Experimentation Priorities with Biased Experts*

Yiying Tan

November 2, 2026

Abstract

Decision makers conducting costly experiments often seek advice from experts with a stake in the outcome: executives evaluating competing product designs or policymakers designing new regulation consult specialized engineers or industry lobbyists who have frontier expertise. I study how a decision maker experimenting across multiple prospects may interact productively with a biased expert who strategically discloses verifiable evidence about each prospect. The expert favors a subset of prospects, regardless of their viability for the decision maker, and neither player can commit to future behavior. The order of experimentation shapes the expert's incentives: testing favored prospects earlier can induce disclosure about alternatives if experimentation continues. Pareto-efficient equilibria feature disclosure in batches, generally with delay. Between disclosures, the decision maker experiments optimally according to an adjusted Gittins index. Under refinements that rule out implausible disclosure behavior, optimal experimentation follows an endogenous index ordering across all prospects. Comparative statics highlight that faster learning or greater alignment with the expert need not improve the decision maker's welfare.

*Click here for the latest version. I thank Leeat Yariv, Alessandro Lizzeri, Mira Frick, Fedor Sandomirskiy, Eran Shmaya, Sylvain Chassang, Pietro Ortoleva, many friends and colleagues at Princeton University, and participants at the 37th Stony Brook International Conference on Game Theory and the 4th International Workshop on Game Theory and Economic Applications for insightful comments.

1 Introduction

In many settings involving costly experimentation, decision makers work closely with expert advisors whose interests shape the guidance they provide. An R&D executive deciding which product designs to develop may consult knowledgeable engineers, who know the latest technical benchmarks; they may have developed the proof of concepts for particular designs and favor furthering those candidates. A policymaker deliberating regulatory standards for an emerging industry may consult industry participants and lobbyists, who are well placed assess feasibility but favor standards that are less costly to meet. While the decision maker holds ultimate authority over experimentation, and so controls the flow of new knowledge, the expert has superior command of frontier knowledge and may hold sway with control over the stock of existing evidence. How do these two sources of influence interact to shape the course and outcomes of experimentation? What makes this relationship more or less productive?

This paper develops a tractable framework to study the dynamic interaction between a decision maker (he) facing a multi-armed bandit problem with n competing risky prospects, and a biased expert (she) who seeks to steer experimentation through strategic disclosure. Crucially, these two problems jointly evolve, and must remain mutually consistent throughout. At each point in time, the decision maker's incentives depend on both forward-looking expectations about what he expects the expert to disclose as experimentation unfolds, and inferences made about remaining prospects given past disclosure or the lack thereof. The expert's incentives depend on how she expects experimentation to evolve. Although the expert may be initially reluctant to disclose evidence of viable alternatives, she may become willing to do so after some of her favored prospects have already been tested and ruled out. I identify a class of equilibria I call *Batched Disclosure Equilibria*, which satisfy properties necessary for Pareto efficiency. Expert disclosure occurs periodically: roughly, disclosure is delayed and tied to the failure of the expert's targets. Between disclosure events, the decision maker experiments optimally, in a sequence consistent with an adjusted Gittins index. Furthermore, under refinements that rule out implausible disclosure behavior, optimal experimentation in DM-optimal equilibria can be characterized by an endogenous index ordering across all prospects.

A key qualitative insight is that dynamic incentives rationalize experimentation paths that would be suboptimal under a static communication model, and suggest that long-term rela-

tionships between decision makers and informed stakeholders may be beneficial. The analysis demonstrates that the decision maker benefits by prioritizing expert-targeted prospects, relative to what their individual merits warrant, in exchange for better information about the remaining alternatives should prioritized prospects fail. In comparison, if limited to a static exchange, the decision maker ought to delay the expert's targets. Equilibria with dynamic engagement, where disclosure occurs with delay, yield Pareto improvements over a benchmark corresponding to static interaction (*Minimum Disclosure Equilibrium*). An interpretation is that compared to one-off consultations, long-term relationships sustained by mutually consistent norms of engagement may help both parties. Saliently, this holds even as both sides are driven by transparently orthogonal interests. Commercial labs may retain top scientists and indulge their agendas, even after early ideas fail to generate commercial success, without being either benevolent or inherently invested in long-shot research. Policymakers may build productive ongoing relationships with interested lobbyists, and even take on their suggestions in early phases of regulation, without being either naive or avaricious.

The model describes a decision maker (DM) who continually chooses one experiment among n heterogeneous risky prospects, and a biased expert (Expert) who may choose at any point to verifiably disclose any subset of her private evidence confirming the viability of particular prospects. Each party is forward-looking and discount future payoffs at a common rate $r > 0$; neither party has commitment power. The objective of the DM is always to find and implement a viable prospect as soon as possible: he receives a positive flow payoff only when the chosen prospect is viable. Every viable solution is equally satisfactory: thus, experimentation concludes once a viable prospect is discovered. The Expert receives a positive flow payoff only when the DM chooses a targeted prospect. She always prefers that her targets are prioritized, even if they are unviable and thus delay the implementation of a known viable solution. While choosing some prospect j, the DM collects an unobserved flow payoff; experimental news that reveals the payoff arrives at a state-independent Poisson rate λ_j. Thus, without the expert, "no news is no news": the DM's posteriors remain constant until news arrival.

The key challenge of characterizing feasible equilibria arises as each player's strategy must be dynamically consistent; dynamic consistency depends on their counterpart's strategy, which is in turn determined by incentives shaped by each player's own strategy. In equilib-

rium, experimentation must remain consistent with credible expectations about future disclosure, as well as with correctly updated posteriors ex-post. Simultaneously, conditional probabilities of evidence disclosure must be incentive-compatible throughout, given continued experimentation. Nevertheless, the framework supports tractable analysis of Markov Perfect equilibria. Three helpful observations facilitate progress. First, Markovian strategies may be defined on the sequence of observed failures, rather than the DM's multidimensional posterior. Second, we may focus on deterministic experimentation strategies if our focus is to consider Pareto efficient equilibria. Third, in the main analysis, it remains weakly optimal throughout for the Expert to withhold further evidence on non-targeted prospects in continuation: this provides a tool of punishment that helps sustain equilibrium experimentation. Without imposing further restrictions, checking for feasiblity boils down to inspecting incentives on n equilibrium-path states: disclosure must be credible relative to the experimentation path, and that the DM's continuation payoffs at each equilibrium-path state must defeat his "minimax" outcome of independent experimentation in continuation without further input.

The set of Pareto-efficient outcomes can be achieved within the class of Batched Disclosure Equilibria. Expert disclosure occurs as soon as possible, subject to her incentive constraints and extensive strategic concerns. Generally, disclosures are tied to the failure of each targeted prospect, and when they occur, happen at the outset of each subsequent "batch" of non-targeted prospects. Even if no evidence is disclosed, the DM is empowered to experiment more efficiently in continuation by making inferences from non-disclosure. Between disclosures, the DM experiments optimally according to an adjusted Gittins index, reminiscent of established intuitions for single-agent experimentation problems. One may interpret Batched Disclosure Equilibria with delayed disclosure as a series of repeated trades. Intuitively, there are greater gains from trade if the decision maker is allowed to access information more quickly and use it more efficiently. The DM willingly expedites targeted prospects due to correctly anticipating the relative reward of greater subsequent disclosure, hedging against the possible failure of more promising non-target prospects. The Expert leverages contingent alignment to credibly promise disclosure after her targets fail and secure the DM's cooperation early on.

I characterize DM-optimal equilibria after introducing two refinements. In applying these refinements, we restrict implausible off-path punishments by the Expert, and focus on strate-

gies that do not demand extensive contingent reasoning. I obtain an endogenous index characterization of DM-optimal experimentation among refined Batched Disclosure Equilibria: optimal experimentation orders all prospects in descending order of Gittins index values that account for equilibrium disclosure. If we fix disclosure, the decision maker ultimately still faces an exploration-exploitation problem where he balances the probability of finding success with the speed of arrival of experimental news. Thus, it may seem sensible that an index capturing these trade-offs still characterizes optimal experimentation. Yet, in this setting, the decision maker's perception of each prospect's attributes is non-separable across prospects and endogenously sustained. Hence, the emergence of an endogenous index characterization is not a forgone conclusion. Optimal experimentation in feasible equilibria generally does not abide by an index ordering across all prospects. However, I also show that if news arrival rates and expert knowledge are homogeneous across prospects, then optimal experimentation in DM-optimal equilibria actually yields a stronger characterization: the order of experimentation coincides with following the Whittle index policy in an auxiliary restless-arm problem.

