R}_{i,t}$	Cumulative accepted, rejected, and resolved sets.
$\mathcal{C}_{i,t}$, $\mathcal{F}_{i,t}$	Retained candidate set and private test-and-observation history.
Acquisition and evaluation	
π, $a_{i,t}$	Acquisition rule and hypothesis selected after t completed tests.
β, $u_{i,t}(h)$, $\Lambda_t(c)$	Softmax inverse temperature, score, and competitor mass.
$\widetilde{\mu}(h)$	Thompson posterior draw.
z, $r_t(h)$	LUCB radius multiplier and confidence radius.
B, B_i, M	Total budget, local budgets, and number of states.
$\mathcal{R}^+_{\text{known}}$	Known true hypotheses, excluded from evaluation.
D_B, D^-_B, $J(B,M)$, $M^\star(B)$	Cumulative true and false acceptances, expected true acceptances, and optimal state count within the equal-split family.
D^{fin}_B, $J^{\text{fin}}(B,M)$	True recovery supported by final posteriors, and its expectation.
Bridge construction	
K, X, s_k, w_j, $H^\star$	Bridge width, number of noise variables, seeds, noise variables, and target.
$\rho(\mathcal{S})$, $N^+(\mathcal{S})$	Bridge count and number of neighbours that increase it.
m, $\bar{p}$, $F^+(t)$	Minimum count $n_{\min}$, final-stage uniform selection bound, and stagewise success bound.
Discovery times and surrogates	
T, F, S, λ, Ψ, $t^\star$	First-acceptance time, distribution, survival, discrete hazard, cumulative hazard, and optimal local duration.
α, Q, p, $F_{Q,p}$	Favourable-commitment probability, advance quota, advance probability, and binomial completion probability.
L, q_s, $p_{A,s}$, $p^\star_A$, $W_{A,s}$	Stage count, quotas, stage probabilities, slowest-stage probability, and waiting times.

and $\mathcal{R}_{i,t} = \mathcal{R}^+_{i,t} \cup \mathcal{R}^-_{i,t}$. Both sets are non-decreasing in t and record historical threshold crossings, not permanent decisions. They may overlap: a hypothesis that has crossed one threshold remains testable and may later cross the other. Seeds are included in $\mathcal{R}^+_{i,0}$ exogenously and are exempt from the observation requirement; if tested, a seed may also enter $\mathcal{R}^-_{i,t}$. Because the prior factorises and only $a_{i,t}$ is observed at step $t+1$, only the tested hypothesis can change its resolution status, so $\mathcal{R}_{i,t+1} \setminus \mathcal{R}_{i,t} \subseteq \{a_{i,t}\}$. Candidate sets evolve by equation 3. Exposure occurs when a hypothesis first enters $\mathcal{R}_{i,t}$, whether by acceptance or rejection, and is not reversed by later evidence, since both $\mathcal{R}_{i,t}$ and $\mathcal{C}_{i,t}$ are non-decreasing. Every non-seed resolved hypothesis has been tested and seeds lie in $\mathcal{C}_{i,0}$, so $\mathcal{R}_{i,t} \subseteq \mathcal{C}_{i,t}$. All candidates remain eligible for selection regardless of their resolution history.

Known set, seed schedule, and admissible architectures. A fixed set $\mathcal{R}^+_{\text{known}}$ of known true hypotheses is excluded from evaluation under every architecture. All seed sets are drawn from it, so initialisation determines access rather than what counts as a discovery. In the K-bridge world, $\mathcal{R}^+_{\text{known}} = \{s_1, \ldots, s_K\}$. For $M \leq K$, seeds are partitioned into M blocks whose sizes differ by at most one. For $M > K$, each state receives one seed assigned round-robin. The seed schedule is deterministic. The pooled state receives all seeds and prior $P(\omega_0)$ with $\omega_0 \sim \nu$; it is the $M = 1$ member of the family below, so $J_{\text{single}}(B) = J(B, 1)$. Isolated priors have the same marginal distribution, with shared or independent draws as specified. Architectures satisfy $1 \leq M \leq B$, with integer local budgets $B_i \geq 1$ split as evenly as possible. Writing $B = Mq + \ell$ with $0 \leq \ell < M$, the first ℓ states in a fixed schedule receive $q + 1$ tests and the remaining states receive q. All experiments use M dividing B, so the remainder convention does not affect the reported comparisons. The quantity $M^\star(B)$ optimises within this specified equal-split family, not over prior families, acquisition rules, or adaptive budget reallocations. Different initialisation schemes are therefore compared separately rather than jointly optimised. Propositions 3.2 and B.18 hold for arbitrary integer splits $\sum_i B_i = B$, not only equal ones.

Symmetric states. Isolated states are *symmetric* when their local discovery-time laws coincide, $F_i = F_{\text{iso}}$ for every i, where each F_i is averaged over the state's own independent prior draw. In the K-bridge world, equal-sized seed blocks (that is, M divides K, or $M \geq K$) give symmetric states if the latent means, the prior distribution, and the acquisition rule, including its tie-breaking, are invariant under the bridge-index relabellings that map those blocks to one another. Invariance of the truth labels alone is not enough, because relabelled hypotheses with different latent means resolve at different rates. The symmetry concerns laws averaged over prior draws; conditional on realised draws, states are generally not symmetric. Fixed-order tie-breaking can also break invariance, for example among untested candidates with identical priors, which tie exactly under the LUCB-style rule. When $M < K$ and M does not divide K, seed blocks differ in size and the individual laws F_i must be used.

Evaluation. The cumulative true and false acceptance counts are

$$D_B = \sum_{h \in \mathcal{H} \setminus \mathcal{R}^+_{\text{known}}} Z_h \, \mathbf{1}\Big\{h \in \textstyle\bigcup_{i=1}^{M} \mathcal{R}^+_{i,B_i}\Big\},$$

$$D_B^- = \sum_{h \in \mathcal{H} \setminus \mathcal{R}^+_{\text{known}}} (1 - Z_h) \, \mathbf{1}\Big\{h \in \textstyle\bigcup_{i=1}^{M} \mathcal{R}^+_{i,B_i}\Big\},$$

and $J(B, M) = \mathbb{E}[D_B]$. Each hypothesis counts at most once, however many states accept it. Both counts are cumulative: a hypothesis that crossed η_+ and later crossed η_- still counts. Section B.8 defines the final-posterior counterparts used in the simulations.

Resource. The budget counts experiments only, following the convention of Bayesian experimental design (Rainforth et al., 2024). Maintaining additional states, priors, rankings, or memory is free in this model. Thus $M^\star$ is an experiment-budget optimum, not a compute-optimal agent count. An initialisation cost c_{init} per state would replace $\sum_i B_i = B$ by $\sum_i B_i + c_{\text{init}} M \leq B$; we do not analyse this extension.

Random initialisations. When $\mathcal{I}$ is random, conditional formulas are evaluated first and then averaged. In particular,

$$\mathbb{E}_{\mathcal{I}}\left[\prod_i \big(1 - F_i(t \mid \mathcal{I})\big)\right]$$

cannot generally be replaced by the product of marginal failure probabilities when states share components of $\mathcal{I}$. Independent prior draws and independent trajectory randomness do give independence after averaging over those draws, conditional on the world and deterministic seed schedule. In that case the product formulas below hold with each F_i replaced by its average over $\omega_i \sim \nu$; this is the form used in Proposition 3.2. Proposition 3.4 treats the difference between shared and independent draws.

Shared datasets. If observations are resampled from a fixed finite dataset rather than generated afresh, trajectories remain conditionally independent given that dataset when sampling randomness is independent across states. Unconditional dependence can enter through the shared data. Communication or coordinated selection would introduce additional dependence not included here.

Finiteness. Every non-seed acceptance requires at least $n_{\min} \geq 1$ observations of that hypothesis by the accepting state. Hypotheses counted by D_B or D_B^- lie outside $\mathcal{R}^+_{\text{known}}$, so none is a seed, and distinct counted hypotheses use disjoint tests. Since the states perform B tests in total, $D_B + D_B^- \leq B/n_{\min} \leq B$, and all expectations below are finite.

Lemma B.1 (Restart equivalence)**.** *Consider a sequential schedule that starts state i from its assigned seeds and prior, runs it for B_i tests with private history, records its acceptances, and then proceeds to state $i+1$. The joint law of $(\mathcal{R}^+_{i,B_i}, \mathcal{R}^-_{i,B_i})_{i=1}^M$ is the same as under isolated parallel execution. In particular, cumulative true and false acceptances are equal in law. The same conclusion holds for any evaluation determined by the terminal private histories, including final-posterior recovery.*

Proof. Under either schedule, each terminal private history is the same measurable function of

$$\big(\mathcal{W},\, \mathcal{R}^+_{i,0},\, \omega_i,\, (\xi_{i,h,k})_{h,k},\, (U_{i,t})_t\big).$$

These arguments have the same joint law under both schedules. □

B.2 Bridge Universe, Access Depth, and Exposure

Definition B.2 (K-bridge world)**.** Let $K \geq 3$, $X \geq 0$, and

$$\mathcal{V} = \{v_1, \ldots, v_K\} \cup \{v'_1, \ldots, v'_K\} \cup \{w_1, \ldots, w_X\}, \qquad n = 2K + X.$$

Hypotheses are identified with scopes,

$$\mathcal{H} = \{\mathcal{S} \subseteq \mathcal{V} : 2 \leq |\mathcal{S}| \leq K\},$$

with one-edit adjacency equation 1. The seeds $s_k = \{v_k, v'_k\}$ and target $H^\star = \{v_1, \ldots, v_K\}$ are true, and every other hypothesis is false. The *bridge count* is $\rho(\mathcal{S}) = |\mathcal{S} \cap \{v_1, \ldots, v_K\}|$.

The following lemma holds in any hypothesis graph and under any acquisition rule.

Lemma B.3 (Exposure chains)**.** *Let $h \in \mathcal{C}_{i,t} \setminus \mathcal{C}_{i,0}$. There exist $\ell \geq 1$ and non-seed hypotheses $h_1, \ldots, h_\ell \in \mathcal{R}_{i,t}$ such that, with $h_0 = h$, the hypotheses $h_0, \ldots, h_\ell$ are distinct, $h_{j+1} \in \Gamma(h_j)$ for $0 \leq j < \ell$, and $h_\ell \in \mathcal{C}_{i,0}$. Moreover, for each $0 \leq j < \ell$, the first $n_{\min}$ tests of h_{j+1} all precede every test of h_j.*

Proof. Since $\mathcal{R}_{i,s+1} \setminus \mathcal{R}_{i,s} \subseteq \{a_{i,s}\}$ (Section B.1), equation 3 gives $\mathcal{C}_{i,s+1} \setminus \mathcal{C}_{i,s} \subseteq \Gamma(a_{i,s})$, with a nonempty difference only if $a_{i,s}$ is first resolved at time $s+1$. For $g \notin \mathcal{C}_{i,0}$ that enters the candidate set at time $s+1$, let $\mathrm{par}(g) = a_{i,s}$ be its *exposing parent*. Then $\mathrm{par}(g) \in \Gamma(g)$, $\mathrm{par}(g)$ is a non-seed first resolved at time $s+1$, and $\mathrm{par}(g)$ entered the candidate set strictly before g. Set $h_{j+1} = \mathrm{par}(h_j)$ while $h_j \notin \mathcal{C}_{i,0}$. Candidate-entry times strictly decrease along the sequence, so it terminates at some $h_\ell \in \mathcal{C}_{i,0}$ with distinct terms. Every test of h_j occurs after h_j enters the candidate set, hence after the first resolution of h_{j+1}, which requires at least $n_{\min}$ tests of h_{j+1}. □

Lemma B.4 (Access depth and connectivity)**.**

(i) If $(\mathcal{S}, \mathcal{T}) \in \mathcal{E}$, then $|\rho(\mathcal{T}) - \rho(\mathcal{S})| \leq 1$; in particular, $\rho(\mathcal{T}) \leq \rho(\mathcal{S}) + 1$.

(ii) The target cannot enter a candidate set before the state has resolved at least $K-2$ distinct non-seed hypotheses, all of which are false. The path $\{v_1, v'_1\} \to \{v_1, v_2\} \to \{v_1, v_2, v_3\} \to \cdots \to H^\star$ attains this bound.

(iii) The hypothesis graph is connected.

Proof. For (i), one-edit adjacency gives $|\mathcal{T} \setminus \mathcal{S}| \leq 1$ and $|\mathcal{S} \setminus \mathcal{T}| \leq 1$, so at most one bridge variable is gained and at most one is lost. For (ii), apply Lemma B.3 to $h = H^\star$, which is not in $\mathcal{C}_{i,0}$ because its neighbours have bridge count at least $K - 1 \geq 2$, whereas seeds have bridge count 1. The chain

ends at $h_\ell \in \mathcal{C}_{i,0}$, a non-seed neighbour of a seed, so $\rho(h_\ell) \le 2$. Since $\rho(H^\star) = K$, part (i) gives $\ell \ge K-2$. For $1 \le j \le K-2$, the ancestor h_j satisfies $\rho(h_j) \ge K-j \ge 2$, so it cannot be any known seed, all of which have bridge count one. These first $K-2$ ancestors are distinct resolved hypotheses different from the target and are therefore false. The displayed path consists of one replacement followed by additions within the permitted arities. For (iii), any two scopes of size two are joined by at most two replacements, and every larger scope is connected to one of size two by deletions. □

Lemma B.5 (Exposure counts)**.** *Let $|\mathcal{S}| = j$ and $\rho(\mathcal{S}) = b$. Then*

$$|\Gamma(\mathcal{S})| = j\,\mathbf{1}\{j \ge 3\} + (n-j)\,\mathbf{1}\{j \le K-1\} + j(n-j).$$

No neighbour has bridge count exceeding $b+1$, and the number with bridge count $b+1$ is

$$N^+(\mathcal{S}) = (K-b)\big(\mathbf{1}\{j \le K-1\} + j - b\big).$$

Proof. Neighbours are deletions, additions, and replacements, of arities $j-1$, $j+1$, and j, respectively. There are j valid deletions when $j \ge 3$, $n-j$ valid additions when $j \le K-1$, and $j(n-j)$ replacements. The three classes are disjoint, and distinct replacement pairs give distinct scopes. Deletions never increase bridge count. An addition increases it when the added variable is one of the $K-b$ missing bridge variables. A replacement increases it when one of the $j-b$ non-bridge variables is replaced by a missing bridge variable. □

Corollary B.6 (Final-stage exposure)**.** *If $\rho(\mathcal{S}) = K-1$, then $H^\star$ is its unique neighbour that increases bridge count and $|\Gamma(\mathcal{S})| \ge K(n-K+1)$. Equality holds when $|\mathcal{S}| = K$ or $K = 3$. After such a scope is resolved, $|\mathcal{C}_{i,t}| \ge K(n-K+1)+1$ at every subsequent selection time. Under uniform selection,*

$$\Pr(a_{i,t} = H^\star \mid \mathcal{F}_{i,t}) \le \bar{p} := \frac{1}{K(n-K+1)+1}.$$

The number of target tests in any fixed-length block of w subsequent tests is stochastically dominated by $\mathrm{Bin}(w, \bar{p})$. Before any scope with $K-1$ bridge variables is resolved, $H^\star$ is not exposed. Finally, every state satisfies $|\mathcal{C}_{i,t}| \ge |\mathcal{C}_{i,0}| \ge 3n-5$ for all t, so under uniform selection each candidate is selected with conditional probability at most $1/(3n-5)$.

Proof. The target is the only permitted scope with bridge count K. If $|\mathcal{S}| = K-1$, Lemma B.5 gives

$$|\Gamma(\mathcal{S})| = (K-1)\mathbf{1}\{K \ge 4\} + K(n-K+1), \qquad N^+(\mathcal{S}) = 1.$$

If $|\mathcal{S}| = K$, it contains one non-bridge variable and $|\Gamma(\mathcal{S})| = K + K(n-K) = K(n-K+1)$, again with $N^+(\mathcal{S}) = 1$. After resolution, the retained candidate set contains both $\mathcal{S}$ and all its neighbours. Uniform selection therefore has the stated conditional bound. Coupling the target-selection indicators with independent uniforms gives binomial domination. Every neighbour of $H^\star$ has bridge count $K-1$, and no seed is its neighbour for $K \ge 3$. Finally, a seed has $|\mathcal{S}| = 2$ and $\rho = 1$, so Lemma B.5 gives $3(n-2)$ neighbours; with the seed itself, $|\mathcal{C}_{i,0}| \ge 3n-5$, and candidate sets are non-decreasing. □

Example ($K = 4$)**.** Along the canonical path, a seed has $3(n-2)$ neighbours, six of which increase bridge count. The scope $\{v_1, v_2\}$ also has $3(n-2)$ neighbours, but only two increase bridge count. The scope $\{v_1, v_2, v_3\}$ has $4(n-3)+3$ neighbours, of which only $H^\star$ increases bridge count. For $X = 10$ and $n = 18$, these neighbourhood sizes are 48, 48, and 63.

B.3 Gaussian Beliefs and Resolution

Posterior updates. Write $n_{i,t}(h)$ for the number of observations of h in $\mathcal{F}_{i,t}$ and $\bar{Y}_{i,t}(h)$ for their sample mean. Under the subjective model P_i, Gaussian conjugacy gives a normal posterior for μ_h given $\mathcal{F}_{i,t}$, with mean and variance

$$m_{i,t}(h) = \frac{\kappa_{0,i}(h)\mu_{0,i}(h) + n_{i,t}(h)\bar{Y}_{i,t}(h)}{\kappa_{0,i}(h) + n_{i,t}(h)}, \qquad v_{i,t}(h) = \frac{\sigma^2}{\kappa_{0,i}(h) + n_{i,t}(h)},$$

where the observation contribution $n_{i,t}(h)\bar{Y}_{i,t}(h)$ is zero when $n_{i,t}(h) = 0$. The belief equation 2 follows because $(\mu_h - m_{i,t}(h))/\sqrt{v_{i,t}(h)}$ is standard normal under the posterior. Adaptive selection does not change this posterior update: conditional on the recorded history, the acquisition rule contributes no additional likelihood factor for μ_h. Because the prior factorises across hypotheses and each observation concerns one hypothesis, the posterior also factorises, so an untested hypothesis retains its state-specific prior belief and $b_{i,t}(h)$ changes only when h is tested.

Standardised thresholds. Let $z_+ = \Phi^{-1}(\eta_+)$ and $z_- = \Phi^{-1}(1-\eta_-)$. Once $n_{i,t}(h) \geq n_{\min}$, the acceptance criterion $b_{i,t}(h) \geq \eta_+$ is equivalent to $(m_{i,t}(h) - \tau)/\sqrt{v_{i,t}(h)} \geq z_+$, and the rejection criterion $b_{i,t}(h) \leq \eta_-$ to $(\tau - m_{i,t}(h))/\sqrt{v_{i,t}(h)} \geq z_-$. With symmetric thresholds, $\eta_- = 1 - \eta_+$, we write $z_{\text{dec}} = z_+ = z_-$.

Lemma B.7 (Consistency)**.** *If state i tests h infinitely often, then almost surely $m_{i,t}(h) \to \mu_h$, $v_{i,t}(h) \to 0$, and $b_{i,t}(h) \to Z_h$. Thus a true hypothesis eventually crosses the acceptance threshold, and a false hypothesis eventually crosses the rejection threshold. If h is tested finitely often, its posterior is eventually constant.*

Proof. The strong law holds for every observation stack, so along infinitely many tests of h its sample mean converges to μ_h. The posterior formulas then give the stated mean and variance limits. Since $\mu_h \neq \tau$, the standardised posterior mean tends to the appropriate signed infinity. The minimum observation count is eventually met. □

B.4 Policy-Facing Competition

All probabilities in this section condition on the world, initialisation, and current history $\mathcal{F}_t = \mathcal{F}_{i,t}$. For notational simplicity, we suppress the state index where convenient. The candidate set $\mathcal{C}_t$ and acquisition scores are functions of the history.

B.4.1 Uniform selection

Uniform selection draws from the entire retained candidate set. For any $\mathcal{F}_t$-measurable subset $\mathcal{G} \subseteq \mathcal{C}_t$,

$$\Pr(a_t \in \mathcal{G} \mid \mathcal{F}_t) = \frac{|\mathcal{G}|}{|\mathcal{C}_t|}.$$

Candidate-set size is non-decreasing, and Corollary B.6 gives the final-stage bridge bound.

B.4.2 Softmax selection

For a history-measurable score $u_t(h)$ and fixed inverse temperature $\beta \geq 0$ ($\beta = 10$ in our simulations), softmax selects

$$\Pr(a_t = h \mid \mathcal{F}_t) = \frac{\exp\{\beta u_t(h)\}}{\sum_{g \in \mathcal{C}_t} \exp\{\beta u_t(g)\}}, \qquad h \in \mathcal{C}_t.$$

For a designated candidate c, define the competitor mass

$$\Lambda_t(c) = \sum_{g \in \mathcal{C}_t \setminus \{c\}} \exp\{\beta[u_t(g) - u_t(c)]\}.$$

Then $\Pr(a_t = c \mid \mathcal{F}_t) = 1/(1 + \Lambda_t(c))$. For a useful set, the numerator sums its members' weights. Our simulations use the Bernoulli entropy of the current belief,

$$u_t(h) = -b_t(h)\log b_t(h) - \big(1 - b_t(h)\big)\log\big(1 - b_t(h)\big) \in [0, \log 2],$$

which favours hypotheses whose truth is most uncertain.

Lemma B.8 (Softmax and uniform selection)**.** *Whenever all scores lie in $[u_{\text{lo}}, u_{\text{hi}}]$, every $\mathcal{F}_t$-measurable $\mathcal{G} \subseteq \mathcal{C}_t$ satisfies*

$$e^{-\beta(u_{\text{hi}} - u_{\text{lo}})}\frac{|\mathcal{G}|}{|\mathcal{C}_t|} \leq \Pr(a_t \in \mathcal{G} \mid \mathcal{F}_t) \leq e^{\beta(u_{\text{hi}} - u_{\text{lo}})}\frac{|\mathcal{G}|}{|\mathcal{C}_t|}.$$

If the score range is bounded by deterministic constants uniformly over histories, Corollary B.6 and Proposition 3.1(ii) hold with $1/(3n-5)$ *and* $\bar{p}$ *replaced by* $\min\{1, e^{\beta(u_{\mathrm{hi}}-u_{\mathrm{lo}})}/(3n-5)\}$ *and* $\min\{1, e^{\beta(u_{\mathrm{hi}}-u_{\mathrm{lo}})}\bar{p}\}$.

Proof. Every weight lies between $e^{\beta u_{\mathrm{lo}}}$ and $e^{\beta u_{\mathrm{hi}}}$. Taking $\mathcal{G}$ to be a single candidate bounds each conditional selection probability by $e^{\beta(u_{\mathrm{hi}}-u_{\mathrm{lo}})}/|\mathcal{C}_t|$. The proof of Proposition 3.1(ii) uses uniform selection only through these per-candidate bounds. □

For the entropy score, $e^{\beta(u_{\mathrm{hi}}-u_{\mathrm{lo}})} = 2^{\beta}$, so this extension is informative only when 2^{β} is small relative to $3n-5$ and $1/\bar{p}$.

A deterministic score-range bound is stronger than pathwise boundedness. For example, Bernoulli posterior entropy lies in $[0, \log 2]$, whereas posterior means are bounded along almost every realised trajectory but need not have a deterministic uniform bound.

B.4.3 THOMPSON SAMPLING

Conditional on the current state, draw $\widetilde{\mu}(h) \sim \mathcal{N}(m(h), v(h))$ independently across candidates, with $v(h) > 0$. Thompson sampling selects the candidate with the largest draw; ties have probability zero.

Lemma B.9 (Exact Thompson selection probability)**.** *For $c \in \mathcal{C}$, let f_c and F_g denote the density and distribution functions of the posterior draws. Then*

$$p_{\mathcal{C}}(c) = \int_{-\infty}^{\infty} f_c(x) \prod_{g \in \mathcal{C}\setminus\{c\}} F_g(x)\, dx.$$

If a nonempty set $\mathcal{U}$ of additional independent nondegenerate Gaussian candidates is exposed while existing draw distributions are held fixed, then $p_{\mathcal{C}\cup\mathcal{U}}(c) < p_{\mathcal{C}}(c)$.

Proof. Condition on $\widetilde{\mu}(c) = x$ and integrate. Additional candidates multiply the positive integrand by $\prod_{u\in\mathcal{U}} F_u(x) \in (0,1)$. □

Lemma B.10 (Useful-set competition)**.** *Let $\varnothing \neq \mathcal{G} \subseteq \mathcal{C}$ be a fixed useful set and define*

$$U = \max_{h\in\mathcal{G}} \widetilde{\mu}(h), \qquad V = \max_{g\in\mathcal{C}\setminus\mathcal{G}} \widetilde{\mu}(g),$$

with $V = -\infty$ when the competitor set is empty. Then $p_{\mathrm{useful}} = \Pr(U > V)$. Adding independent non-useful candidates weakly decreases this probability. Increasing useful posterior means or decreasing competitor posterior means, with variances fixed, weakly increases it.

Proof. Couple the draws using $\widetilde{\mu}(h) = m(h) + \sqrt{v(h)}\,\epsilon_h$ with shared standard normals across the compared systems. The stated changes respectively increase V, increase U, or decrease V pointwise. □

Identically distributed competitors. If a prerequisite has posterior $\mathcal{N}(m_c, s_c^2)$ and N distractors have independent posteriors $\mathcal{N}(m_0, s_0^2)$, then

$$p_N(c) = \int \frac{1}{s_c}\phi\left(\frac{x-m_c}{s_c}\right)\Phi\left(\frac{x-m_0}{s_0}\right)^N dx,$$

where ϕ is the standard normal density. This equals $1/(N+1)$ when all draw distributions coincide. For every x,

$$p_N(c) \le 1 - \Phi\left(\frac{x-m_c}{s_c}\right) + \Phi\left(\frac{x-m_0}{s_0}\right)^N. \tag{5}$$

The prerequisite can win only if it exceeds x or all distractors fall below x. Dominated convergence applied to the integral gives $p_N(c) \to 0$.

Corollary B.11 (Distractor scaling)**.** *For $N \geq 2$, let $z_N = \bar{\Phi}^{-1}(\log N/N)$, where $\bar{\Phi} = 1 - \Phi$. Then*

$$p_N(c) \leq \bar{\Phi}\left(\frac{m_0 - m_c + s_0 z_N}{s_c}\right) + \frac{1}{N}, \qquad z_N = \sqrt{2\log N}\,(1 + o(1)).$$

The bound becomes small when $s_0 z_N$ exceeds the prerequisite's mean advantage $m_c - m_0$ by several multiples of s_c. For a positive mean advantage, the leading competitor scale is roughly $\exp\{(m_c - m_0)^2/(2s_0^2)\}$.

Proof. Take $x = m_0 + s_0 z_N$ in equation 5. Then $\Phi(z_N)^N = (1 - \log N/N)^N \leq 1/N$. The asymptotic expression follows from the Gaussian tail approximation. □

A sum of pairwise competitor probabilities does not determine the probability of winning against all competitors. For one candidate and N identically distributed competitors, that sum is $N/2$, but the winning probability is $1/(N+1)$ rather than $1/(1+N/2)$. The integral accounts for the dependence between pairwise comparison events.

B.4.4 LUCB-STYLE THRESHOLDING

Let

$$r_t(h) = \frac{z\sigma}{\sqrt{\kappa_0(h) + n_t(h)}}, \qquad z > 0,$$

and define lexicographic priority

$$\text{priority}_t(h) = \left(\mathbf{1}\{|m_t(h) - \tau| \leq r_t(h)\},\ r_t(h),\ -\frac{|m_t(h) - \tau|}{r_t(h)}\right).$$

The rule selects a maximum-priority candidate, breaking remaining ties by a fixed order. A candidate whose interval contains τ is said to *straddle*. This is the specific LUCB-style thresholding rule analysed here, adapted to the non-stationary nature of the available arms to be tested.