Equilibrium dynamics demonstrate how the Expert's conditional disclosure can be wielded to shape the DM's incentives. Notably, experimentation also doubles as an incentive instrument for the DM, given how it changes the willingness of the expert to communicate. Importantly, the effect of expected disclosure on the DM's posteriors introduces strategic constraints on both levers of influence. To illustrate, suppose the DM considers a strategy where he chooses the Expert's targets only when all other alternatives have failed. If he follows through, the Expert would be indifferent about immediately revealing the viability of any prospect, including on non-targets: the DM would have discovered some viable non-target anyway, before trying any of the expert's targets. However, this strategy may not be dynamically consistent. If the Expert indeed discloses any evidence on non-targets immediately, no disclosure would be a negative signal. Absent upfront disclosure, the DM may prefer ex-post to instead delay non-targets. Then, the Expert may do better by withholding evidence on alternatives she anticipates will be delayed after her targets. Similarly, the Expert may hope to encourage immediate experimentation on her most promising target by promising to disclose all available evidence once it fails. Yet, if the DM expects this, he may be sufficiently pessmistic about non-targets when no disclosure occurs, such that he delays them after a remaining Expert-targeted prospect. The Expert's overpromise would not be credible.

Comparative statics on DM welfare reveal interesting non-monotonicities that stem from strategic constraints on credible upfront disclosure. In Section 5, I consider comparative statics on DM welfare, focusing on DM-optimal refined equilibria, with respect to rates of experimental news arrival ("learning rates"), expert knowledge, and expert misalignment. As an overview, the decision maker always benefits from faster learning about non expert-targeted prospects, and from greater expert knowledge on the expert's targets. However, welfare does not always improve with a faster rate of learning about expert-targeted prospects, nor with greater expert alignment. At a high level, while these may improve the DM's continuation payoffs if prioritized prospects are indeed unviable, they may also induce a decision maker to expedite the expert's targets before a larger set of potentially viable non-targeted prospects.The same rate of disclosure may make marginal non-targeted prospects less attractive than the expert's targets ex-post, and disclosure on these marginal prospects may no longer be credible. While a viable solution will be inevitably discovered, its implementation would be delayed.

The paper proceeds as follows. I first provide an outline of related literature. Then, Section 2 sets out the main model. Section 3 characterizes feasible equilibria, and introduces the class of Batched Disclosure Equilibria. Further, I introduce refinements maintained in subsequent analysis of optimal equilibria. Section 4 analyzes DM-optimal equilibria. I describe the endogenous index ordering that characterizes experimentation, and provide a connection to restless bandits. In addition, I show how DM-optimal welfare may be computed. Section 5 considers comparative statics, and considers how Expert evidence, news arrival rates, and alignment affect DM welfare. Finally, Section 6 concludes and discusses key extensions.

Related Literature

The model of experimentation in this framework draws from canonical exponential bandit models originated in Keller et al. (2005) and Keller and Rady (2010). Characterizations of optimal experimentation in this paper also build upon classic insights on multi-armed bandit problems stemming from Gittins (1979), and connect to ideas in Whittle (1988). This paper develops a tractable and parsimonious framework that reconciles the analysis of rich strategic issues with the flexibility of accommodating an arbitrary number of n heterogeneous prospects. The ability to consider n prospects allows my analysis to comment on the direction of experimentation: which experiments occur first and in what order. This departs from much

of the existing literature in strategic experimentation, which primarily focus on the extent and intensity of experimentation effort.

An important strand of the strategic experimentation literature addresses incentives surrounding joint experimentation (Bolton and Harris, 1999; Keller et al., 2005; Keller and Rady, 2010; Bonatti and Hörner, 2011). In a model of experimentation with voting, Strulovici (2010) shows how learning changes the incentives for continued experimentation through future voting power. Here, experimentation is also a joint problem: both the decision maker have direct stakes which experiments are chosen and what is ultimately implemented; each hold some discretion over publicly available information. However, here, each party controls disjoint information sources: the decision maker experiments while the expert controls the stock of existing evidence.

Another key focus in the literature concerns the design of incentives for experimentation: a principal seeks to institute incentives to motivate efficient experimentation by agents. For example, Guo (2016) studies how a principal should delegates experimentation to an informed agent with an unobserved posterior. Other papers also study how principals may design efficient contracts or approval mechanisms in lieu of direct control over experimentation (Hörner and Samuelson, 2013; Halac et al., 2016; Henry and Ottaviani, 2019; McClellan, 2022). In this paper, the DM retains direct control of experimentation and final implementation. In the main analysis, neither party provides financial incentives, nor wield commitment power. Instead, experimentation is a direct instrument of influence by the DM over the asymmetrically informed Expert; both are constrained by ex-post incentives.

My paper is closely connected to relatively recent literature work studying how disclosure may shape subsequent learning and experimentation. Knoepfle and Salmi (2025) and Che and Hörner (2018) show how a benevolent designer may control public news to encourage adoption and evidence generation. In both cases, the principal is a sender who but leverages disclosure to influence experimentation by multiple receivers; free-riding and underexperimentation are key features. In this paper, the DM is the sole evidence-generating experimenter. The Expert controls established evidence, and influences experimentation via disclosure; but crucially the experimenter now directly influences disclosure incentives. Neither party can design incentives that require commitment power. In Orlov et al. (2020) a biased sender grad-

ually discloses information, and seeks to delay the receiver's action; the key constraint is the receiver's patience. The Expert in this framework does not inherently benefit from delaying the decision-maker's implementation of a viable prospect. While she seeks to steer the decision maker toward her targets, she may become more or less aligned as outcomes unfold.

This paper connects more broadly to strategic communication between principals and experts, stemming from seminal papers such as Crawford and Sobel (1982); Krishna and Morgan (2001) and Dewatripont and Tirole (1999). The framework in this paper draws from canonical models of disclosure developed in Grossman (1981); Milgrom (1981); Dye (1985): inference by the DM absent disclosure corresponds to the concept of "unraveling". Other work shows how principals' decision rules can encourage information acquisition, for instance in Che and Kartik (2009) and Li (2001). The role of experimentation in shaping disclosure incentives shares a similar spirit. This paper embeds and extends insights on principal-expert interactions in a dynamic framework where communication incentives continually interact with experimentation.

The qualitative results obtained from this framework offers an interpretation of policy entrepreneurship and innovation. Policy entrepreneurs promote alternatives and organize support for them (Kingdon, 1984; Mintrom, 1997; Mintrom and Norman, 2009), while expertise and social learning shape policy choices (Haas, 1992; Hall, 1993). A key theme in work on innovation is that ownership shapes R&D incentives (Aghion and Tirole, 1994), and economic incentives direct technical change (Acemoglu, 2002). The insights here highlight how expertise may shape the direction of innovation through dynamic incentives: experts use the expected utility of their continued guidance and involvement to secure resources to prioritize their favored prospects. The details of the mechanism also suggest that governments, firms, or funders may benefit more from expertise over a complex terrain with many possible approaches and no obvious winners: here, extended search may be more likely, and better information in continued experimentation would be more important.

2 Model

A decision-maker ("DM") faces a menu $M = \{1, 2, \ldots, n\}$ of uncertain prospects. The viability of each prospect $j \in M$ is described by a binary state, $\theta_j \in \{g, b\}$: $\theta_j = g$ indicates that

a prospect is viable for the DM ("good"), and $\theta_j = b$ indicates that it is not ("bad"). Each alternative prospect—for example, competing regulatory standards, product designs, or external contractors—either satisfies the DM's requirements or does not. Qualities are mutually independent, with commonly known priors $p_j := \Pr(\theta_j = g) \in (0,1)$.

The DM implements one prospect at a time. At each $t \geq 0$, he chooses some $a(t) \in M$. While prospect j is chosen, the DM collects a flow payoff $u(j,\theta) := \mathbf{1}\{\theta_j = g\}$, which he does not observe while its viability remains uncertain. At the same time, public news revealing prospect j's quality arrives at Poisson rate $\lambda_j > 0$.[1] The rate of news arrival is independent of the value of θ_j. Thus, the absence of experimental news alone is uninformative.[2] For instance, regardless of the eventual outcome, similar time and effort are required to properly evaluate a prospect: facilitating a regulatory sandbox, creating a product prototype, or collecting sufficient data on a pilot program. Before evaluation concludes, the duration of activity provides no information about viability.