Lemma B.12 (Uncertainty-balancing delay)**.** *Assume a common prior concentration on $\mathcal{U} \cup \{c\}$. Let $c \in \mathcal{C}_t$ have $n_c \geq 1$ observations, and let $\mathcal{U} \subseteq \mathcal{C}_t \setminus \{c\}$ satisfy $n_t(g) < n_c$ for every $g \in \mathcal{U}$. Suppose each such g remains selectable and straddles whenever its count is below n_c. Then before c is selected again, at least $\sum_{g\in\mathcal{U}}(n_c - n_t(g))$ tests must be allocated to $\mathcal{U}$.*

Proof. Until c is selected, its count and radius are fixed. Any $g \in \mathcal{U}$ below count n_c straddles and has strictly larger radius. It therefore beats c on the first priority coordinate if c does not straddle, and on the second otherwise. Thus every such g must reach count n_c before c can be selected. □

Corollary B.13 (Newly exposed competitors)**.** *Assume a common prior concentration on c and the new candidates. If c has one observation and L newly exposed candidates initially straddle, all L receive a first observation before c is selected again.*

Proof. Apply Lemma B.12 with $n_c = 1$. Before its first observation, a candidate's posterior and straddling status remain fixed. □

Priors and the straddling class. A fresh candidate g straddles for state i iff $|\mu_{0,i}(g) - \tau| \leq z\sigma/\sqrt{\kappa_{0,i}(g)}$. Thus priors partition newly exposed hypotheses into acquisition-priority classes. The number of fresh neighbours of h that initially straddle is

$$A_i(h;t) = \sum_{g\in\Gamma(h)\setminus\mathcal{C}_{i,t}} \mathbf{1}\left\{|\mu_{0,i}(g) - \tau| \leq \frac{z\sigma}{\sqrt{\kappa_{0,i}(g)}}\right\}.$$

Some may be useful continuations rather than distractors. With symmetric decision thresholds and the minimum count met, a hypothesis can currently satisfy a decision threshold while its acquisition interval still straddles iff

$$z_{\text{dec}} \leq \frac{|m_t(h) - \tau|}{\sqrt{v_t(h)}} \leq z.$$

This window is nonempty iff $z \geq z_{\text{dec}}$. Consequently, first resolution, exposure, and departure from the straddling class need not coincide.

B.4.5 EXPLORATION

A rule *explores* if, almost surely, every hypothesis that enters $\mathcal{C}_{i,t}$ is selected infinitely often as $t \to \infty$.

Lemma B.14 (Exploration of the four rules)**.** *Assume $\mathcal{H}$ is finite, candidate sets are retained, $\mu_h \neq \tau$, and $\kappa_{0,i}(h) > 0$. Uniform selection, Thompson sampling, LUCB-style acquisition, and softmax with fixed finite β and an almost surely bounded score along the trajectory (including the entropy score) explore. Bernoulli posterior entropy is bounded by $\log 2$, and posterior means are almost surely bounded along each trajectory.*

Proof. Every exposed hypothesis enters at a finite time. For each hypothesis, its posterior mean either converges along infinitely many observations or becomes constant after finitely many observations. Since the hypothesis universe is finite, $\bar{U} = \max_h \sup_t |m_{i,t}(h)| < \infty$ almost surely. By the conditional Borel–Cantelli lemma (Williams, 1991, Thm. 12.15), a hypothesis is selected infinitely often on the event that the sum of its conditional selection probabilities diverges.

For uniform selection, every exposed hypothesis has selection probability at least $1/|\mathcal{H}|$. For softmax, if $\bar{u} = \sup_{t,g\in\mathcal{C}_{i,t}} |u_t(g)| < \infty$, every exposed hypothesis has probability at least $e^{-2\beta\bar{u}}/|\mathcal{H}|$.

For Thompson sampling, suppose an exposed h is selected finitely often. Its posterior is eventually fixed at (m_h, v_h) with $v_h > 0$. Let $\kappa_{\min} = \min_g \kappa_{0,i}(g)$ and $c = \sigma/\sqrt{\kappa_{\min}}$. Other candidates have posterior means bounded by $\bar{U}$ and standard deviations at most c. If $\widetilde{\mu}(h) > \bar{U} + c$ and $\widetilde{\mu}(g) < m(g) + c$ for every other candidate g, then h has the largest draw. Independence of the posterior draws therefore gives the eventual lower bound

$$\Pr(a_{i,t} = h \mid \mathcal{F}_{i,t}) \geq \bar{\Phi}\left(\frac{\bar{U} + c - m_h}{\sqrt{v_h}}\right) \Phi(1)^{|\mathcal{H}|} > 0.$$

Conditional Borel–Cantelli contradicts finite selection.

For the LUCB-style rule, suppose an exposed h is selected finitely often. Because the eventual candidate pool is finite, after some time only hypotheses in a nonempty set I of infinitely selected hypotheses are chosen. For every $g \in I$, consistency gives $m_t(g) \to \mu_g \neq \tau$ and $r_t(g) \to 0$. Eventually none of them straddles, and all have radii smaller than the fixed positive radius of h. Then h beats each member of I on the first priority coordinate if it straddles, and on the second otherwise. This contradicts selection from I alone. □

B.5 COMPLETION-TIME STRUCTURE

The following processes are completion-time surrogates rather than exact reductions of the Gaussian search model. They replace adaptive selection and random posterior threshold crossings by explicit progress probabilities and quotas. They therefore establish sufficient crossover mechanisms without asserting that Gaussian bridge trajectories obey these laws.

B.5.1 THE STAGEWISE SURROGATE

Definition B.15 (Stagewise bridge process)**.** Fix $L \geq 1$, quotas $q_1, \ldots, q_L \in \mathbb{N}$, and probabilities $p_{A,1}, \ldots, p_{A,L} \in (0, 1)$ for architecture A. The process begins at stage 1. At stage s, each test independently produces an advance with probability $p_{A,s}$. After q_s advances, the process enters the next stage; completion of stage L records discovery. Randomness is independent across stages and isolated processes. Write

$$Q = \sum_{s=1}^{L} q_s, \qquad p_A^\star = \min_{1\leq s\leq L} p_{A,s}.$$

Lemma B.16 (Stagewise completion distribution)**.** *The stage waiting times are independent with $W_{A,s} \sim \mathrm{NegBin}(q_s, p_{A,s})$ on $\{q_s, q_s + 1, \ldots\}$ and*

$$\Pr(W_{A,s} = w) = \binom{w-1}{q_s - 1} p_{A,s}^{q_s} (1 - p_{A,s})^{w - q_s}.$$

For $T_A = \sum_s W_{A,s}$,

$$\mathbb{E}[\zeta^{T_A}] = \prod_{s=1}^{L} \left(\frac{p_{A,s}\zeta}{1-(1-p_{A,s})\zeta} \right)^{q_s},$$

$$\mathbb{E}[T_A] = \sum_s \frac{q_s}{p_{A,s}}, \qquad \mathrm{Var}(T_A) = \sum_s \frac{q_s(1-p_{A,s})}{p_{A,s}^2}.$$

Proof. Each stage waits for a fixed number of independent Bernoulli successes. Independence gives the product generating function and additive moments. □

The completion distribution can be computed by

$$f_0(t) = \mathbf{1}\{t=0\}, \qquad f_s(t) = \sum_{w=q_s}^{t} f_{s-1}(t-w) \binom{w-1}{q_s-1} p_{A,s}^{q_s} (1-p_{A,s})^{w-q_s},$$

with $F_A(B) = \sum_{t=0}^{B} f_L(t)$. If all stage probabilities equal p_A, then $F_A(B) = \Pr\{\mathrm{Bin}(B, p_A) \geq Q\}$.

B.5.2 Increasing failure rate and splitting

Write $\mathbb{N} = \{1, 2, \dots\}$. For $T \in \mathbb{N} \cup \{\infty\}$, let $S(t) = \Pr(T > t)$ and, whenever $S(t-1) > 0$,

$$\lambda(t) = \Pr(T = t \mid T \geq t), \qquad t \geq 1.$$

The law is IFR when $\lambda(t)$ is non-decreasing. The variable $T = \infty$ almost surely has zero hazard and is IFR.

Lemma B.17 (Log-concavity implies IFR)**.** *(i) A proper distribution on $\mathbb{N}$ with a log-concave mass function and no internal zeros is IFR. (ii) The stagewise completion time T_A has such a mass function and is IFR.*

Proof. For (i), let $\pi_k = \Pr(T = k)$ on its interval of support. Before that support begins, the hazard is zero. Within the support, log-concavity makes π_{k+1}/π_k non-increasing. Hence, for each $\ell \geq 0$, the ratio $\pi_{t+\ell}/\pi_t$ is non-increasing in t, with terms beyond a finite support taken as zero. Therefore $1/\lambda(t) = \sum_{\ell \geq 0} \pi_{t+\ell}/\pi_t$ is non-increasing. For (ii), T_A is a sum of Q independent geometric variables on $\{1, 2, \dots\}$. Their mass functions are log-concave without internal zeros, and convolution preserves this property (Hoggar, 1974). □

Proposition B.18 (No splitting advantage under IFR completion times)**.** *Let $T_1, \dots, T_M$ be independent copies of an IFR variable T, and let the single process have the same discovery-time law. For all integer budgets $B_i \geq 0$ with $\sum_i B_i = B$,*

$$1 - \prod_{i=1}^{M} S(B_i) \leq 1 - S(B).$$

Thus neither equal nor unequal splitting improves single-target discovery.

Proof. If $S(B) = 0$, the result is immediate. Otherwise, for $t \leq B$,

$$\Psi(t) = -\log S(t) = -\sum_{s=1}^{t} \log(1 - \lambda(s))$$

has non-decreasing increments and $\Psi(0) = 0$. For $a + b \leq B$,

$$\Psi(a+b) - \Psi(b) \geq \Psi(a) - \Psi(0),$$

so Ψ is superadditive. Induction gives $\sum_i \Psi(B_i) \leq \Psi(B)$. Exponentiating yields the result. □

B.5.3 STATIONARY EFFICIENCY CROSSOVER

Lemma B.19 (Completion-time tail rate)**.** *For the stagewise process,*

$$\lim_{t\to\infty} \frac{1}{t} \log \Pr(T_A > t) = \log(1 - p_A^\star).$$

Proof. For a slowest stage $s^\star$,

$$\Pr(T_A > t) \geq \Pr(W_{A,s^\star} > t) = \sum_{j=0}^{q_{s^\star}-1} \binom{t}{j} (p_A^\star)^j (1 - p_A^\star)^{t-j}$$

for sufficiently large t. This has logarithmic rate $\log(1 - p_A^\star)$. For the upper bound, the moment generating function of T_A is finite whenever $0 < \theta < -\log(1 - p_A^\star)$. Chernoff's inequality and passage to the endpoint give the matching rate. □

Proposition B.20 (Stationary efficiency crossover)**.** *Consider two stagewise architectures with the same quotas, $Q \geq 2$, $M \geq 2$, independent isolated processes, and $p_{\mathrm{iso}}^\star > p_{\mathrm{single}}^\star$.*

(i) For budgets divisible by M with $Q \leq B < MQ$, $J_{\mathrm{single}}(B) > J_{\mathrm{iso}}(B, M) = 0$.

(ii) There is a finite B_0 such that $J_{\mathrm{iso}}(B, M) > J_{\mathrm{single}}(B)$ for every divisible $B \geq B_0$.

(iii) If stage probabilities are homogeneous within each architecture, write them as p_{s} and p_{i}. Let

$$\Delta_p = \log \frac{1 - p_{\mathrm{s}}}{1 - p_{\mathrm{i}}} > 0, \qquad g(B) = B\Delta_p - M \log Q - M(Q-1) \log \frac{Bp_{\mathrm{i}}}{M(1 - p_{\mathrm{i}})}.$$

Define

$$B_{\min} = \max \left\{ \frac{M(1 - p_{\mathrm{i}})}{p_{\mathrm{i}}}, \frac{M(Q-1)}{\Delta_p} \right\}.$$

Then g is non-decreasing on $[B_{\min}, \infty)$, and the isolated system is strictly better for every divisible $B \geq B_{\min}$ with $g(B) > 0$. For fixed p_{i}, the least such budget is $O(\Delta_p^{-1} MQ \log(MQ/\Delta_p))$ as $\Delta_p \to 0$.

Proof. For (i), each process needs at least Q tests. The pooled process completes in exactly Q tests with positive probability, while each isolated budget is below Q. A divisible budget in the specified interval exists because $M\lceil Q/M \rceil < MQ$ for $Q \geq 2$.

For (ii), independence and Lemma B.19 give

$$\frac{1}{B} \log[1 - J_{\mathrm{iso}}(B, M)] \to \log(1 - p_{\mathrm{iso}}^\star),$$

whereas the pooled failure rate tends to $\log(1 - p_{\mathrm{single}}^\star)$. The isolated-to-pooled failure ratio tends to zero.

For (iii), pooled failure is at least $(1 - p_{\mathrm{s}})^B$. With $t = B/M$ and $x = tp_{\mathrm{i}}/(1 - p_{\mathrm{i}}) \geq 1$,

$$\Pr\{\mathrm{Bin}(t, p_{\mathrm{i}}) < Q\} \leq (1 - p_{\mathrm{i}})^t \sum_{j=0}^{Q-1} \frac{x^j}{j!} \leq (1 - p_{\mathrm{i}})^t Q x^{Q-1}.$$

Thus isolated failure is at most $(1 - p_{\mathrm{i}})^B Q^M x^{M(Q-1)}$, which is below the pooled lower bound when $g(B) > 0$. Also, $g'(B) = \Delta_p - M(Q-1)/B \geq 0$ on the stated interval. Substitution of $B = A\Delta_p^{-1} MQ \log(MQ/\Delta_p)$ with sufficiently large fixed A gives the order bound. □

B.6 PROOFS FOR SECTION 3

Proof of Proposition 3.2. Run each state without a budget limit and let $T_i = \inf\{t : H^\star \in \mathcal{R}^+_{i,t}\} \in \mathbb{N} \cup \{\infty\}$ be its discovery time. Because the acquisition rule does not depend on the budget (Section B.1), a state with local budget B_i has accepted $H^\star$ by the end of its budget iff $T_i \le B_i$. In the bridge world, D_B is the indicator that some state has accepted $H^\star$, so J is a success probability. The pooled system succeeds iff $T_0 \le B$, so $J_{\mathrm{single}}(B) = F_0(B)$. The isolated system fails iff $T_i > B_i$ for every i. Conditional on the world, the events $\{T_i > B_i\}$ are independent when initialisations are fixed, because states use disjoint independent stacks and randomisation variables, and remain independent after averaging over independent prior draws. Hence $1 - J_{\mathrm{iso}}(B, M) = \prod_i S_i(B_i)$, where S_i is averaged over $\omega_i \sim \nu$ when the prior is random. For symmetric states and M dividing B, this equals $S_{\mathrm{iso}}(B/M)^M$, and $J_{\mathrm{iso}}(B, M) > J_{\mathrm{single}}(B)$ iff $S_{\mathrm{iso}}(B/M)^M < S_0(B)$. Taking negative logarithms, with the convention $-\log 0 = \infty$, gives $M\Psi_{\mathrm{iso}}(B/M) > \Psi_0(B)$; when both sides are infinite, both systems succeed almost surely and neither is preferred. □

Allocation criterion. Substituting $S_i = \exp(-\Psi_i)$, Proposition 3.2 reads

$$J_{\mathrm{iso}} = 1 - \exp\left\{-\sum_i \Psi_i(B_i)\right\},$$

so for fixed local laws, maximising success over integer splits with $\sum_i B_i = B$ is equivalent to maximising the total cumulative hazard $\sum_i \Psi_i(B_i)$. The local laws may themselves depend on the architecture, for example through the seed blocks. Writing Ψ_M for the local cumulative hazard of a symmetric M-state architecture, with $\Psi_1 = \Psi_0$, the preferred state counts among those dividing B are

$$\arg \max_{\substack{1 \le M \le B \\ M \mid B}} M\Psi_M(B/M).$$

These do not necessarily include $M^\star(B)$, which ranges over all $1 \le M \le B$; Section B.7 treats the unrestricted equal-split family and random initialisations.

Proof of Proposition 3.1. For (i), Lemma B.4(ii) shows that at least $K - 2$ distinct non-seed hypotheses must be resolved before the target is exposed, each after at least m tests of that hypothesis. The target therefore cannot be tested before $d = (K - 2)m$ tests have been completed. Acceptance then requires at least m target tests, so $T_i \ge (K - 1)m$.

For positive pooled success, consider m consecutive tests of each hypothesis along the canonical path, ending with the target. Require each intermediate to be rejected at its m-th observation and the target to be accepted at its m-th observation. Every required hypothesis is exposed when needed. Uniform selection gives each prescribed test positive conditional probability. Each required posterior crossing also has positive probability because a fixed-count Gaussian sample mean has full support and the posterior belief ranges continuously over $(0, 1)$. This event therefore has positive probability. Under cumulative evaluation, later observations cannot remove the acceptance.

For (ii), fix a state and index its tests by $k = 1, 2, \ldots$, where test k selects $a_{i,k-1}$. Suppose the state accepts $H^\star$ within t tests. Lemma B.3, applied to $H^\star$ as in the proof of Lemma B.4(ii), gives distinct resolved non-seed hypotheses $h_1, \ldots, h_{K-2}$ with the ordering property of that lemma. Let Λ_0 be the steps of the first m tests of $H^\star$ and, for $1 \le j \le K - 2$, let Λ_j be the steps of the first m tests of h_j. These $(K - 1)m$ steps are distinct and lie in $\{1, \ldots, t\}$, and the ordering property gives $\Lambda_{K-2} < \Lambda_{K-3} < \cdots < \Lambda_0$ elementwise. Hence, listing $\Lambda = \bigcup_j \Lambda_j$ in increasing order, its consecutive blocks of m steps are $\Lambda_{K-2}, \ldots, \Lambda_1, \Lambda_0$.

For a fixed set $\Lambda \subseteq \{1, \ldots, t\}$ of size $(K - 1)m$, split into consecutive blocks in this way, let E_Λ be the event that every test in each of the first $K - 2$ blocks selects the same hypothesis as the first test of that block, and every test in the last block selects $H^\star$. Process the steps of Λ in increasing order. At a step that is not the first of one of the first $K - 2$ blocks, the required hypothesis is determined by the history, so uniform selection and Corollary B.6 bound its conditional probability by $1/(3n - 5)$. At a step of the last block, the conditional probability of selecting $H^\star$ is at most $\bar{p}$, again by Corollary B.6. Successive conditioning therefore gives

$$\Pr(E_\Lambda) \le (3n - 5)^{-(K-2)(m-1)}\, \bar{p}^{\,m}.$$

Acceptance within t tests implies E_Λ for the set Λ constructed above, and there are $\binom{t}{(K-1)m}$ possible sets, with the convention $\binom{t}{q} = 0$ for $t < q$, so a union bound gives $\Pr(T_i \le t) \le F^+(t)$. Since F^+ is non-decreasing, a union bound over states gives

$$J_{\mathrm{iso}}(B, M) \le \sum_i F^+(B_i) \le M\, F^+(\lceil B/M \rceil).$$

When M divides B, $\binom{t}{k} \le (et/k)^k$ with $t = B/M$ and $k = (K-1)m$ gives

$$M\, F^+(B/M) \le M^{1-(K-1)m} \left(\frac{eB}{(K-1)m} \right)^{(K-1)m} (3n-5)^{-(K-2)(m-1)}\, \bar{p}^{\,m}.$$

The final stage alone gives a second bound that is tighter when $m = 1$ or when t is large relative to $3n - 5$. Whenever the target is exposed, list it first in the candidate set and realise uniform selection using $U_{i,k-1}$. Since the target cannot be tested at steps $k \le d$, Corollary B.6 gives $\{a_{i,k-1} = H^\star\} \subseteq \{k > d,\ U_{i,k-1} \le \bar{p}\}$, so the number of target tests within t tests is dominated by $\mathrm{Bin}([t-d]_+, \bar{p})$, where $[x]_+ = \max\{x, 0\}$. Hence $\Pr(T_i \le t) \le \min\{F^+(t), \Pr\{\mathrm{Bin}([t-d]_+, \bar{p}) \ge m\}\}$, and the union bound over states applies to this minimum as well.

Finally, both parts hold under final-posterior recovery. The zero in (i) and both bounds in (ii) follow from $D_B^{\mathrm{fin}} \le D_B$ (Proposition B.28(i)), and positive pooled success follows from Proposition B.28(ii). □

Surrogate conventions. In Propositions 3.3, B.22 and B.26, $J(B, M)$ denotes the surrogate objective rather than the objective equation 4 of the Gaussian search process. Each state, including the single state at $M = 1$, draws its commitment independently with the same favourable probability α, and local budgets follow the equal-split convention of Section B.1. The surrogate therefore does not credit the pooled state for holding all seeds. If the single state is instead favourable with probability $\alpha_0 \in [\alpha, 1)$, part (i) of Proposition 3.3 is unchanged, and part (ii) holds for every M with $1 - (1-\alpha)^M > \alpha_0$. The divergence $M^\star(B) \to \infty$ also still holds, although comparing M_0 with $M_0 + 1$ states no longer suffices, since α_0 may exceed $1 - (1-\alpha)^{M_0+1}$. Instead, for any fixed M_0, choose $M' > M_0$ with $1 - (1-\alpha)^{M'} > \max\{\alpha_0, 1 - (1-\alpha)^{M_0}\}$; the limiting success of M' states exceeds the success ceiling of every architecture with at most M_0 states, so M' eventually beats all of them. Thus the mechanism does not require pooled and isolated commitments to be equally good.

Proof of Proposition 3.3. A state is favourable with probability α and then completes by time t with probability $F_{Q,p}(t)$. Its failure probability is therefore $1 - \alpha F_{Q,p}(t)$, and independence gives the product formula. Part (i) follows from $F_{Q,p}(t) = 0$ for $t < Q$ and positivity for $t \ge Q$.

For fixed M, equal splitting can be indexed so that increasing B by one increases one local budget by one. Thus success is non-decreasing in B. Since $F_{Q,p}(t) \to 1$, the limit is $1 - (1-\alpha)^M$, strictly exceeding α for $M \ge 2$. For any fixed M_0, every $M \le M_0$ has success at most $1 - (1-\alpha)^{M_0}$. The success of $M_0 + 1$ states eventually exceeds this bound. Hence every maximizing state count tends to infinity. □

Lemma B.21 (Product perturbation)**.** *For $a_i, b_i \in [0, 1]$,*

$$\left| \prod_i a_i - \prod_i b_i \right| \le \sum_i |a_i - b_i|.$$

Consequently, suppose independent alternative processes have local success functions satisfying

$$0 \le F_i^{\mathrm{alt}}(t) - \alpha F_{Q,p}(t) \le \varepsilon_t,$$

where ε_t is non-decreasing. Then

$$0 \le J^{\mathrm{alt}}(B, M) - J(B, M) \le \sum_i \varepsilon_{B_i} \le M \varepsilon_{\lceil B/M \rceil}.$$

Proof. Telescope the difference of products one factor at a time. Every remaining product factor lies in $[0, 1]$. Apply the bound to the failure probabilities. □

If $J_M^{\mathrm{alt}} \in [J_M, J_M + e_M]$ and $J_1^{\mathrm{alt}} \in [J_1, J_1 + e_1]$, then $J_M - J_1 > e_1$ is sufficient for an isolated advantage under the alternative processes. For two-sided approximation errors, the sufficient margin is $e_M + e_1$. These are conditional transfer statements; no such approximation bound for Gaussian bridge trajectories is assumed.

Proposition B.22 (Slow commitments). *Let $0 \le \varepsilon < p$, and let poor commitments advance with probability ε per test rather than never advancing. The local completion distribution is*

$$F_\varepsilon(t) = \alpha F_{Q,p}(t) + (1-\alpha) F_{Q,\varepsilon}(t).$$

Write J_ε for the resulting objective and J_0 for the permanent-failure surrogate. Then:

(i)

$$0 \le J_\varepsilon(B, M) - J_0(B, M) \le (1-\alpha) \sum_i \binom{B_i}{Q} \varepsilon^Q.$$

(ii) If $\delta = J_0(B, M) - J_0(B, 1) > 0$ and $(1-\alpha)\binom{B}{Q}\varepsilon^Q < \delta$, then $J_\varepsilon(B, M) > J_\varepsilon(B, 1)$.

(iii) For $\varepsilon > 0$ and fixed M, $J_\varepsilon(B, M) \to 1$ as $B \to \infty$. The fragmentation statement in Proposition 3.3(i) remains valid.

Proof. A union bound over Q-subsets gives $F_{Q,\varepsilon}(t) \le \binom{t}{Q}\varepsilon^Q$. Part (i) follows from Lemma B.21. For (ii), use $J_\varepsilon(B, M) \ge J_0(B, M)$ and apply (i) to the pooled system. For (iii), both favourable and poor completion probabilities tend to one when $\varepsilon > 0$, while completion still requires at least Q tests. □

Part (ii) preserves an advantage at a fixed finite budget for sufficiently small ε. It does not transfer the permanent-failure surrogate's asymptotic success ceiling or optimal-state-count scaling to every positive ε.

Proof of Proposition 3.4. Given ω, all states have failure probability $S(\omega, t)$ and independent trajectory randomness. A shared draw therefore gives joint failure probability $\mathbb{E}_\nu[S(\omega, t)^M]$. Independent draws give $(\mathbb{E}_\nu S(\omega, t))^M$. Strict convexity of $x \mapsto x^M$ on $[0, 1]$ gives the inequality and equality condition. For $M = 2$, the difference is the variance. □

Corollary B.23 (Attribution to diversity). *Compare a pooled state with the same seeds, $\omega_0 \sim \nu$, and budget $B = \sum_i B_i$ against states with local budgets B_i and the same conditional discovery-time law. If $T \mid \omega$ is IFR for ν-almost every ω, then*

$$J_M^{\mathrm{sh}} \le J_1, \qquad J_M^{\mathrm{div}} - J_1 \le J_M^{\mathrm{div}} - J_M^{\mathrm{sh}}.$$

Proof. Conditionally on ω, Proposition B.18 gives $\prod_i S(\omega, B_i) \ge S(\omega, B)$. Average over ν. □

B.6.1 Different seeds per state

With different seeds, write $S_i(\omega, t)$ for the state-specific failure probability. Then

$$1 - J_M^{\mathrm{sh}} = \mathbb{E}_\nu \left[\prod_i S_i(\omega, t) \right], \qquad 1 - J_M^{\mathrm{div}} = \prod_i \mathbb{E}_\nu[S_i(\omega, t)].$$

The diversity ordering holds iff the first expression is at least the second; it is not automatic.

Proposition B.24 (Diversity gain under association). *Let $\omega = (\omega^{(1)}, \ldots, \omega^{(d)})$ have independent real-valued coordinates under ν. Suppose every $S_i(\cdot, t)$ is coordinatewise non-decreasing, or every one is coordinatewise non-increasing. Then*

$$\mathbb{E}_\nu \left[\prod_i S_i(\omega, t) \right] \ge \prod_i \mathbb{E}_\nu[S_i(\omega, t)],$$

and $J_M^{\mathrm{div}} \ge J_M^{\mathrm{sh}}$.

Proof. Independent coordinates are associated (Esary et al., 1967). Products of non-negative coordinatewise non-decreasing functions remain non-negative and coordinatewise non-decreasing. Apply the association inequality inductively. The non-increasing case follows by replacing ω with $-\omega$. □

B.7 Allocation, Restarts, and Convergence

Integer budgets. Let $q = \lfloor B/M \rfloor$ and $\ell = B - Mq$. An equal split gives ℓ states $q+1$ tests and $M - \ell$ states q tests, assigned as in Section B.1. For symmetric states,

$$J(B, M) = 1 - \exp\{-H(B, M)\}, \qquad H(B, M) = (M - \ell)\Psi_M(q) + \ell\Psi_M(q+1).$$

Thus $M^\star(B) \in \arg\max_M H(B, M)$. Pilot estimates of Ψ_M yield a corresponding plug-in allocation rule, although true discoveries may not be directly observable in deployment. When seed assignments make states asymmetric, the individual hazards must be used instead.

Averaging over initialisations. For shared random components of $\mathcal{I}$, ex-ante success is

$$1 - \mathbb{E}_{\mathcal{I}}\left[\exp\left\{-\sum_i \Psi_i(B_i \mid \mathcal{I})\right\}\right].$$

It is not generally obtained by maximizing an expected sum of conditional hazards. For independently drawn initialisations, the product formula can instead be applied to marginal local survival functions conditional on the world.

Proposition B.25 (Restart criterion)**.** *Suppose the common local law does not depend on the architecture, so $\Psi_M \equiv \Psi$. For M dividing B and $t = B/M$,*

$$J(B, M) = 1 - \exp\{-B\Psi(t)/t\}.$$

(i) If the law is IFR, $\Psi(t)/t$ is non-decreasing and $M = 1$ is optimal. (ii) If $\sup_t \Psi(t) < \infty$ and $\Psi \not\equiv 0$, then $\Psi(t)/t$ attains a positive maximum at some finite $t^\star$, not necessarily unique.