The DM discounts at a constant rate $r > 0$ and maximizes expected discounted flow utility. For an allocation process $a = (a(t))_{t\geq 0}$, his normalized payoff is

$$U_{DM} = \mathbb{E}\left[r \int_0^\infty e^{-rt} u(a(t),\theta)\, dt \right].$$

The expectation includes uncertainty about viability, the timing of experimental news, and any strategy randomization. If all prospects fail, the DM gets a continuation payoff of zero. If he acts alone, the DM solves a multi-armed bandit problem, trading off higher current expected payoffs against faster learning. An optimal policy implements unresolved prospects in descending order of their Gittins indices,

$$G_j = \frac{p_j}{1-(1-p_j)\rho_j}, \qquad \rho_j := \frac{\lambda_j}{r+\lambda_j},$$

continuing with each until news arrives and implementing a confirmed good prospect thereafter.[3] In many settings, however, decision-makers do not experiment in isolation. Policymakers consult lobbyists and researchers; product managers rely on engineers and other techni-

[1] Let $\tau_j \in [0,\infty]$ denote the arrival time of experimental news about j. Before τ_j, the arrival rate is λ_j while j is chosen and zero otherwise, so only time spent implementing j contributes to learning. The underlying Poisson processes are independent across prospects and of qualities, evidence, and private randomization.

[2] This feature may be termed "balanced news", following Lizzeri, Shmaya, and Yariv (June 2026 draft).

[3] See Gittins (1979). Appendix A.1 derives the index and ordering directly for this environment.

cal specialists. These informed stakeholders may influence experimentation despite lacking decision-making authority.

An informed stakeholder ("Expert") observes the same menu. She is publicly biased towards a nonempty proper subset $M_H \subset M$; alternatives outside this subset belong to $M_L := M \setminus M_H$. I call the Expert's targets H-prospects and the remaining prospects L-prospects. During implementation of prospect j, she receives flow payoff $x(j) := \mathbf{1}\{j \in M_H\}$, regardless of its quality. For instance, a lobbyist may receive remuneration while a client's preferred regulatory standard is in place; an engineer may benefit from organizational visibility while her preferred project is in development. The Expert discounts at the same rate, r and conditional on her private evidence ω described below, evaluates an allocation by

$$U_{E,\omega} = \mathbb{E}\left[r \int_0^\infty e^{-rt} x(a(t))\, dt \,\middle|\, \omega \right].$$

Thus, she values earlier and longer implementation of her favored prospects, including the time spent experimenting with a favored prospect that ultimately proves bad. If all prospects fail, the Expert gets a continuation payoff of zero. Let $U_E = \mathbb{E}_\omega[U_{E,\omega}]$ denote the Expert's ex-ante payoff.

With some probability, the Expert has additional private news about θ, in the form of preexisting verifiable evidence that confirms a prospect's viability. In particular, let $\omega_j \in \{0,1\}$ indicate whether she possesses evidence that $\theta_j = g$. I assume that the pairs (θ_j, ω_j) are mutually independent across prospects.[4] Expert knowledge is captured by evidence capacity μ, specifying the probability with which the Expert possesses positive evidence on prospect j when it is viable for the DM. For each prospect $j \in M$, this probability is $\mu_j := \Pr(\omega_j = 1 \mid \theta_j = g) \in [0,1]$. For instance, an engineer may know about a successful instance of deployment published in a technical journal, or deduce the feasibility of a design by observing competitors' products; a policy entrepreneur may point to records of successful policy implementation in a comparable setting. This "good-news" model of Expert knowledge reflects a natural asymmetry. While private experimentation may yield both positive and negative conclusions, successful implementations can leave publicly verifiable evidence, whereas records of failed attempts

[4]This permits latent factors correlated with qualities and affecting evidence availability across prospects, provided they induce the same joint law of qualities and evidence. Appendix E gives a brief construction. The latent factors are unobserved by both players and have no additional effects on payoffs or experimentation.

may remain confidential.[5]

At each $t \geq 0$, the Expert observes which prospects are chosen by the DM and any news from experimentation. Given her private evidence $\omega = (\omega_j)_{j \in M}$ and what she has observed thus far, she may disclose a subset $D(t) \subseteq \{j \in M : \omega_j = 1\}$. I assume that disclosure is verifiable: a successful prototype or positive report cannot be fabricated.

Information and Strategies

Both players observe which prospects the DM chooses, and what experimental news arrives. The DM additionally observes whatever the Expert has disclosed. In general, strategies may depend on all observable actions and outcomes. However, I focus on Markovian strategies, which depend only on the sequence of observed failures. Intuitively, observed outcomes should contain all payoff relevant information from experimentation: since the rate of news arrival is uncorrelated with viability, the duration of experimentation should not be inherently informative.

Let $\mathcal{K}$ denote the collection of ordered lists of distinct prospects, of any length, including the empty list. The state of public knowledge, $\kappa \in \mathcal{K}$, describes the sequence of prospects that have been chosen and observed to fail, that is, for which bad news has arrived. Let $K(\kappa)$ denote the set of failed prospects.

Remark (Markovian strategies)**.** *A strategy is Markovian if it depends only on $\kappa = (j_1, \ldots, j_i)$, the sequence of observed failures.*

The arrival of experimental news either corresponds to the discovery of a viable solution and ends experimentation (good news), or changes the state by extending the sequence of observed failures (bad news). I now describe Markovian disclosure and experimentation strategies.

Disclosure. A deterministic disclosure strategy $d \in \mathcal{D}$ specifies, at each state, a set of prospects $d(\kappa) \subseteq M \setminus K(\kappa)$ about which the Expert should reveal any evidence she holds. She

[5]For example, the complaint in *Waymo v. Uber* (paragraph 36) describes knowledge of "dead-end" designs as valuable trade secrets. The order in *Genentech v. JHL Biotech* (March 5, 2019, pp. 19, 31) discusses proprietary "negative know-how" that can help a competitor avoid unsuccessful development paths.

can reveal only evidence she possesses, so the disclosed set is

$$D(\kappa) = d(\kappa) \cap \{j : \omega_j = 1\}.$$

A mixed disclosure strategy $\phi(\omega) \in \Delta(\mathcal{D})$ specifies a distribution over these strategies. I assume that disclosure about each prospect depends only on the Expert's evidence about that prospect, with independent randomization across prospects.[6] Let $\phi_j(\omega) : \mathcal{K} \to [0,1]$ denote conditional probabilities with which the Expert with evidence ω discloses about prospect j at each state $\kappa \in \mathcal{K}$.[7] As separable disclosure across prospects is maintained throughout the main analysis, I generally suppress the argument ω. For an unresolved prospect j, $\phi_j(\kappa)$ denotes the conditional probability of disclosure at state κ, given that the Expert holds private evidence about j.

Experimentation. A deterministic experimentation strategy a selects an unresolved prospect $a(\kappa) \in M \setminus K(\kappa)$ at each continuing state and implements it until news arrives. Let $\mathcal{A}$ denote the set of these deterministic strategies. A mixed experimentation strategy $\alpha \in \Delta(\mathcal{A})$ specifies a distribution over $\mathcal{A}$. The DM's randomization is independent across states and independent of the Expert's randomization. When the DM can be assumed to use a deterministic strategy, I use a rather than α to denote his strategy.

Consistent Posterior Beliefs. Posterior updating when there is no observed disclosure depends on consistent beliefs about the Expert's disclosure strategy. If the Expert is expected to disclose evidence about j, non-disclosure lowers the DM's belief that positive evidence exists, and therefore also his posterior on j. For an unresolved prospect j, let $\varphi_j(\kappa)$ denote the DM's probability that the Expert holds good news about j and $\widetilde{p}_j(\kappa)$ denote his belief that j is good if no-disclosure has been observed at state κ. The Expert's posterior on prospect j is either 1 if she has evidence on its viability, or p_j^- if not, where

$$p_j^- := \frac{p_j(1-\mu_j)}{1-p_j\mu_j}.$$

[6] Together with the independence of the pairs (θ_j, ω_j), prospect-separable disclosure and independent randomization ensure that the DM's belief updating is separable across unresolved prospects.

[7] Under the strategy, conditional on ω and earlier non-disclosure, the probability of disclosing exactly the set $S \subseteq M \setminus K(\kappa)$ at κ is $\prod_{j\in S} \phi_j(\omega)(\kappa) \prod_{j\in M\setminus(K(\kappa)\cup S)}[1-\phi_j(\omega)(\kappa)]$. If $\omega_j = 0$, then $\phi_j(\omega)(\kappa) = 0$. Appendix A.2 gives the product restriction on the full mixed strategy.

The DM's posterior on j is a garbling of the Expert's posterior:

$$\widetilde{p}_j(\kappa) = \varphi_j(\kappa) + (1 - \varphi_j(\kappa))p_j^-.$$

Let $\widetilde{p}_j(\emptyset)$ denote the DM's posterior at $\kappa = \emptyset$. This differs from his prior whenever immediate disclosure about j occurs with positive probability.

Whenever disclosure is allowed to vary across observed outcomes, the order of failures matters, not just which prospects have failed. Even if the Expert only conditions disclosure on the set of failures, the probability with which the Expert is expected to disclose evidence on remaining prospects may differ between different sequences of experimentation over the same early prospects. Then, what the DM infers about remaining prospects may depend on the sequence of observed failures, and his strategy may not be fully specified over just the set of failures.[8]

Definition 1 (Markov Perfect Equilibrium (MPE))**.** *A pair of Markovian strategies* (α, ϕ)*, together with consistent Bayesian beliefs, is a Markov Perfect Equilibrium if each player's prescribed continuation maximizes that player's expected discounted payoff at every decision history, given the other's strategy.*

In equilibrium, experimentation must be consistent with credible expectations about future disclosure ex-ante to experimentation, as well as with correctly updated posteriors ex-post. At every state, the chosen prospect should maximize the DM's expected continuation payoffs, given equilibrium continuation. Similarly, disclosure must be incentive compatible given continued experimentation at every stage. The Expert's disclosure must maximize her expected continuation payoffs. Hence, her strategy should never prescribe disclosure on non-targeted L-prospects at state κ if withholding that evidence would leave a positive probability that a targeted H-prospect is implemented in continuation from κ.

[8] Appendix A.3 includes a more detailed discussion. Under the maintained assumption of prospect-separable disclosure and independent randomization, the sequence of failures is the coarsest state that is uniformly closed under Bayesian updating over state-measurable strategy profiles. When strategies condition on the state, posterior beliefs and best responses must also be measurable with respect to that state.

3 Feasible Equilibrium Outcomes

What patterns of experimentation and disclosure can arise in equilibrium? I begin with a natural benchmark, the *Minimum Disclosure equilibrium*, in which the Expert reveals her H-evidence at the outset and withholds all L-evidence. Disclosure is minimal in the sense that the Expert generally discloses evidence only when it is strictly beneficial to her ex-post.[9] I then illustrate, using a simple example, why better equilibria may exist: in the example, the Expert discloses any remaining private evidence once her sole target fails.

I then characterize feasible equilibrium outcomes. I begin with lemmas that make the analysis tractable; in particular, Pareto-efficient payoffs can be attained with a deterministic experimentation order. These results lead to necessary and sufficient conditions for feasibility. Next, I introduce prompt disclosure and interval-optimal experimentation, which define *Batched Disclosure equilibria*, and show why Pareto efficiency requires both. Even within this class, particular equilibria are sustained by relatively implausible strategies that require extensive contingent reasoning or punish the DM for choosing the Expert's favored prospects early. I therefore introduce refinements that restrict such strategies.

3.1 A Simple Benchmark

There always exists a straightforward equilibrium in which the Expert discloses all her good news about H-prospects at the outset, and nothing else.

Definition 2 (Minimum Disclosure Strategy)**.** *Under a* minimum disclosure strategy $d_{\min}$*, the Expert discloses all her H-evidence at the outset and no L-evidence.*

Proposition 1 (Minimum Disclosure Equilibrium)**.** *There exists a Markov perfect equilibrium in which the Expert follows the minimum disclosure strategy $d_{\min}$.*

All viable prospects are equally satisfactory to the DM: once a prospect is confirmed viable, he ceases experimentation and implements it perpetually. Disclosing H-evidence therefore induces immediate implementation of an expert-targeted prospect, attaining the Expert's highest feasible payoff, one. Revealing this evidence is ex-post optimal. Revealing only L-

[9] An exception is where H-prospects are first to be chosen absent disclosure even when the DM assumes all H-evidence is disclosed and becomes pessimistic absent disclosure, in which case the Expert is indifferent.

evidence instead ends all further H-experimentation, so the Expert cannot benefit by doing so. Minimum disclosure is thus a best response regardless of the DM's experimentation choices, given his response to confirmed viability.[10]

Under minimum disclosure, the DM either immediately implements a viable H-prospect or continues experimenting independently. In the latter case, his best response solves a single-agent experimentation problem with beliefs adjusted for the absence of immediate disclosure. Since the Expert would reveal any H-evidence she has, non-disclosure establishes that she has none. This weakly lowers the DM's belief in each H-prospect to $\widetilde{p}_h(\emptyset) = p_h^-$. No disclosure is expected about L-prospects, so their beliefs remain at their priors. The DM therefore experiments in decreasing order of Gittins indices,

$$G_j\big(\widetilde{p}_j(\emptyset)\big) = \frac{\widetilde{p}_j(\emptyset)}{1-\rho_j[1-\widetilde{p}_j(\emptyset)]}, \qquad G_j := G_j(p_j).$$

News arrives independently of viability, so these indices remain constant until news arrives. The DM continues each experiment until its outcome is known, moving to the next prospect after a failure.

Minimum disclosure need not be the best equilibrium for either player. Once all the Expert's targets have failed, she has nothing to gain from further experimentation and is indifferent about revealing L-evidence. She can therefore promise to disclose this evidence after her targets fail. Anticipating a better-informed choice among the alternatives, the DM may be willing to choose a target earlier. The following example shows how this exchange can improve both players' payoffs.

Example 1 (Tail Disclosure)**.** *There are two identical L-prospects and one H-prospect, with*

$$p_{\ell_1} = p_{\ell_2} = 0.6, \quad p_h = 0.3, \qquad \mu_{\ell_1} = \mu_{\ell_2} = 0.9, \quad \mu_h = 0.1, \qquad \rho_{\ell_1} = \rho_{\ell_2} = 0.4, \quad \rho_h = 0.8.$$

Under minimum disclosure, the DM chooses ℓ_1, then ℓ_2, and only then h, whose index is lowest even at its prior. There is another equilibrium in which the Expert discloses all L-evidence she holds as soon as h fails, and the DM chooses h first. Both players are strictly better off in the second equilibrium.[11]

[10] Appendix B.1 constructs the equilibrium, including behavior after earlier concealment.

[11] Appendix B.7 verifies both equilibria and gives the payoffs.

The H-prospect is a long shot that resolves quickly, the L-prospects are more promising but slow to test, and the Expert is likely to know when one of them is good. Testing h first therefore costs the DM little delay and, should h fail, brings him the Expert's knowledge of the alternatives. The Expert, for her part, sees her prospect chosen first with certainty rather than last and only after both L-prospects have failed.

3.2 Equilibrium Experimentation and Disclosure

The preceding example shows how later disclosure can change the DM's experimentation priorities. I now characterize which combinations of experimentation and disclosure can be sustained in equilibrium.[12]

Markovian strategies depend only on the state of public knowledge, and this allows equilibrium experimentation to be described compactly. Whenever experimentation continues, the DM's choice in a Pareto-optimal MPE can be taken to be deterministic and therefore to follow a fixed sequence.

Lemma 1 (Equilibrium Experimentation as Priority Policy)**.** *Among Pareto-optimal MPE, equilibrium payoffs can be attained with a priority ordering $\pi : M \to M$: a fixed sequence in which the DM implements prospects until a viable prospect is confirmed or all prospects fail.*

Two observations underlie the lemma. First, randomization by the DM can be dispensed with. Suppose that at some state the DM is willing to mix between prospects. Each prospect in the support of his mixture is then optimal for him, so he loses nothing by selecting one of them deterministically. The Expert's incentives are unaffected by the selection: at any state where she is willing to disclose L-evidence with positive probability, she must have nothing to gain by withholding along any continuation supported by his mixture. Selecting among these continuations therefore preserves her willingness to disclose. Breaking ties in the Expert's favor preserves both players' payoffs at a Pareto optimum.

Second, given deterministic experimentation, the state can change in only two ways while search progresses: the prospect being tested fails, or good news arrives, whether from the experiment or from disclosure. In the latter case the DM ceases nontrivial experimentation

[12] As in the benchmark, I maintain initial disclosure of all H-evidence, including when the Expert is indifferent. L-evidence may be disclosed initially or later.

and implements a confirmed good prospect. In the former, the failed prospect is appended to κ, and the DM's strategy again prescribes a single prospect at the new state. The next experiment in the event of failure is thus predetermined at every point, and the sequence of experiments absent disclosure is a fixed priority order π.[13]

For a priority order π, write

$$\kappa_\pi^{(i)} = (\pi(1), \ldots, \pi(i)), \qquad K_\pi^{(i)} = K(\kappa_\pi^{(i)}), \qquad i = 0, \ldots, n,$$

with $\kappa_\pi^{(0)} = \emptyset$. When equilibrium experimentation follows π, these are the states reached with positive probability. Let ϕ_π denote the Expert's equilibrium-path disclosure: her conditional probabilities of disclosing evidence about each unresolved prospect at these states. The experimentation–disclosure pair (π, ϕ_π) is *feasible* if it is induced by some MPE (a, ϕ). Feasibility requires the Expert to be willing to disclose as anticipated and the DM to be willing to follow the prescribed order.