Proof. Under IFR, the increments of Ψ are non-decreasing, so their running average $\Psi(t)/t$ is non-decreasing. For (ii), boundedness gives $\Psi(t)/t \to 0$, while the ratio is positive somewhere. Its maximum is therefore attained on a finite set. □

This is a fixed-budget success-probability analogue of the fixed-cutoff restart problem of Luby et al. (1993). Their expected-runtime criterion is $\mathbb{E}[\min(T, t)]/F(t)$ and generally selects a different cutoff. Neither uniqueness of $t^\star$ nor a rise-then-fall hazard shape is assumed.

Proposition B.26 (Linear growth of the preferred state count)**.** *In Proposition 3.3, fix $\alpha \in (0, 1)$, $p \in (0, 1)$, and $Q \in \mathbb{N}$. Under equal integer budget splits, there are constants $0 < c_1 \le c_2$ depending only on these parameters such that*

$$c_1 B \le M^\star(B) \le c_2 B$$

for all sufficiently large B.

Proof. Let $\varphi(t) = -\log(1 - \alpha F_{Q,p}(t))$. It is non-decreasing, vanishes for $t < Q$, and is bounded above by $-\log(1 - \alpha)$. Hence $c^\star = \max_{t \ge 1} \varphi(t)/t > 0$ is attained at some finite $t^\star \ge Q$. Choose $T_0 \ge t^\star$ such that $\varphi(t)/t \le c^\star/2$ for every $t \ge T_0$.

For $M_1 = \lfloor B/t^\star \rfloor$, every local budget is at least $t^\star$, so

$$\sum_i \varphi(B_i) \ge M_1 \varphi(t^\star) \ge c^\star(B - t^\star).$$

If $\lfloor B/M \rfloor \ge T_0$, every local budget is at least T_0 and the total hazard is at most $(c^\star/2)B$. For $B > 2t^\star$, this is below the displayed achievable value. Thus $\lfloor B/M^\star \rfloor < T_0$, implying a positive linear lower bound on $M^\star$. The admissibility condition $M^\star \le B$ gives the linear upper bound. □

Proposition B.27 (Convergence of true discoveries)**.** *Assume $\mathcal{H}$ is finite, candidate sets are retained, $\mu_h \ne \tau$, and each state's rule explores. Let Reach_i be the hypotheses connected to its seeds. Almost surely, every $h \in \mathrm{Reach}_i$ eventually becomes exposed, and every true such hypothesis is eventually accepted. For fixed M and initialisation,*

$$\lim_{B \to \infty} J(B, M) = \left|\left\{h \notin \mathcal{R}^+_{\mathrm{known}} : Z_h = 1,\ h \in \textstyle\bigcup_i \mathrm{Reach}_i\right\}\right|.$$

In the bridge world, this limit is one for every fixed M. For false cumulative acceptances,

$$\lim_{B\to\infty} \mathbb{E}[D_B^-] = \sum_{\substack{h\notin\mathcal{R}^+_{\text{known}}\\ Z_h=0}} \left[1-\prod_i (1-\pi_{i,h})\right], \qquad \pi_{i,h} = \Pr\left(h \in \bigcup_t \mathcal{R}^+_{i,t}\right).$$

Proof. Every exposed hypothesis is tested infinitely often and, by consistency, eventually resolved in the correct direction. Its neighbours are then exposed. Induction along finite graph paths gives exposure of every reachable hypothesis. True reachable hypotheses eventually enter the cumulative accepted set, regardless of earlier rejections. For fixed M, every local budget tends to infinity with B, and bounded convergence gives the limit. The bridge graph is connected and its only unknown true hypothesis is $H^\star$. For false acceptances, apply conditional independence across states to each hypothesis's eventual historical acceptance event and sum. □

B.8 Final-Posterior Recovery and Its Relation to the Theory

The simulations retain the same testing, belief-updating, and exposure dynamics but evaluate acceptance at the final local budget. For a non-seed hypothesis, define

$$A^{\text{fin}}_{i,t}(h) = \mathbf{1}\{n_{i,t}(h) \geq n_{\min},\ b_{i,t}(h) \geq \eta_+\}.$$

The final-posterior objective is

$$D^{\text{fin}}_B = \sum_{h\notin\mathcal{R}^+_{\text{known}}} Z_h\, \mathbf{1}\left\{\exists i:\ A^{\text{fin}}_{i,B_i}(h)=1\right\}, \qquad J^{\text{fin}}(B,M) = \mathbb{E}[D^{\text{fin}}_B].$$

False final-posterior acceptances replace Z_h by $1-Z_h$.

Proposition B.28 (Checkpoint evaluation)**.** *Under the same trajectories:*

(i) $D^{\text{fin}}_B \leq D_B$ *pathwise, and likewise for false acceptances.*

(ii) Under uniform selection, pooled final-posterior success is positive for every $B \geq (K-1)m$*. Together with (i), this gives the final-posterior form of Proposition 3.1.*

(iii) Define

$$G_i(t) = \Pr(A^{\text{fin}}_{i,t}(H^\star)=1).$$

Conditional independence gives

$$J^{\text{fin}}_{\text{iso}}(B,M) = 1-\prod_i [1-G_i(B_i)].$$

(iv) For matched seeds, rules, and local budgets, Proposition 3.4 holds with $S^{\text{fin}}(\omega,t) = 1-G(\omega,t)$ *in place of* $S(\omega,t)$*. In particular, independent prior draws weakly improve final-posterior recovery over a shared draw.*

(v) Under the exploration assumptions of Proposition B.27, final-posterior true recovery has the same asymptotic limit as cumulative recovery for fixed M*, while expected false final-posterior acceptances converge to zero.*

Proof. For (i), a non-seed hypothesis satisfying the acceptance criterion at time t has crossed that criterion by time t. For (ii), at any admissible B, spend the first $B-(K-1)m$ tests on a seed, then follow the canonical-path event used in the fragmentation proof. At the final checkpoint, the target has just received its m accepting observations. This event has positive probability.

For (iii), system failure at the checkpoint is the intersection of independent local failure events. For (iv), repeat the shared-versus-independent draw calculation:

$$J^{\text{fin,sh}}_M = 1-\mathbb{E}_\nu[(S^{\text{fin}})^M], \qquad J^{\text{fin,div}}_M = 1-(\mathbb{E}_\nu S^{\text{fin}})^M.$$

Jensen's inequality gives the ordering and the same equality and variance statements.

For (v), every reachable hypothesis is eventually tested infinitely often. Its posterior converges to its truth value, so its checkpoint acceptance indicator converges to one if true and zero if false. Unreachable non-seed hypotheses have no observations and cannot count. The hypothesis universe is finite, so bounded convergence completes the proof. □

What does not transfer automatically. Unlike first-acceptance probability, $G_i(t)$ need not be non-decreasing in t. It therefore need not be the distribution function of a discovery time. Although $-\log[1 - G_i(t)]$ still provides an algebraic product transformation at each checkpoint, the IFR and first-passage arguments do not automatically apply to it. Likewise, $J^{\mathrm{fin}} \leq J$ for each architecture does not imply that an ordering between architectures under J is preserved under J^{fin}.

For the bridge target, define

$$\Delta_M(B) = J(B, M) - J^{\mathrm{fin}}(B, M) \geq 0.$$

Then

$$J^{\mathrm{fin}}(B, M) - J^{\mathrm{fin}}(B, 1) = J(B, M) - J(B, 1) - \Delta_M(B) + \Delta_1(B).$$

Thus a cumulative advantage transfers whenever $J(B, M) - J(B, 1) > \Delta_M(B)$, but no such finite-budget discrepancy bound is assumed for the simulations.

Interpretation of the surrogate and experiments. In the completion-time surrogates, completion is absorbing, so completion by the checkpoint and historical completion are the same event. Their crossover results are exact for those processes. The Gaussian simulations instead test whether the proposed tradeoff persists under adaptive posterior dynamics and the more demanding requirement of final-posterior support. An observed crossover under that criterion is empirical evidence for the mechanism, not a consequence of the pointwise inequality between the two metrics. The matched-seed prior-diversity comparison directly tests an inequality that is valid under either metric.

C Additional Experimental Details: SciGym

Reporting conventions. Uncertainty terms denote sample standard deviations across algorithmic runs. For SciGym, systems are averaged within each seed before computing SD. CausalBench paired intervals are two-sided 95% Student-t intervals over matched seeds.

C.1 Structural-search protocol

Task and representation. SciGym (Duan et al., 2025) supplies hidden reaction systems, a simulator, and a sequential experiment loop. We retain the simulator and structural recovery metric, but replace the native free-text interface with explicit search over directed reaction networks. This makes the comparison a controlled test of search organization within the simulated environment. We allow at most two reactants and two products per reaction. Of the 137 systems, 90 satisfy this representation: 65 have fewer than four actionable species and 25 have at least four.

Search conditions. Each epistemic state maintains a beam of candidate directed reaction networks. A candidate is expanded through the method-defined one-edit neighbourhood, and subsequent experiments are selected from the current beam. The pooled condition uses one beam initialized with four candidates ($M = 1$). The private condition uses four beams initialized with one candidate each ($M = 4$). The comparison matches aggregate initialization, total beam capacity, charged simulator budget, and output capacity. Total beam capacity is 16, with one beam of width 16 for pooled search, or four beams of width four for private search.

Evaluation. We evaluate budgets $B \in \{0, 4, 8, 12, 16, 20, 24\}$. Initialization is complete at $B = 0$; thus B counts subsequent sequential experiments. Repeated interventions are cached and do not incur a second simulator charge. The primary metric is directed reactant/product (RP) F_1, averaged over the 65 smaller or 25 larger systems for each paired seed. Differences are computed before rounding and reported in percentage points as $F_1(M = 4) - F_1(M = 1)$. Other metrics include Symmetric Mean Absolute Percentage Error (SMAPE), which measures trajectory error, and reactant/product/modifier (RPM) F_1.

C.2 Budget-dependent structural recovery

Budget crossover. Pooled search leads at small budgets, whereas private search continues improving and is higher at the largest evaluated budget in both size strata (Table 4). The mean difference becomes positive earlier in larger systems. At $B = 24$, the private-minus-pooled difference is $+1.03$ points in smaller systems and $+3.73$ points in larger systems.

	Smaller systems ($n = 65$)			Larger systems ($n = 25$)		
B	$M = 1$	$M = 4$	Δ (pp)	$M = 1$	$M = 4$	Δ (pp)
0	$.417 \pm .008$	$.398 \pm .020$	-1.89	$.216 \pm .020$	$.197 \pm .020$	-1.85
4	$.428 \pm .006$	$.413 \pm .021$	-1.43	$.230 \pm .030$	$.206 \pm .015$	-2.31
8	$.438 \pm .007$	$.421 \pm .014$	-1.73	$.242 \pm .023$	$.230 \pm .017$	-1.21
12	$.443 \pm .008$	$.432 \pm .013$	-1.07	$.240 \pm .023$	$.247 \pm .014$	$+0.74$
16	$.446 \pm .009$	$.438 \pm .014$	-0.82	$.245 \pm .017$	$.265 \pm .019$	$+1.98$
20	$.445 \pm .008$	$.445 \pm .012$	-0.01	$.242 \pm .017$	$.271 \pm .017$	$+2.89$
24	$.443 \pm .007$	$.453 \pm .009$	$+1.03$	$.250 \pm .010$	$.288 \pm .017$	$+3.73$

Table 4: SciGym structural recovery across sequential-experiment budgets. Recovery entries are mean $\pm$ sample SD across ten seed-level means. Δ is private-minus-pooled recovery in F_1 percentage points. Budget zero follows initialization.

Condition	SMAPE $\downarrow$	RP $F_1 \uparrow$	RPM $F_1 \uparrow$
Single agent	$.391 \pm .016$	$.400 \pm .028$	$.219 \pm .015$
Reconciled pair, same context	$.400 \pm .018$	$.391 \pm .013$	$.218 \pm .010$
Reconciled pair, different context	$\mathbf{.342} \pm .024$	$\mathbf{.417} \pm .014$	$\mathbf{.231} \pm .016$

Table 5: Native SciGym extension over 137 systems and 20 matched seeds. Values are mean $\pm$ sample SD across seed-level means. SMAPE evaluates simulated trajectories; RP evaluates reactants/products, and RPM additionally includes modifiers. Both pair conditions communicate and submit one reconciled model.

C.3 Native LLM-agent extension

Task and information conditions. We run SciGym's native free-text experimentation loop over all 137 systems with 20 matched API seeds. Agents use `gpt-5.6-luna` with `medium` reasoning effort, propose runnable reaction models in SBML, and use the simulator to test experiments before submitting a final model. The single agent receives the supplied partial model and an allowance of 20 ReAct loops. Each pair contains two agents, each allowed 10 loops and retaining its own experimental history. In the same-context condition, both agents receive the full supplied context. In the different-context condition, one agent receives the full context and the other receives a redacted context. For the latter, species initial amounts/concentrations and constancy flags are removed from both the initial prompt and the exposed partial-model string; the underlying experimental simulator is unchanged. Neither member receives additional initial information unavailable to the single agent.

Results and interpretation. Different-context pairs outperform same-context pairs on all three mean metrics (Table 5). SMAPE decreases by .057 (95% CI $[.035, .080]$), and RP F_1 increases by .027 ($[.015, .038]$). These results are exploratory: both pair conditions allow communication and submit a single reconciled model, unlike the isolated states studied above. Moreover, context redaction changes the information available to one agent, so the comparison does not isolate prior diversity. The results suggest that heterogeneous initial contexts can benefit collaborative search, but do not directly test the isolation mechanism of `ISLAND`.

D Additional Experimental Details: CausalBench

D.1 Datasets and evaluation

Datasets. We use the filtered K562 and RPE1 datasets from CausalBench (Chevalley et al., 2025), with the benchmark preprocessing and evaluator. Table 6 summarizes their sizes.

Metrics. For a predicted edge set G_p, the primary metric is the benchmark mean Wasserstein score,

$$\mathrm{W}_1(G_p) = \frac{1}{|G_p|} \sum_{(i,j) \in G_p} W_1\Big(P^{\mathrm{do}(i)}(X_j), P^{\mathrm{obs}}(X_j)\Big). \tag{6}$$

Dataset	Genes	Directed edges	Observational cells	Interventional cells
K562	622	386,262	8,553	121,647
RPE1	383	146,306	9,188	120,998

Table 6: Filtered CausalBench datasets. Directed-edge counts include all ordered non-self gene pairs. Observation-only searches use only the observational expression matrix; perturbation data are used by the benchmark evaluator.