Without H-evidence, revealing that an L-prospect is viable ends experimentation before she can benefit from any remaining H-prospect. She is willing to disclose if withholding would likewise lead the DM to choose that viable L-prospect before any remaining H. If an H-prospect comes first, she can instead benefit by withholding the evidence.

Lemma 2 (Credible Disclosure)**.** *Fix a deterministic experimentation policy. A disclosure strategy is a best response if the Expert follows these rules at every state κ:*

(i) *She discloses any H-evidence she has.*

(ii) *She discloses evidence about an L-prospect only if, without further disclosure, the DM would choose it before any remaining H-prospect.*

The DM must in turn be willing to follow the prescribed experimentation strategy. *Incentive compatibility* (IC) requires his continuation payoff at every continuing state to be at least as high as that from choosing a different prospect and then experimenting optimally under the specified disclosure strategy ϕ.

[13] See Appendix B.2, particularly Lemmas B.2 and B.1, which state the precise hypotheses and prove that the Expert remains willing to follow her disclosure strategy after the DM's policy is selected. Selection here precedes the additional refinements introduced below.

The Expert can impose maximum punishment by cancelling further disclosure after an observed deviation. I interpret the corresponding on-path incentive constraint as an *individual rationality* (IR) condition. The DM must be willing to follow the prescribed experimentation order in exchange for further disclosure. His alternative is to choose a different prospect and then experiment optimally without further disclosure.

At an on-path state κ after non-disclosure, write $V_{\pi,\phi_\pi}(\kappa)$ for the DM's value from following π under ϕ_π. Let $V_{\min}(\kappa;\pi,\phi_\pi)$ be his optimal value without further disclosure at his current beliefs, and $V^{\text{dev}}_{\min}(\kappa;\pi,\phi_\pi)$ the same value restricted to a distinct first prospect. Individual rationality requires

$$V_{\pi,\phi_\pi}(\kappa) \geq V^{\text{dev}}_{\min}(\kappa;\pi,\phi_\pi). \tag{1}$$

IR *binds* if some distinct first prospect attains equality.

Proposition 2 (Feasible Markov Perfect Equilibria)**.** *The pair (π,ϕ_π) is induced by an MPE if and only if:*

1. ***Credible disclosure.*** *Under ϕ_π, the Expert discloses evidence about an L-prospect only if, without further disclosure, the DM would choose it before any remaining H-prospect.*
2. ***DM individual rationality.*** *Inequality* (1) *holds at every continuing on-path state with a distinct available action.*

Proposition 2 shows that we can determine whether (π,ϕ_π) can be sustained in equilibrium by checking credibility and individual rationality only along π. The Expert must be willing to disclose as promised, and the promised information must make the DM willing to follow the order. To sustain the proposed experimentation and disclosure, the Expert can cancel further L-evidence disclosure after an observed deviation. The DM then experiments optimally using what he has already learned, including from earlier non-disclosure.[14]

Individual rationality ensures that the DM is at least as well off following the equilibrium as deviating and experimenting optimally without further disclosure. The next two examples show how a large range of outcomes can be sustained given the laxity afforded by the possibility of extreme punishment. While the examples are by definition an improvement over

[14] Appendix B.3 verifies this continuation and both players' incentives at every state.

minimum disclosure, they are clearly inefficient.

In each example, every prospect has $\rho_j = 1/2$, and the H-prospect has prior $p_h = 0.9$ and evidence capacity $\mu_h = 0$. Thus, h has the highest Gittins index. Once h fails, the Expert is indifferent about subsequent experimentation and disclosure. After an observed deviation, the Expert cancels further disclosure.

Example 2 (Delayed disclosure)**.** *Consider one H-prospect h and two L-prospects ℓ_a, ℓ_b, with $(p_{\ell_a}, p_{\ell_b}) = (3/5, 2/5)$ and $(\mu_{\ell_a}, \mu_{\ell_b}) = (0, 4/5)$. Prescribe the order (h, ℓ_a, ℓ_b), disclosing the Expert's evidence about ℓ_b only after h and ℓ_a fail. After h fails, the DM prefers ℓ_a even without further disclosure: it has the higher prior and the same learning rate. Delaying disclosure therefore leaves him willing to follow this order.*

However, revealing the evidence as soon as h fails can spare the DM an unsuccessful experiment on ℓ_a, without harming the Expert.

Example 3 (Inefficient ordering)**.** *Consider one H-prospect h and four L-prospects $\ell_a, \ell_b, \ell_c, \ell_d$, with $(p_{\ell_a}, p_{\ell_b}, p_{\ell_c}, p_{\ell_a}) = (0.5, 0.51, 0.4, 0.4)$ and $(\mu_{\ell_a}, \mu_{\ell_b}, \mu_{\ell_c}, \mu_{\ell_d}) = (0, 0, 1, 1)$. Prescribe the order $(h, \ell_a, \ell_b, \ell_c, \ell_d)$, disclosing all the Expert's evidence about ℓ_c, ℓ_d only after h, ℓ_a, ℓ_b fail in that order. Although $p_{\ell_b} > p_{\ell_a}$, the gain from choosing ℓ_b first is smaller than the loss of the promised disclosure, so the DM prefers to follow the prescribed order.*[15]

Yet swapping ℓ_a and ℓ_b while retaining disclosure after both fail improves the DM's payoff without changing the Expert's.

3.3 Batched Disclosure Equilibria

Pareto efficiency requires *interval-optimal experimentation* and *prompt disclosure*. I now introduce *Batched Disclosure equilibria*, which are characterized by these two properties. These properties formalize the intuitions about inefficiency illustrated by Examples 2 and 3. I begin by outlining key definitions.

Definition 3 (Batching)**.** *The batching induced by π is the ordered partition $\mathcal{P}_\pi$ of M into maximal nonempty sets of consecutive prospects of the same type. Each partition element is called a* batch.

[15] Appendix B.4 verifies equilibrium support for both arrangements and their improvements.

Let n_B denote the number of L-batches, and let $L_\pi^{(i)}$ denote the i^{th} L-batch under π. Write $H_\pi^{(i)}$ for the H-batch immediately following $L_\pi^{(i)}$, when present, and $H_\pi^{(0)}$ for an initial H-batch. Then, π induces the partition

$$\mathcal{P}_\pi = \left(H_\pi^{(0)}, L_\pi^{(1)}, H_\pi^{(1)}, \ldots, L_\pi^{(n_B)}, H_\pi^{(n_B)}\right),$$

with any absent initial or terminal H-batch omitted. The two type-specific subcollections, in their inherited order, are the *ordered L-batches* and *ordered H-batches*.

The Expert is indifferent among equilibrium priority orders that induce the same batching. Her flow payoff is constant within each batch, and experimentation proceeds to the next batch only if every prospect in the current batch fails. Once all H-prospects preceding an L-batch have failed, the Expert is willing to disclose evidence about any unresolved prospect in that batch. She can disclose before the DM chooses its first prospect or after further failures within the batch; for an initial L-batch, she can disclose at the outset. Withholding cannot benefit her, since the DM would still choose the known viable L-prospect before any remaining H. This is the credibility condition in Lemma 2; I call such disclosure *batch-admissible*.

Batching does not fully specify the timing of disclosure: evidence about prospects in the same L-batch may be disclosed together or at different states. I therefore distinguish disclosure points and intervals to describe when disclosure occurs and how the DM experiments between disclosures. Let $D(\kappa)$ denote the disclosed set at κ, conditional on reaching it without earlier disclosure:

Definition 4 (Disclosure Points and Intervals)**.** *Fix an experimentation–disclosure pair (π, ϕ_π).*

(i) *A prospect $\pi(i)$ is a* disclosure point *if the Expert discloses with positive probability upon its on-path failure:*

$$\Pr\left(D(\kappa_\pi^{(i)}) \neq \varnothing\right) > 0.$$

(ii) *An* interval *is a maximal nonempty set of consecutive prospects under π containing no disclosure point except possibly its last prospect.*

I now introduce the two properties that characterize batched disclosure equilibria. The first feature concerns the order of experimentation within these intervals. Failures leave the

DM's beliefs about the remaining prospects unchanged, so he can compare their expected quality and learning rates using his existing information. Consecutive prospects of the same type should be ordered optimally for the DM, retaining the same disclosure if all prospects in the interval fail.

Interval-optimal experimentation. Within each interval, consecutive prospects of the same type are chosen in weakly decreasing order of their adjusted indices. The *adjusted index* is the Gittins index evaluated at the DM's belief after non-disclosure:

$$\widetilde{G}_j(\kappa) = \frac{\widetilde{p}_j(\kappa)}{1-\rho_j[1-\widetilde{p}_j(\kappa)]}.$$

These indices remain constant within the interval before its terminal disclosure.