Higher values indicate stronger intervention-induced distributional shifts among predicted source-target pairs. We also report biological precision,

$$\mathrm{BioPrec}(G_p) = \frac{\mathrm{TP}_{\mathrm{bio}}(G_p)}{\mathrm{TP}_{\mathrm{bio}}(G_p) + \mathrm{FP}_{\mathrm{bio}}(G_p)}. \tag{7}$$

D.2 Observation-only search protocol

State representation. Let V be the full gene universe and $\mathcal{H} = \{(i,j) : i,j \in V,\ i \neq j\}$ the directed hypothesis space. Each of $M \in \{1,2,4,8,16\}$ states maintains a seed set S_a, candidate set C_a, transient visible set $U_a^{(t)}$, observed universe O_a, and accumulated edge scores. They satisfy $S_a \subseteq C_a \subseteq O_a \subseteq V$. The observation-only runs sample 32 seed genes uniformly and partition them across states. At each update, a state retains its current candidate genes and fills the visible subspace with randomly sampled outside scouts. Its size is

$$|U_a^{(t)}| = \min\{|V|, \max\{64, |C_a^{(t)}| + 32\}\}. \tag{8}$$

Each update also bootstraps N cells with replacement from the shared N-cell observational matrix.

Within the visible subspace, columns are centered and scaled, a ridge-regularized covariance is formed with $\lambda = .01$, and its pseudoinverse gives $\widehat{\Omega}$. The directed local score is

$$\beta_{ij}^{(a,t)} = \left| -\frac{\widehat{\Omega}_{ij}^{(a,t)}}{\widehat{\Omega}_{jj}^{(a,t)}} \right|, \qquad \beta_{ii}^{(a,t)} = 0. \tag{9}$$

Evidence accumulation and frontier expansion. For each jointly visible pair, the state stores a support count $n_{ij}^{(a)}$ and accumulated score $s_{ij}^{(a)}$. Its mean belief is

$$\mu_{ij}^{(a)} = \frac{s_{ij}^{(a)}}{\max\{n_{ij}^{(a)}, 1\}}. \tag{10}$$

After repeated updates, genes with strong accumulated cross-boundary signal can be promoted from the observed universe into the candidate set. Expansion occurs every ten local updates, adding up to ten genes with a .2 random-expansion fraction and minimum support of two for evidence-based promotion.

Consensus and ranking. States remain separate during search. At evaluation, their scores are pooled by support-weighted averaging,

$$\bar{\mu}_{ij} = \frac{\sum_a n_{ij}^{(a)} \mu_{ij}^{(a)}}{\sum_a n_{ij}^{(a)}}, \tag{11}$$

for pairs with nonzero total support. Unobserved pairs have score zero. Self-edges are excluded before truncating the eligible ranking, with stable descending sorting. The top-ranked edges are passed to the benchmark evaluator.

D.3 K562 observation-only results

Multi-state recovery. At 640 updates, $M = 1$ obtains mean W_1 / BioPrec of .182/.424, while $M = 16$ obtains .209/.550 (Table 7). The paired $M = 16$ minus $M = 1$ changes are $+.027$ in W_1

Method	Mean W_1 ↑	BioPrec ↑
GRNBoost (Moerman et al., 2019)	0.164 ± 0.007	0.276 ± 0.012
DCDI-DSF (Brouillard et al., 2020)	0.164 ± 0.001	0.252 ± 0.010
DCDI-G (Brouillard et al., 2020)	0.170 ± 0.004	0.296 ± 0.008
ISLAND, $M = 1$	0.182 ± 0.001	0.424 ± 0.002
ISLAND, $M = 2$	0.180 ± 0.001	0.409 ± 0.005
ISLAND, $M = 4$	0.178 ± 0.002	0.384 ± 0.007
ISLAND, $M = 8$	0.186 ± 0.005	0.399 ± 0.022
ISLAND, $M = 16$	$\mathbf{0.209 \pm 0.006}$	$\mathbf{0.550 \pm 0.031}$

Table 7: Final top-1,000 observation-only results on filtered K562. The ISLAND rows use the precision-based score at 640 total updates over 20 initialization seeds; external baselines are ranked at the same depth. Values are mean ± sample SD.

Method	Mean W_1 ↑	BioPrec ↑
GRNBoost (Moerman et al., 2019)	0.117 ± 0.004	0.097 ± 0.006
DCDI-DSF (Brouillard et al., 2020)	$\mathbf{0.163 \pm 0.006}$	0.092 ± 0.001
DCDI-G (Brouillard et al., 2020)	0.158 ± 0.008	0.096 ± 0.011
ISLAND, $M = 1$	0.140 ± 0.001	0.146 ± 0.001
ISLAND, $M = 2$	0.144 ± 0.001	0.150 ± 0.002
ISLAND, $M = 4$	0.148 ± 0.001	0.154 ± 0.002
ISLAND, $M = 8$	0.152 ± 0.001	0.154 ± 0.004
ISLAND, $M = 16$	$\mathbf{0.162 \pm 0.002}$	$\mathbf{0.174 \pm 0.005}$

Table 8: RPE1 observation-only replication at 640 total updates. Results evaluate the top 1,000 unique directed non-self edges and report mean ± sample SD over 20 initialization seeds.

(95% CI $[.024, .029]$) and $+.126$ in BioPrec ($[.112, .140]$), and both metrics improve in all 20 paired runs. The benefit is not monotone in M: at this budget, $M = 2$ and $M = 4$ fall below $M = 1$ on both metrics, and $M = 8$ is higher on W_1 (.186) but lower on BioPrec (.399). This follows from how the update budget interacts with frontier expansion. Each state promotes at most ten genes every ten local updates (App. D.2), so with 640 total updates a state can add at most $640/M$ genes to its candidate set: 320 for $M = 2$, 160 for $M = 4$, 80 for $M = 8$, and 40 for $M = 16$. For $M = 1$ this bound exceeds the 622-gene universe, and its endpoint matches a single state given every gene from the start (Table 9). Intermediate M therefore have room to drift towards this global estimator by 640 updates, whereas the $M = 16$ states remain small and evidence-driven. The budget curves in Fig. 4B show the same pattern: each curve rises and then declines as its states expand, and larger M delay the decline. We therefore read the endpoint result as evidence that persistent local trajectories help while they remain local, rather than that more states are uniformly better.

D.4 RPE1 REPLICATION

RPE1 provides a cross-dataset observation-only check. At the final 640-update checkpoint, $M = 1$ obtains $.140/.146$ and $M = 16$ obtains $.162/.174$ on mean W_1 / BioPrec across 20 paired seeds. The paired changes are $+.022$ in W_1 and $+.028$ in BioPrec, positive in all 20 runs. At $M = 16$, the performance is comparable to that of the best performing method in W_1 (DCDI-DSF), with substantially higher biological precision. The endpoint advantage therefore extends beyond K562, although its magnitude and budget trajectory differ between datasets.

D.5 SEARCH-SPACE CONTROLS

We use two K562 controls to examine whether the observation-only gains can be explained by random subspace sampling or restricted single-state initialization. Both use the same precision-based scoring pipeline, 640 total updates, and 20 initialization seeds.

Random subspaces. This control samples random 16-gene subspaces at each update instead of retaining and expanding a candidate frontier. Increasing the number of states from $M = 1$ to

Search organization	M	Mean $W_1 \uparrow$	BioPrec $\uparrow$
Random subspaces	1	0.171 ± 0.002	0.317 ± 0.016
Random subspaces	16	0.175 ± 0.004	0.329 ± 0.007
Global from start	1	0.182 ± 0.001	0.426 ± 0.002
ISLAND	1	0.182 ± 0.001	0.424 ± 0.002
ISLAND	16	$\mathbf{0.209 \pm 0.006}$	$\mathbf{0.550 \pm 0.031}$

Table 9: K562 search-space controls at 640 total updates and 1,000 predicted edges. All rows use 20 initialization seeds. Control rows report mean $\pm$ sample SD; ISLAND means are repeated for reference. Random subspaces contain 16 genes per update; global initialization exposes all 622 genes from the start.

$M = 16$ changes mean W_1 from .171 to .175 and BioPrec from .317 to .329. These gains are substantially smaller than those under ISLAND, and both endpoints remain below the adaptive-search results (Table 9), thus supporting the adaptive-search design as a whole.

Global initialization. As a base-method control, global-from-start $M = 1$ exposes all 622 genes immediately while retaining the same data, ridge, update budget, and evaluator. It obtains mean W_1 / BioPrec of $.182/.426$, compared with $.182/.424$ for the expanding-frontier $M = 1$ condition and $.209/.550$ for $M = 16$. The near equality of the two single-state endpoints indicates that the multi-state advantage is not simply a consequence of restricting the single-state search's initial gene access. Together, the controls support the value of the multi-state adaptive organization, while leaving localization, persistent separation, and final consensus as jointly varying components.

D.6 Intervention selection and network evaluation

Intervention protocol. We explore whether selecting interventions within persistent, evolving gene subspaces improves network recovery. The intervention stage starts from 160 total observational updates and 32 seed genes, as in the observation-only setup. Each method selects 80 unique source genes in five batches of 16, with intervention outcomes retrospectively acquired from the K562 benchmark's recorded perturbation data. The intervention budget is shared across states rather than allocated separately to each state.

Within ISLAND, each state ranks legal intervention sources in its current candidate set by accumulated outgoing edge mass,

$$R_i^{(a)} = \sum_{j \in O_a,\, j \neq i} |\mu_{ij}^{(a)}|. \tag{12}$$

Source self-edges are excluded before ranking. Sources are allocated across states in round-robin order, without repeated source selection. For a selected source i, target evidence is obtained from the absolute mean differential expression,

$$\Delta_{ij} = \left| \mathbb{E}[X_j^{\mathrm{do}(i)}] - \mathbb{E}[X_j^{\mathrm{obs}}] \right|. \tag{13}$$

Method	W1@1,000	BioPrec@1,000	W1@5,000	BioPrec@5,000
Random	0.394 ± 0.026	0.456 ± 0.067	0.245 ± 0.014	0.281 ± 0.033
BALD (Houlsby et al., 2011)	0.417 ± 0.012	0.304 ± 0.019	0.254 ± 0.008	0.211 ± 0.012
CBED (Tigas et al., 2022)	0.440 ± 0.014	0.720 ± 0.045	0.279 ± 0.010	0.485 ± 0.040
DiscoBAX-GP (Lyle et al., 2023)	0.441 ± 0.020	0.675 ± 0.057	0.282 ± 0.010	0.479 ± 0.025
ISLAND, $M = 1$	0.456 ± 0.007	0.779 ± 0.010	0.299 ± 0.006	0.582 ± 0.007
ISLAND, $M = 4$	$\mathbf{0.463 \pm 0.004}$	$\mathbf{0.793 \pm 0.003}$	$\mathbf{0.303 \pm 0.002}$	$\mathbf{0.592 \pm 0.002}$

Table 10: **K562 intervention-target selection at a matched budget.** Selected-source proxy results after five batches of 16 interventions (80 unique source genes), with source self-edges excluded before ranking. Methods are evaluated at exactly 1,000 or 5,000 unique non-self edges. Values are mean $\pm$ sample SD; higher is better for both metrics.

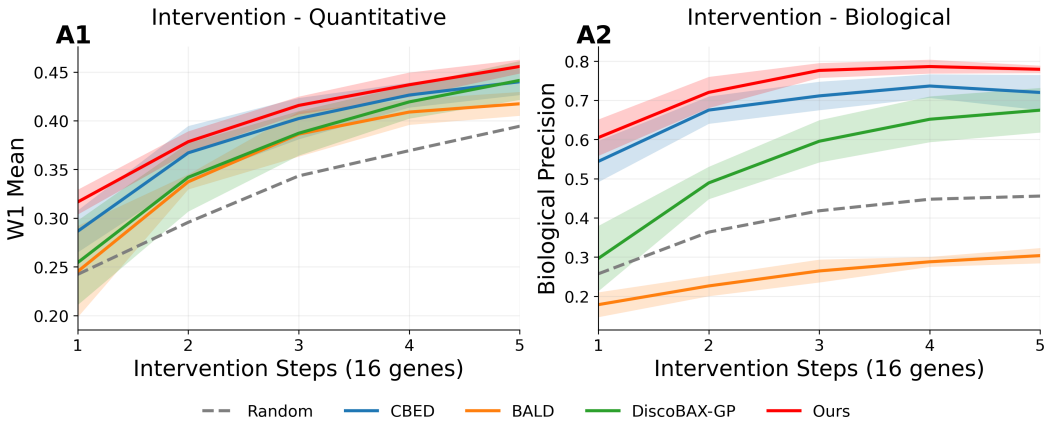


Figure 5: **Exploratory intervention-target selection on K562.** (**A1**) Mean Wasserstein score and (**A2**) biological precision of the common selected-source proxy, evaluated at the top 1,000 non-self edges. Each step adds 16 unique source-gene interventions, reaching 80 after five steps. All methods use the same intervention budget. `ISLAND` (shown here with $M{=}1$) achieves higher endpoint mean scores than Random, BALD, CBED, and the adapted DiscoBAX-GP baseline.

Each batch proceeds through source ranking, round-robin allocation, private evidence updates, and candidate promotion, followed by evaluation-only consensus. Thus, intervention evidence can change both subsequent source rankings and the local candidate frontier.

Matched selection evaluation and baselines. Our primary comparison concerns experiment-target selection, not differences between methods' native network estimators. All methods receive the same source-intervention budget and are evaluated through a common selected-source proxy constructed from their selected genes' perturbation data. We report mean Wasserstein score and biological precision at fixed numbers of unique non-self edges. This proxy measures the network evidence recovered through source selection; it is distinct from evaluating each method's internal belief-state network.

We compare against Random, BALD (Houlsby et al., 2011), CBED (Tigas et al., 2022), and an adapted DiscoBAX-GP (Lyle et al., 2023). Our DiscoBAX-GP implementation uses an observational gene embedding and a scalar transcriptomic-response phenotype, with 16 initial random selections followed by four adaptive batches of 16. This matches the total budget of 80 source interventions but is not an exact reproduction of the original genomic task.

Comparison with selection baselines. At 80 interventions, both `ISLAND` variants achieve higher mean scores than the comparison methods at both evaluation depths (Table 10). At 1,000 edges, $M{=}1$ reaches Wasserstein 0.456 and biological precision 0.779, compared with 0.440 and 0.720 for CBED and 0.441 and 0.675 for DiscoBAX-GP. The baseline trajectories also show that quantitative and biological recovery need not improve together: BALD exceeds Random in Wasserstein score but remains below it in biological precision (Fig. 5).