In Example 3, ℓ_b is more promising than ℓ_a and has the same learning rate. Reversing their order while retaining disclosure after both fail raises the DM's continuation payoff after h fails by 1/200. More generally, the adjusted index balances expected quality and the speed of learning. Exchanging consecutive prospects of the same type when the second has the higher adjusted index improves the DM's payoff without changing the Expert's. The comparison includes the disclosure point: reordering may change which prospect ends the interval. The same disclosure follows failure of all prospects in the interval, and the resulting arrangement remains credible and individually rational.

Prompt disclosure. Every post-initial disclosure point is either an H-prospect or an L-prospect whose individual rationality constraint (1) binds when it is chosen under π. Initial disclosure is unrestricted.

Evidence should be disclosed as soon as doing so is credible: earlier information can benefit the DM without harming the Expert. Further delay is justified only when it is necessary to sustain the DM's incentives along the equilibrium path.

In Example 2, once h fails, evidence about ℓ_b could be revealed before choosing ℓ_a, sometimes avoiding that experiment and raising the DM's continuation payoff from 0.68 to $\approx$ 0.74. Whenever individual rationality is slack, a small amount of disclosure can similarly be brought forward without upsetting the DM's incentives. The Expert loses nothing: the evidence concerns a viable L-prospect that would be chosen before any remaining H-prospect

anyway.

Definition 5 (Batched Disclosure Equilibrium)**.** *A* batched disclosure equilibrium *(BDE) is a deterministic-path MPE with interval-optimal experimentation and prompt disclosure.*

Proposition 3 (Feasible Batched Disclosure Equilibria)**.** *A pair* (π, ϕ_π) *is induced by a BDE if and only if:*

1. ***Individual rationality.*** *At every on-path state with a distinct available action,*

$$V_{\pi,\phi_\pi}(\kappa_\pi^{(i)}) \geq V_{\min}^{\text{dev}}(\kappa_\pi^{(i)}; \pi, \phi_\pi),$$

   with equality whenever the next prospect is an L disclosure point.

2. ***Credible disclosure.*** *On-path disclosure* ϕ_π *is batch-admissible.*

3. ***Within-interval ordering.*** *Experimentation is interval optimal.*

As in Proposition 2, feasibility can be checked using only the proposed order and on-path disclosure. The Expert can sustain any pair satisfying these conditions by cancelling further disclosure after an observed deviation, while preserving promptness and interval optimality. Identifying a BDE therefore does not require solving separately for disclosure and experimentation after every possible deviation.

Proposition 4 (Pareto Efficiency Requires Batched Disclosure)**.** *Every Pareto-efficient MPE with deterministic on-path experimentation is a batched disclosure equilibrium.*

Whenever an MPE with deterministic on-path experimentation fails either property, we can construct a Pareto improvement by extending the logic of Examples 2 and 3. The construction raises the DM's payoff without changing the Expert's payoff or the batching. Repeating these improvements leads to a BDE.

Interval optimality. In Example 3, swapping ℓ_a and ℓ_b allowed the DM to make better use of the available information. More generally, when no disclosure intervenes, efficient experimentation follows the familiar Gittins comparison from single-agent experimentation, balancing expected quality and learning speed. The adjusted indices account for earlier non-disclosure. The Expert is indifferent among priority orders inducing the same batching, so allowing the DM to choose a more efficient order within a batch costs her nothing. If consecutive prospects

of the same type within an interval are in the opposite index order, interchanging them improves the DM's continuation payoff. Each batch retains its members and position, so the batching is unchanged.

When the swap involves a disclosure point, we must also move its disclosure to the end of the reordered interval. The Expert then reveals the same evidence after the same prospects have failed, preserving credibility. Choosing the prospect with the higher index first was already an available deviation. The original incentive constraints therefore also ensure that the DM is willing to continue after it fails in the new order. Thus the improvement in Example 3 extends to the disclosure point itself.

Prompt disclosure. Waiting for an H failure is consistent with prompt disclosure when credibility requires that delay. Example 2 illustrates the payoff gain from earlier disclosure: revealing evidence about ℓ_b as soon as h fails can spare the DM the experiment on ℓ_a. To explain how this improvement extends, consider an L disclosure point ℓ_x and suppose some of the evidence promised at its failure is brought forward to just before it is chosen. The Expert loses nothing. Credibility already requires that, if she withheld this evidence, the viable L-prospect would be chosen before any remaining H-prospect. She therefore has nothing to gain by waiting. If no evidence is disclosed, experimentation follows the original order, so the batching is unchanged.

Bringing all disclosure forward is feasible in Example 2. In general, however, the promise after ℓ_x fails may be needed to induce the DM to choose it. If individual rationality is slack, a small advance preserves that incentive. Earlier constraints remain satisfied: following the order becomes more valuable, while deviating still yields the same payoff without further disclosure. Later constraints are unchanged, since beliefs and subsequent disclosure after ℓ_x fails without disclosure are the same as before. We can therefore advance disclosure until the constraint at ℓ_x binds or none remains after its failure. Repeating this process at the remaining L disclosure points establishes promptness while retaining the improved order.[16]

[16] Appendix B.5 proves the sorting and pre-emption steps and their iteration.

3.4 Refining Batched Disclosure Equilibria

I introduce two refinements of batched disclosure equilibria. The first rules out disclosure after L-failures, motivated by the extensive contingent reasoning such equilibria may require. The second restricts threats to withhold evidence that discourage earlier experimentation on the Expert's favored prospects.

Prompt disclosure can occur initially, after an H failure, or after an L failure where individual rationality binds. In the last case, the promise of evidence may induce the DM to choose an L-prospect early. Revealing that evidence earlier could let him postpone the prospect and choose the Expert's targets sooner, potentially benefiting both players.

Lemma 3 (Gains from Earlier Disclosure)**.** *If individual rationality binds at an L disclosure point in equilibrium, earlier disclosure can strictly increase the DM's payoff without changing the Expert's, holding the experimentation order fixed.*

Consider an L disclosure point ℓ_x. Cancel any disclosure about ℓ_x just before it is chosen and bring forward the evidence promised after its failure. The first change leaves payoffs unchanged; the second can spare the DM an unsuccessful experiment on ℓ_x. The Expert is indifferent because the disclosed L-prospect would otherwise be chosen before any remaining H.[17]

The DM may now prefer to postpone ℓ_x, potentially increasing his payoff further. The Expert is indifferent if it stays in the same batch and benefits if it moves past an H-prospect. But even a revised order that both players prefer may fail to be an equilibrium, as the following example shows.

Example 4 (Sophisticated Prioritization of L Experimentation)**.** *There are five L-prospects and one H-prospect. A BDE prescribes the order*

$$\ell_x, \ell_a, \ell_b, h, \ell_d, \ell_e.$$

The Expert discloses her H-evidence initially, part of her evidence about ℓ_a, ℓ_b initially and the remainder after ℓ_x fails, and her evidence about ℓ_e after h fails along this order. Following initial non-disclosure, ℓ_x has index 0.65*, below those of ℓ_a, ℓ_b, each approximately* 0.70*. Individual ratio-*

[17] Appendix B.6 proves Lemma 3.

nality (1) *binds: the DM is indifferent between choosing ℓ_x and his best deviation without further disclosure.*[18]

The proposed improvement is to disclose all evidence about ℓ_a, ℓ_b initially and postpone ℓ_x until after h, retaining full disclosure about ℓ_e whenever h fails. The proposed order is $(\ell_a, \ell_b, h, \ell_x, \ell_d, \ell_e)$. If followed, both players would be better off: the DM receives useful evidence earlier, and the Expert's favored prospect is chosen sooner. After ℓ_a and ℓ_b fail, however, the DM prefers to choose ℓ_e and continue without further disclosure rather than choose h. The postponed ℓ_x remains available, making that deviation more attractive. In the original equilibrium, ℓ_x had already failed and the DM preferred h. Cancelling further disclosure cannot deter the new deviation, since it relies on no further evidence.

If equilibria may be understood as norms established through communication, then in practice either the Expert or the DM could propose an improved outcome whenever the existing equilibrium involves disclosure after an L-prospect fails and the improvement is sustainable. They could simply discuss and agree to bring disclosure forward, so it no longer depends on that failure.