Single-state versus multi-state selection. The effect of search organisation is smaller than in observation-only recovery (Fig. 6). For Wasserstein score, $M{=}1$ leads over the first three batches, while $M{=}4$ overtakes it at the fourth batch and finishes at 0.463 versus 0.456. Biological precision follows a different trajectory: $M{=}4$ leads from the second batch onward and plateaus near 0.79, finishing at 0.793 versus 0.779 for $M{=}1$. At 5,000 edges, the corresponding endpoint scores are 0.303 versus 0.299 in Wasserstein score and 0.592 versus 0.582 in biological precision.

These descriptive results are consistent with a modest benefit from distributed intervention search, including a qualitative crossover in Wasserstein score. However, they do not establish a general or statistically resolved advantage from retaining multiple states during intervention selection. Because each variant enters the intervention phase with its own observational search history, this comparison

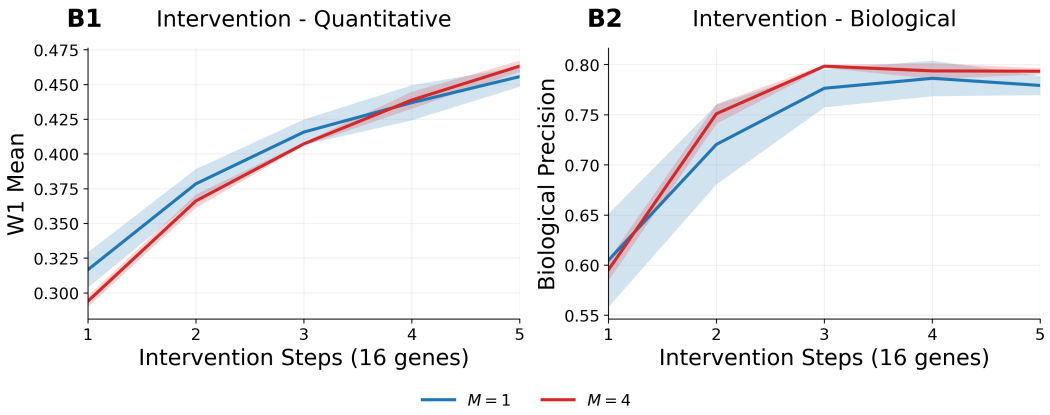


Figure 6: **Search multiplicity during intervention-target selection.** (**B1**) Mean Wasserstein score and (**B2**) biological precision of the selected-source proxy at the top 1,000 non-self edges, comparing $M{=}1$ with $M{=}4$. Both variants start from 160 total observational updates and share a total intervention budget of five batches of 16 unique source genes. Wasserstein score shows an early single-state advantage followed by a modest multi-state advantage; biological precision favours $M{=}4$ from the second batch onward and subsequently plateaus.

also does not isolate intervention-stage multiplicity from the states inherited from observational search.

E Synthetic Simulations: Details and Extended Results

This appendix describes the simulator, evaluation protocol, and experiments supporting Section 4. We distinguish individual acceptance, acceptance by any state, and aggregate-score positivity. These are different outcomes and are not used interchangeably.

E.1 Simulator and experimental protocol

Hypotheses and observations. A world consists of variables $\mathcal{V}$, queryable hypotheses $\mathcal{H}$, and a fixed latent effect μ_h for each hypothesis. As in Section 2, the truth label is

$$Z_h = \mathbf{1}\{\mu_h > \tau\}.$$

We write

$$\mathcal{H}_{\text{true}} = \{h \in \mathcal{H} : Z_h = 1\}$$

for the true-hypothesis set, reserving $H^\star$ for the bridge target. The implementation's effect threshold τ_μ is denoted τ throughout.

Each query by state i returns

$$Y \mid h \sim \mathcal{N}(\mu_h, \sigma^2),$$

using that state's random stream. True hypotheses have effects given by their factor strengths. False hypotheses receive latent effects drawn uniformly from $[\tau - \delta, \tau)$, fixed for the world and shared across compared methods.

The principal bridge experiments use

$$(\tau, \delta, \mu_{\text{true}}, \sigma) = (0.7, 0.1, 0.8, 0.22).$$

Larger-world defaults are

$$(\tau, \delta, \sigma) = (0.6, 0.5, 0.22),$$

with true strengths drawn uniformly from $[\tau, 1]$ unless otherwise specified. Under this continuous distribution, equality with τ has probability zero. Here σ is the observation standard deviation.

K-bridge construction. For the principal bridge experiments,

$$\mathcal{V} = \{v_1, \dots, v_K, v'_1, \dots, v'_K, w_1, \dots, w_X\}, \qquad n = 2K + X,$$

where X denotes the number of noise variables (distractors in the search space), and

$$\mathcal{H} = \{h \subseteq \mathcal{V} : 2 \le |h| \le K\}.$$

The true hypotheses are the known seed pairs

$$s_k = \{v_k, v'_k\}$$

and the target

$$H^\star = \{v_1, \dots, v_K\}.$$

There is therefore one non-seed target, and

$$|\mathcal{H}| = \sum_{r=2}^{K} \binom{n}{r}.$$

Increasing X adds distractor hypotheses without changing observation variance or adding true targets.

Hypothesis adjacency. We use the edit-distance-one graph from Eq. 1. For scopes $\mathcal{S}_h$ and $\mathcal{S}_g$,

$$d_{\text{edit}}(\mathcal{S}_h, \mathcal{S}_g) = \max\{|\mathcal{S}_h \setminus \mathcal{S}_g|, |\mathcal{S}_g \setminus \mathcal{S}_h|\},$$

and

$$\Gamma(h) = \{g \in \mathcal{H} : d_{\text{edit}}(\mathcal{S}_h, \mathcal{S}_g) = 1\}.$$

Thus a neighbour is obtained by exactly one of:

- adding one variable not already in the scope;
- deleting one variable from the scope;
- replacing one variable in the scope with one outside it.

Prior family. Each state uses a prior that factorises over hypotheses. A draw $\omega = (\xi, \epsilon) \sim \nu$ specifies a random directional anchor ξ and hypothesis-specific jitter:

$$\mu_{0,\omega}(h) = \tau - 2\delta_{\text{prior}} + \delta_{\text{prior}} \exp\{-d(h, \xi)/\ell\} + \epsilon_h, \qquad \kappa_{0,\omega}(h) = \kappa_{\text{prior}}.$$

The distance is

$$d(h, \xi) = \frac{1}{|h|} \sum_{v \in h} d_G(v, \xi),$$

where d_G denotes shortest-path distance in the variable graph G, with unreachable distances replaced by n. The locality parameter ℓ controls how quickly the directional preference decays.

The principal bridge simulations use $\delta_{\text{prior}} = 0.10$, $\kappa_{\text{prior}} = 10$, and $\ell = 1.5$, with independent Gaussian jitter $\epsilon_h \sim \mathcal{N}(0, 10^{-8})$. Thus the jitter standard deviation is 10^{-4}, and the prior variance of the latent effect is $\sigma^2/\kappa_{\text{prior}}$.

Initialisation and budget allocation. Seed hypotheses are selected randomly where the world construction permits; the principal bridge experiment uses its designated seed pairs. Let $\mathcal{S}_i$ be the seed set supplied to state i. Its initial candidate set is

$$\mathcal{C}_{i,0} = \mathcal{S}_i \cup \Gamma(\mathcal{S}_i), \qquad \Gamma(\mathcal{A}) = \bigcup_{h \in \mathcal{A}} \Gamma(h).$$

The pooled state receives the union of the seeds supplied to the isolated states and one randomly drawn prior from ν. In the diverse-prior condition, isolated states receive independent prior draws from the same marginal distribution.

Within each replicate, random prior draws are reused across architecture comparisons to reduce comparison variance. This coupling across compared architectures does not imply that states within a diverse-prior architecture share a prior.

For the bridge assignment with $M \le K$, seed k is assigned to state

$$1 + ((k-1) \bmod M).$$

If $M > K$, seed pairs are reused round robin.

At a total-budget checkpoint B, local budgets are

$$B_i = \left\lfloor \frac{B}{M} \right\rfloor + \mathbf{1}\{i \le B \bmod M\}, \qquad \sum_{i=1}^{M} B_i = B.$$

Budget grids are chosen to be divisible by the tested state counts where possible. In the larger-world state-count comparisons, a common collection of 16 seeds is partitioned across $M \in \{1, 2, 4, 8, 16\}$. All known seeds are excluded from evaluation, regardless of which state receives them. Isolated states exchange neither observations nor posterior beliefs during search.

Posterior updates and individual classifications. For state i, let $n_i(h)$ be the number of observations of h and $S_i(h)$ their sum. The posterior mean, variance, and threshold-exceedance probability are

$$m_i(h) = \frac{\kappa_{0,i}(h)\mu_{0,i}(h) + S_i(h)}{\kappa_{0,i}(h) + n_i(h)}, \qquad v_i(h) = \frac{\sigma^2}{\kappa_{0,i}(h) + n_i(h)},$$

$$b_i(h) = \Pr(\mu_h \ge \tau \mid \mathcal{F}_i) = \Phi\left(\frac{m_i(h) - \tau}{\sqrt{v_i(h)}}\right).$$

In the principal bridge simulations, a hypothesis is accepted when $n_i(h) \ge n_{\min} = 2$ and $b_i(h) \ge \eta_+ = 0.8$, and rejected when $n_i(h) \ge n_{\min} = 2$ and $b_i(h) \le \eta_- = 0.2$. Hypotheses that satisfy neither condition remain unresolved. Previously resolved hypotheses remain eligible for testing, and their current classification may change with subsequent observations.

Frontier expansion and persistent exposure. Expansion is recorded separately from current classification. Let $\mathcal{E}_{i,t}$ be the set of hypotheses that have been resolved at least once by local time t, including the initial seeds. After a query, define

$$\mathcal{E}_{i,t+1} = \mathcal{E}_{i,t} \cup \{h : h \text{ is accepted or rejected after the query}\}.$$

The candidate set is updated by

$$\mathcal{C}_{i,t+1} = \mathcal{C}_{i,t} \cup \Gamma(\mathcal{E}_{i,t+1} \setminus \mathcal{E}_{i,t}).$$

A generator therefore exposes its neighbours only once. Those neighbours remain accessible if its classification later changes. Repeated exposure does not create duplicate candidates.

Acquisition policies. Each policy selects from the state's current selectable candidate set.

Thompson sampling. For each candidate h, draw independently

$$\widetilde{\mu}_i(h) \sim \mathcal{N}(m_i(h), v_i(h))$$

and query the hypothesis with the largest draw.

Softmax uncertainty sampling. Let

$$u_i(h) = -b_i(h)\log b_i(h) - (1 - b_i(h))\log(1 - b_i(h)).$$

The selection probability is

$$\Pr(h \mid \mathcal{C}_i) = \frac{\exp\{\beta u_i(h)\}}{\sum_{g \in \mathcal{C}_i} \exp\{\beta u_i(g)\}}, \qquad \beta = 10.$$

This favours uncertainty about whether the effect exceeds τ.

LUCB-style threshold sampling. Define

$$r_i(h) = z\sqrt{v_i(h)}, \qquad z = 1.96.$$

Experiment	World / initialisation	Budget / states	Replication / policy
Bridge crossover (Fig. 2A)	$K = 4$; $n \in \{14, 19\}$	$B = 512{:}128{:}12672$; $M \in \{1, 4\}$	100 replicates; softmax
Bridge complexity sweep (Fig. 2B, 7)	$K = 4$; $X = 0, \ldots, 21$	$B = 128{:}128{:}8064$; $M \in \{1, 4\}$	100 replicates; all policies
Larger-world budget sweep (Fig. 2C)	$n = 50$; 50 true factors; 16 seeds; arities $\{2, 3\}$	$B = 0$–16384; $M \in \{1, 2, 4, 8, 16\}$	10 seed sets $\times$ 10 trials; Thompson
Trajectory diagnostics (Fig. 3A)	$K = 4$ bridge world	$B = 200$; $M \in \{1, 4\}$	Illustrative trajectory
Candidate size growth (Fig. 3B)	$K = 4$ bridge world	$B = 8192$; $M \in \{1, 4\}$	100 replicates; LUCB-style
Seed–prior factorial (Fig. 3C)	$K = 4$, $X = 18$; shared/distributed seeds and priors	$B = 512{:}128{:}8192$; $M = 4$	100 replicates; LUCB-style
Supporting controls			
Fixed-budget policy checks	Random-20 and block-4$\times$5 worlds; 20 true factors; 16 seeds; arities $\{2, 3\}$	$B = 4096$; $M \in \{1, 2, 4, 8, 16\}$	10 worlds $\times$ 10 seed sets; all three policies

Table 11: **Controlled-discovery experimental configurations.** Settings for the main synthetic experiments and supporting controls.

Candidates are ranked lexicographically by whether

$$[m_i(h) - r_i(h), m_i(h) + r_i(h)]$$

straddles τ, then by decreasing $r_i(h)$, then by increasing

$$|m_i(h) - \tau|/r_i(h).$$

Remaining ties follow the supplied candidate order. This is a LUCB-style heuristic for thresholding on an expanding candidate set, rather than the original fixed-arm best-arm identification algorithm, due to the dynamic nature of the expanding frontier.

Experimental configurations. Table 11 summarises the experiments supporting the main figures and extensions. The notation $a : s : b$ denotes an inclusive grid with step s.

E.2 Evaluation and statistical reporting

Individual acceptance and any-state acceptance. Let $\mathcal{A}_i(t)$ denote the hypotheses currently accepted by state i after t queries. Let $\mathcal{S}$ be the union of all known seeds in the comparison.

For the bridge world, the checkpoint acceptance indicator is

$$I_B^{\mathrm{acc}} = \mathbf{1}\left\{ H^\star \in \bigcup_{i=1}^{M} \mathcal{A}_i(B_i) \right\}.$$

Averaging this indicator across replicates estimates

$$J^{\mathrm{acc}}(B, M) = \mathbb{E}[I_B^{\mathrm{acc}}].$$

Aggregate posterior-mean score. At total budget B, define

$$N_B(h) = \sum_{i=1}^{M} n_{i,B_i}(h).$$

For hypotheses with $N_B(h) > 0$, the evaluation score is

$$g_B(h) = \frac{\sum_{i=1}^{M} n_{i,B_i}(h) m_{i,B_i}(h)}{N_B(h)}.$$

States that never tested h have zero weight. This aggregation is used only for evaluation; states do not receive the aggregate score or one another's evidence during search.