Example 4 demonstrates that not all such improvements are sustainable: leaving an L-prospect unresolved can subtly change downstream incentives. At the same time, the alternative experimentation path and its potential gains are relatively obvious. Maintaining an equilibrium of this kind therefore requires extensive reasoning about off-path contingencies: both players must recognize why the apparent improvement cannot be sustained without commitment.[19] To the extent that such reasoning may be excessively demanding for either party, it makes sense to rule out strategies where disclosure is tied to L-experiments. When the proposed improvement is sustainable, the original outcome is Pareto dominated; otherwise, maintaining it depends on implausibly sophisticated reasoning about subsequent deviations.

Refinement 1 (Disclosure after H failure). An *h-BDE* is a BDE whose every post-initial disclosure point is an H-prospect. Write $\mathcal{E}_{h\text{-BDE}}$ for this class.

Even an h-BDE may rely on a harsh punishment: the Expert withholds promised evidence

[18] Appendix B.8 specifies the primitives and verifies the original equilibrium and the payoff comparisons below.

[19] Ongoing work examines when equilibria with disclosure after L-failures can be replaced by Pareto-improving equilibria without such disclosure, and when subsequent incentive constraints prevent such improvements.

after an H-failure because the DM chose that prospect earlier than prescribed. Yet this earlier choice benefits the Expert ex post. Example 5 illustrates how such threats can sustain counterintuitive equilibria that both players would likely seek to renegotiate in practice.

Example 5 (Wasteful Punishment of Early H Experimentation)**.** *There are three L-prospects ℓ_x, ℓ_d, ℓ_e and one H-prospect h. A BDE discloses H-evidence initially and follows the experimentation order absent disclosure*

$$\ell_x, h, \ell_d, \ell_e.$$

Evidence about ℓ_e is disclosed only after both ℓ_x and h have failed. Testing h early therefore postpones this disclosure until ℓ_x also fails. Conditional on absence of initial disclosure, the indices satisfy

$$\widetilde{G}_h \approx 0.57, \qquad G_{\ell_x} \approx 0.58, \qquad G_h^+ \approx 0.59, \qquad \widetilde{G}_h < G_{\ell_x} < G_h^+,$$

where G_h^+ includes the disclosure available at the prescribed H boundary. Under this strategy, choosing ℓ_x first is strictly better for the DM.

Releasing the same evidence whenever h fails instead supports the order $h, \ell_x, \ell_d, \ell_e$. Both players gain ex ante.[20]

The threat is credible because withholding evidence after h fails does not cost the Expert. Punishing the DM for a deviation that benefits her may nevertheless be implausible in practice. The second refinement therefore rules out such punishments for the early H experiments specified below.

Refinement 2 (No h-discouragement). If the DM chooses an H-prospect earlier than prescribed, skipping only some of the last L-prospects of the batch before it, the Expert discloses at its failure whatever she would have disclosed had he followed the order. The same restriction applies to each favored prospect in turn: once he has caught up on the skipped prospects, her later promises stand as well. After any other observed deviation, the Expert permanently cancels further L-evidence disclosure.[21]

Restricting disclosure to H failures rules out the L-triggered release in Example 4; no h-

[20] Appendix B.9 gives the primitives and the payoffs, and verifies both complete equilibria.

[21] Definition B.1 in Appendix B.10 specifies the corresponding states and disclosure equalities, including partial advancement and interleaving within H-batches.

discouragement rules out the punishment in Example 5. These restrictions define the class studied below. Such punishments can also help the DM commit to an experimentation order and secure more early disclosure. The paper does not establish that ruling them out preserves the best payoff available among all BDE.

Definition 6 (Refined h-BDE)**.** *An outcome (π, ϕ_π) is induced by a* refined h-BDE *if it is induced by an MPE satisfying no h-discouragement, its equilibrium disclosure occurs only initially or after an H failure, and experimentation is interval optimal. Write $\widehat{\mathcal{E}}_{h\text{-BDE}}$ for this refined subset of $\mathcal{E}_{h\text{-BDE}}$.*

The Expert can no longer deter these early H experiments by withholding promised evidence. Suppose an L-prospect ℓ immediately precedes an H-prospect h. Choosing h first brings forward the evidence promised at its failure, before ℓ has been tested. Let Q^+ be the conditional probability of any such disclosure on the prospects surviving in both orders. The value of this information is included in the *augmented index*

$$G_h^+(Q^+) = \frac{\widetilde{p}_h + \delta_h Q^+}{1 - \delta_h(1 - Q^+)}, \qquad \delta_h = \rho_h(1 - \widetilde{p}_h).$$

A larger Q^+ makes an early test of h more attractive even if h fails. Individual rationality must therefore account for the disclosure that the Expert must retain after an observed deviation. Let $V_{\text{nhd}}^{\text{dev}}$ denote the DM's best payoff from choosing a different prospect first and then experimenting optimally, when the Expert cancels all further disclosure except what no h-discouragement requires.[22]

Proposition 5 (Necessary Conditions for Refined Batched Disclosure)**.** *If (π, ϕ_π) is induced by a refined h-BDE, then:*

1. ***Individual rationality.*** *At every on-path state κ with a distinct available action,*

$$V_{\pi,\phi_\pi}(\kappa) \geq V_{\text{nhd}}^{\text{dev}}(\kappa; \pi, \phi_\pi).$$

2. ***Credible disclosure.*** *Whenever ϕ_π calls for revealing evidence about an L-prospect, π places that prospect before every remaining H-prospect.*

3. ***Within-interval ordering.*** *Experimentation is interval optimal.*

[22] Appendix B.10 defines this continuation under maximum punishment subject to no h-discouragement. Earlier non-disclosure remains informative.

The individual rationality comparison implies that, whenever an L-prospect ℓ immediately precedes an H-prospect h,

$$\widetilde{G}_\ell \geq G_h^+(Q^+). \tag{2}$$

Otherwise the DM would gain by choosing h first and receiving the disclosure at its failure, as shown in Lemma B.3.

In a BDE, the Expert can deter early H experimentation by cancelling further disclosure. The refinement rules out this punishment, so the preceding L-prospect must remain preferable to the DM even when choosing H first would bring forward the promised information.

Unlike the BDE conditions in Proposition 3, these conditions need not be sufficient. The Expert must also remain willing to disclose after an early H failure, given the DM's optimal continuation. They nevertheless narrow the search for the DM-optimal equilibrium. The index ordering results below identify a smaller class for which equilibrium support can be established.[23]

4 DM-Optimal Equilibrium

Disclosure can save the DM time by revealing a viable prospect before he experiments on other prospects. The DM would welcome more evidence ex post, but anticipating its disclosure can change his earlier choices. For example, more disclosure after an H-failure can make choosing H early more attractive, reducing the early L-evidence disclosure that can be sustained. Experimentation priorities and disclosure quantities must therefore be chosen jointly.

Throughout this section, I restrict attention to the refined h-BDE introduced in Section 3.4. An equilibrium is *DM-optimal* if it maximizes the DM's ex ante payoff within this class. Given a focus on h-BDE, the order π determines when on-path disclosure can occur. Evidence about an L-prospect is disclosed, if at all, just before the DM chooses the first prospect in its L-batch. Let σ_ℓ denote the probability of this disclosure conditional on the Expert holding the evidence, and set $\sigma_h = 1$ for initial disclosure of H-evidence. Given π, the vector $\sigma = (\sigma_j)_{j \in M}$ therefore suffices to specify the quantity of disclosure about each prospect on the equilibrium path.

[23] Proposition B.1 in Appendix B.10 retains the complete equilibrium test. Proposition B.2 illustrates why the adjacent bound alone does not generally rule out a profitable deviation.

I first derive necessary properties of the experimentation order. The relevant indices incorporate the information associated with each experiment; when L-prospects have common evidence and learning parameters, an optimum can also be chosen in decreasing L prior order. I then provide an auxiliary bandit interpretation of these indices and use the index ordering results to organize the joint choice of experimentation and disclosure.

4.1 Necessary Properties of Experimentation Priorities

I first derive the ordering conditions that a DM-optimal equilibrium must satisfy for its disclosure probabilities σ. These conditions will then guide the choice of disclosure quantities in Section 4.3: the DM benefits from more evidence whenever it saves experimentation, but both players must remain willing to follow the resulting strategies. The endogenous indices below balance expected quality and speed of learning, as in single-agent experimentation, while accounting for disclosure.

For L-prospects, the ordering comparison uses the adjusted Gittins index after non-disclosure. For H-prospects, each H-batch is effectively treated as a single prospect when comparing experimentation priorities.[24] The Expert cannot credibly disclose L-evidence while another H in the batch would still be chosen first. The batch's augmented index accounts for both its experimentation payoff and the value of the disclosure that follows if all its prospects fail.