Aggregate-score positivity and TP/FP counts. A tested hypothesis is classified positive when

$$g_B(h) \geq \tau.$$

Define the non-seed positive set

$$\mathcal{P}_B = \{h \in \mathcal{H} \setminus \mathcal{S} : N_B(h) > 0,\ g_B(h) \geq \tau\}.$$

The larger-world metrics are

$$\mathrm{TP}_B = \sum_{h \in \mathcal{P}_B} Z_h, \qquad \mathrm{FP}_B = \sum_{h \in \mathcal{P}_B} (1 - Z_h).$$

Each scope contributes at most once, regardless of how many states tested it.

Bridge positivity. The bridge "positive" outcome uses the same aggregate-score rule:

$$I_B^{\mathrm{pos}} = \mathbf{1}\{N_B(H^\star) > 0,\ g_B(H^\star) \geq \tau\}.$$

This is distinct from I_B^{acc}. The bridge may be positive under the aggregate score without any state meeting the 0.8 acceptance threshold. Conversely, one state may accept the bridge while the weighted contribution of other states lowers the aggregate score below τ.

Ranking quality. We report average precision (AP) as a complementary measure of how final scores rank true hypotheses above false ones. This is the non-interpolated area under the precision–recall curve

Within each comparison and replicate, all methods are evaluated on the same universe

$$\mathcal{U} = \left(\bigcup_{\text{compared methods}} \{h : h \text{ has at least one observation}\} \right) \setminus \mathcal{S}.$$

Thus the universe is the shared observed union, not the full hypothesis space.

For each method, the ranking score is

$$q_B(h) = \begin{cases} g_B(h), & N_B(h) > 0, \\ \tau, & N_B(h) = 0. \end{cases}$$

The default score τ is used only for ranking hypotheses observed by another compared method but not by the method being evaluated. It does not make an untested hypothesis positive under the TP/FP rule.

Order $\mathcal{U} = \{h_1, \ldots, h_N\}$ by decreasing score, breaking ties by increasing arity and then lexicographic variable order. Writing

$$P = \sum_{k=1}^{N} Z_{h_k},$$

the reported metric is

$$\mathrm{AP} = \frac{1}{P} \sum_{k=1}^{N} Z_{h_k} \frac{\sum_{j=1}^{k} Z_{h_j}}{k}, \qquad P > 0,$$

with AP set to zero when $P = 0$.

False-positive reporting. For larger worlds, false positives are defined by FP_B above. For an any-state acceptance readout, the corresponding false-acceptance count is instead

$$\mathrm{FA}_B^{\mathrm{acc}} = \left| \left(\bigcup_i \mathcal{A}_i(B_i) \right) \cap (\mathcal{H} \setminus (\mathcal{H}_{\mathrm{true}} \cup \mathcal{S})) \right|.$$

Replication and variability. Compared methods share the latent world and seed collection within a replicate where applicable. States use separate random streams, and prior draws vary across replicates. Larger-world experiments additionally vary world configuration and seed allocation as specified in Table 11.

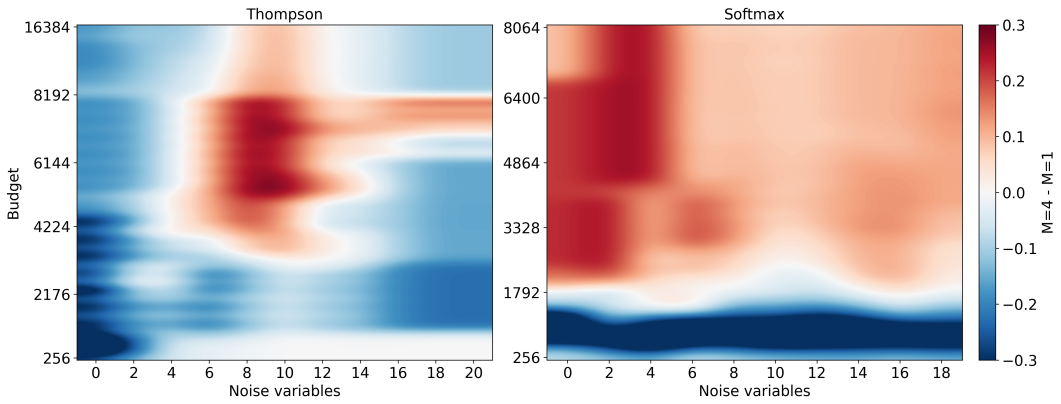


Figure 7: Crossover regimes for Thompson sampling and softmax.

E.3 Bridge experiments and policy robustness

Budget and distractor complexity. The principal bridge curves compare $M = 1$ and $M = 4$ under softmax acquisition across the budget grid in Table 11. The single state starts from all four seed pairs, while each isolated state starts from one. The LUCB-style complexity sweep additionally varies the number of noise variables. Figure 7 shows similar crossover regimes for softmax. Under Thompson sampling the multi-state advantage appears in an intermediate band of budgets rather than as a single crossover, although at higher complexity both settings had poor absolute performance.

Policy dependence. Softmax favours threshold uncertainty, Thompson sampling favours large sampled effects, and LUCB-style acquisition prioritises uncertain threshold crossings. These rules allocate queries differently even on the same exposed candidate set.

The softmax and LUCB-style bridge sweeps exhibit budget-dependent reversals. The larger-world checks in Appendix E.6 compare all three policies at a common budget. Those fixed-budget checks assess architecture rankings under different policies; they do not by themselves establish a budget-dependent crossover for each policy.

E.4 Trajectory and candidate-set diagnostics

Trajectory visualisation. Figure 3A compares one state's first 200 queries with four states' first 50 queries each. The same hypothesis has the same position across plots. The layout places bridge variables and their seed partners in corresponding local regions in order to provide an illustrative view of query allocation.

Candidate-set cardinality. At total budget B, local checkpoints are aligned using

$$t_i(B) = \left\lfloor \frac{B}{M} \right\rfloor + \mathbf{1}\{i \le B \bmod M\}.$$

Let $\mathcal{C}_{i,t}$ be state i's exposed candidate set after t queries. Panel B compares the pooled cardinality with

$$\left| \bigcup_{i=1}^{M} \mathcal{C}_{i,t_i(B)} \right|.$$

The isolated union is computed within each replicate before averaging; it is not the sum of local cardinalities. Each isolated decision is made from its own local set.

E.5 Seed allocation and prior diversity

Factorial design. The factorial crosses two seed allocations with two prior assignments, at fixed $M = 4$, total budget, acquisition rule, and prior concentration:

Seed allocation	Prior assignment	Mean bridge acceptance (95% CI)
Shared	Shared	$0.260\ (0.174,\ 0.346)$
Shared	Independent	$0.360\ (0.265,\ 0.455)$
Distributed	Shared	$0.350\ (0.256,\ 0.444)$
Distributed	Independent	$0.460\ (0.362,\ 0.558)$

Table 12: Seed/prior factorial at $B = 4096$, with $M = 4$, $K = 4$, $X = 18$, and LUCB-style acquisition.

- *Shared seeds:* every state receives the same seed collection.
- *Distributed seeds:* the same collective seed collection is partitioned across states.
- *Shared prior:* one random prior draw $\omega \sim \nu$ is reused across all states.
- *Diverse priors:* states receive independent draws $\omega_i \overset{\text{iid}}{\sim} \nu$.

Shared and diverse conditions therefore have the same marginal prior distribution. Observation and acquisition randomness is independent across states. The experiment uses $K = 4$, $X = 18$, LUCB-style acquisition, and 100 replicates over $B = 512 : 128 : 8192$.

Connection to prior hedging. For current acceptance by any state, condition on a fixed world, use identical seeds and equal local budgets, and let $S(\omega)$ be the conditional probability that one state does not accept the target at its checkpoint. With independent state-level randomness, shared-prior and diverse-prior acceptance probabilities are respectively

$$1 - \mathbb{E}_\nu[S(\omega)^M], \qquad 1 - \big(\mathbb{E}_\nu[S(\omega)]\big)^M.$$

The Jensen argument in Prop. 3.4 therefore applies to this checkpoint acceptance event. If worlds also vary, the inequality holds conditional on each world and can then be averaged over worlds.

Factorial results. Table 12 gives the checkpoint at $B = 4096$ for the budget comparison in Fig. 3C. Diverse priors increase the outcome by 0.10 under shared seeds and 0.11 under distributed seeds. Distributing seeds increases it by 0.09 under a shared prior and 0.10 under diverse priors.

E.6 Larger sparse worlds

The larger-world experiments replace the single bridge target with multiple true hypotheses of arity two or three. They report aggregate-score TP/FP rather than any-state acceptance. The main budget sweep examines how TP counts vary with B and M.

Erdős–Rényi (ER) construction. The ER construction samples a fixed number F of distinct true factors from

$$\mathcal{H} = \{h \subseteq \mathcal{V} : |h| \in \{2, 3\}\}.$$

Factors are sampled uniformly within each arity, with nominal arity weights $(0.7, 0.3)$. Selected seed factors are excluded from this count.

The main budget sweep uses $n = 50$, $F = 50$, and 16 known seed factors. It compares $M \in \{1, 2, 4, 8, 16\}$ under Thompson sampling through $B = 16384$. The common seed collection is partitioned across states, while the single state receives its union. The resulting curves appear in Fig. 2C.

Stochastic block-model (SBM) construction. The SBM construction partitions variables into blocks and samples a fixed total of F distinct true factors. Most factors lie within blocks; a scaffold connects successive blocks in a randomly ordered sequence. All scaffold factors count towards F. We use "SBM" as shorthand for this fixed-count, block-structured hypergraph construction, rather than an independent-edge SBM. The fixed-budget comparison uses four blocks of five variables, denoted SBM-4×5.

For the fixed-budget ER and SBM worlds, connected covers are included before the remaining factors are sampled. Both constructions use arities two and three with nominal weights $(0.7, 0.3)$;

connectivity requirements can alter realised proportions. The same effect distribution and total factor count are used in both families.

Fixed-budget policy checks. The policy checks use $n = 20$, $F = 4$, true strengths uniform on $[0.8, 0.9]$, and total budget $B = 4096$. The same 16 known seed factors are partitioned across $M \in \{1, 2, 4, 8, 16\}$; each world contains four non-seed true targets. Each comparison uses ten worlds and ten seed allocations per world. Methods share the sampled world and seed collection, and use priors from the same family with random directions.

Table 13 reports aggregate-score TP and AP. The largest mean TP occurs at $M = 4, 8, 4$ in the random worlds under softmax, Thompson, and LUCB-style acquisition, respectively, and at $M = 4, 4, 8$ in the block worlds. The AP maxima generally agree, except that block-world LUCB-style acquisition has nearly equal means at $M = 4$ and $M = 8$.

		ER-20			SBM-4 × 5		
Policy	M	TP	FP	AP	TP	FP	AP
Softmax	1	0.50 ± 0.20	3.12 ± 0.37	0.107 ± 0.044	0.82 ± 0.29	2.06 ± 0.22	0.191 ± 0.068
	2	0.50 ± 0.18	4.21 ± 0.38	0.095 ± 0.042	0.69 ± 0.18	3.00 ± 0.36	0.154 ± 0.045
	4	$\mathbf{0.53 \pm 0.19}$	1.90 ± 0.25	$\mathbf{0.114 \pm 0.039}$	$\mathbf{1.08 \pm 0.32}$	1.50 ± 0.22	$\mathbf{0.249 \pm 0.072}$
	8	0.42 ± 0.13	1.21 ± 0.22	0.095 ± 0.031	0.74 ± 0.25	1.05 ± 0.22	0.173 ± 0.058
	16	0.21 ± 0.07	1.33 ± 0.14	0.057 ± 0.018	0.62 ± 0.30	0.67 ± 0.12	0.151 ± 0.069
Thompson	1	0.06 ± 0.08	0.82 ± 0.42	0.019 ± 0.020	0.00 ± 0.00	0.31 ± 0.21	0.003 ± 0.001
	2	0.08 ± 0.07	0.57 ± 0.20	0.025 ± 0.018	0.10 ± 0.08	0.27 ± 0.09	0.028 ± 0.019
	4	0.14 ± 0.06	0.45 ± 0.12	0.037 ± 0.014	$\mathbf{0.19 \pm 0.08}$	0.30 ± 0.12	$\mathbf{0.050 \pm 0.019}$
	8	$\mathbf{0.15 \pm 0.10}$	0.31 ± 0.13	$\mathbf{0.041 \pm 0.024}$	0.16 ± 0.05	0.29 ± 0.10	0.043 ± 0.013
	16	0.05 ± 0.04	0.33 ± 0.14	0.018 ± 0.011	0.15 ± 0.06	0.19 ± 0.08	0.040 ± 0.014
LUCB-style	1	0.54 ± 0.24	1.28 ± 0.54	0.128 ± 0.062	0.62 ± 0.40	0.23 ± 0.12	0.156 ± 0.099
	2	1.18 ± 0.45	4.21 ± 0.79	0.254 ± 0.110	1.32 ± 0.40	1.81 ± 0.31	0.308 ± 0.102
	4	$\mathbf{2.02 \pm 0.35}$	4.43 ± 0.68	$\mathbf{0.391 \pm 0.098}$	2.31 ± 0.48	1.58 ± 0.27	$\mathbf{0.558 \pm 0.122}$
	8	1.57 ± 0.51	3.05 ± 0.33	0.303 ± 0.116	$\mathbf{2.45 \pm 0.35}$	2.90 ± 0.39	0.557 ± 0.093
	16	0.21 ± 0.20	0.75 ± 0.21	0.047 ± 0.041	1.63 ± 0.27	1.24 ± 0.24	0.367 ± 0.062

Table 13: Fixed-budget policy checks at $B = 4096$. TP and FP count non-seed true and false hypotheses, respectively, with at least one observation and an observation-count-weighted posterior-mean score of at least τ. The maximum TP count is four. AP evaluates the union of observed hypothesis universes excluding known seeds, assigning score τ to unobserved hypotheses. Entries are mean ± 95% CI half-width across ten world-level summaries, each averaging ten seed allocations. Bold indicates the largest mean TP or AP across tested state counts within each policy and world family.

F AI use statement

We used large language models (LLMs) to assist with improving the clarity of the manuscript, writing code for the experiments, and refining the formatting of proofs, tables and figures. LLMs were not used for research ideation or any other substantive contributions that would merit authorship. The authors take responsibility for the final content of this work.