For an H-batch B, let $\widetilde{P}_B$ be the DM's expected discounted payoff from testing its prospects in order and stopping if all fail. Let Δ_B be the discounted probability of those failures. The information gained afterward matters when it saves further experimentation. Evidence about the first following L adds no value along the order: if it is good, choosing it immediately already gives the DM the successful flow payoff. Evidence about a later L can instead spare him intervening unsuccessful tests. Let Q_B be the probability of revealing evidence about at least one such later L, conditional on all prospects in B failing without earlier disclosure.[25]

Definition 7 (Batch-Endogenous Index)**.** *For an order π and disclosure probabilities σ, the* batch-endogenous index *is defined as follows.*

[24] If H-prospects are never chosen consecutively, every H-batch already consists of a single prospect.

[25] If no L follows, $Q_B = 0$. Appendix C.1 derives the batch payoff and explains the different disclosure probability used when an H is chosen early.

(i) *For an L-prospect ℓ, it is its adjusted index at the belief after non-disclosure when ℓ is chosen under π:*

$$\widehat{G}_\ell(\pi,\sigma) = \widetilde{G}_\ell.$$

(ii) *Each prospect in an H-batch B has the common index*

$$\widehat{G}_B(\pi,\sigma) = \frac{\widetilde{P}_B + \Delta_B Q_B}{1 - \Delta_B(1 - Q_B)}. \tag{3}$$

Thus $\widehat{G}_h = \widehat{G}_B$ for every $h \in B$.

In a DM-optimal equilibrium, prospects must be chosen in weakly decreasing order of these indices.

Theorem 1 (Global Endogenous Index Ordering)**.** *Let (π,σ) be induced by a DM-optimal refined h-BDE. Then prospects are chosen in weakly decreasing order of their batch-endogenous index:*

$$i < k \quad \Longrightarrow \quad \widehat{G}_{\pi(i)}(\pi,\sigma) \geq \widehat{G}_{\pi(k)}(\pi,\sigma).$$

If the first L following an H-batch has a higher index, choosing it earlier can improve the DM's payoff. It brings forward that prospect's chance of success; if it fails, it delays the remaining H experiments and their promised disclosure. The index comparison accounts for both effects. The proof constructs an equilibrium that realizes the improvement, with both players willing to follow the revised experimentation and disclosure.[26]

The batch-endogenous indices depend on disclosure as well as initial quality. When L-prospects share evidence and learning parameters, disclosure can be arranged so that higher-prior L-prospects are chosen first without reducing the DM's payoff.

Proposition 6 (DM-Optimal Ordering with L-Homogeneity)**.** *Suppose all L-prospects have common μ_L and ρ_L, with priors allowed to differ. There exists a DM-optimal refined h-BDE in which L-prospects are chosen in weakly decreasing order of p_ℓ, both within and across batches.*

Disclosure can reverse the ordering by prior quality: non-disclosure can lower a more promising L-prospect's posterior more sharply. With common μ_L and ρ_L, the Expert can re-

[26] Appendix C.1 proves Theorem 1.

distribute disclosure between an inverted pair so that exchanging their positions leaves the sequence of posteriors, the probability of disclosure, and both players' payoffs unchanged. The proof verifies that the exchange also preserves their incentives. Sorting within each batch in this way, and then applying the index comparisons across H-batches, orders the entire L subsequence by prior.[27]

4.2 An Auxiliary Index Interpretation

The index ordering results connect to the restless-bandit framework of Whittle (1988). In this framework, an arm can change state while passive, and its index is the subsidy that makes activity and passivity equally attractive. Whittle relates this subsidy to the retirement-income interpretation of bandit indices. Applied here, that comparison explains how disclosure changes the DM's willingness to undertake an experiment.

To make the connection, hold the equilibrium order and disclosure strategy fixed. The *retirement value* is the safe payoff that would make the DM willing to give up an experiment and its associated information. Treat each L experiment and each H-batch as an option with value $A + Dv$, where v is its continuation value. For an L-prospect,

$$A = \widetilde{p}_\ell, \qquad D = \rho_\ell(1 - \widetilde{p}_\ell),$$

and for an H-batch B,

$$A = \widetilde{P}_B + \Delta_B Q_B, \qquad D = \Delta_B(1 - Q_B).$$

The H-batch calculation uses Q_B from Definition 7, valuing the first following L before disclosure about that prospect.

In an auxiliary problem, give the DM the option to retire and receive a constant flow payoff s, whose normalized discounted value is also s. Retiring immediately gives s; undertaking the option and then retiring when continuation is needed gives $A+Ds$. The gain from undertaking it is therefore

$$A + Ds - s = (1 - D)\left(\frac{A}{1 - D} - s\right).$$

[27] Appendix C.2 gives the redistribution and its equilibrium implementation.

The DM prefers the option when $s < A/(1-D)$ and retirement when $s > A/(1-D)$. At equality he is indifferent. Thus the batch-endogenous index is precisely the retirement payoff that makes the option just worth undertaking. The index balances expected quality against delay from unsuccessful tests. Faster learning lets the DM move sooner to another prospect after failure. Disclosure changes this comparison: non-disclosure lowers an L-prospect's expected quality, while the evidence promised after an H-batch makes even unsuccessful experiments informative about the alternatives.

An L-prospect's retirement value can therefore change before the DM chooses it. As other experiments fail, promised evidence about that prospect may fail to appear. This is the passive evolution allowed in Whittle's restless-bandit framework. The auxiliary L-arm records these non-disclosure updates; the H-batch arm has a constant value before implementation, already including its promised terminal disclosure. Whittle's subsidy comparison permits later activation of a passive arm. Here, waiting without disclosure can only lower an L-arm's value, while an H-batch's value stays fixed. The subsidy that makes passivity optimal is therefore the same $A/(1-D)$ as in the retirement comparison. This is the arm's *Whittle index*.[28]

Choosing the largest Whittle index need not be optimal in a general restless-bandit problem. Here, the additional issue is that global endogenous ordering compares indices when prospects are chosen, while a later L-prospect awaiting disclosure may still have its higher prior index. Common L evidence and learning parameters allow an optimum with L-prospects in prior order; common learning rates make this also their order of prior indices. The first L of each future batch therefore has the highest prior index in that batch. The preceding H-batch's index is at least as high, so it also bounds every other L awaiting disclosure in the batch. Together with global ordering, these bounds imply the comparison of current indices in the following proposition.

Proposition 7 (Whittle Index Representation)**.** *Suppose L-prospects have common μ_L and ρ_L. Fix a DM-optimal refined h-BDE with weakly decreasing L prior order, as in Proposition 6, and experimentation order π. With disclosure fixed, π selects a remaining auxiliary arm with the highest current Whittle index whenever, following non-disclosure, the DM is to choose an L-prospect or the first prospect in an H-batch. Its Whittle index equals its batch-endogenous index.*

[28] Appendix C.3 constructs the auxiliary arms, verifies their indices and proves Proposition 7.

The DM chooses the remaining prospect or H-batch with the highest current Whittle index, which is also its retirement value. With ties broken according to π and each H-batch's internal order preserved, this reproduces π. These values are themselves sustained in equilibrium. Changing the experimentation order can change what the Expert is willing to reveal, and hence the value of the remaining prospects. The additional assumptions in Proposition 7 delimit this Whittle-index interpretation. The DM's optimization must therefore choose experimentation and disclosure jointly.

4.3 Finding the DM-Optimal Equilibrium

The preceding results establish necessary conditions on disclosure timing and on experimentation for given disclosure probabilities. They leave open how much evidence should be disclosed about each prospect.

With the global ordering in Theorem 1, the on-path conditions in Proposition 5 are also sufficient. Even if further disclosure is cancelled, the ordering lets the DM choose each disclosure target before the remaining H-prospects in an optimal independent order. We can therefore support the candidate while preserving the Expert's willingness to disclose.[29] When the refined class is nonempty, maximizing over these candidates finds the DM-optimal equilibrium.

Practical restrictions might limit initial disclosure or require all available evidence to be revealed once every H-prospect has failed.[30] Here I impose no additional restrictions on disclosure quantities within the refined h-BDE class.

Fix the experimentation order. Revealing that a later L-prospect in a batch is viable lets the DM choose it immediately, avoiding delay from earlier experiments. The Expert is indifferent: even without disclosure, the DM chooses a viable L-prospect before any remaining H-prospect.[31] Thus, at a DM optimum, disclosure that benefits the DM must be full unless an equilibrium constraint prevents even a small increase.

[29] Appendix C.4 constructs the equilibrium using the full on-path $V^{\text{dev}}_{\text{nhd}}$ comparison.

[30] Example 7 considers such a disclosure requirement.

[31] Disclosure about the first L-prospect does not improve the DM's payoff along the prescribed order: he chooses it immediately anyway.

Proposition 8 (Complementary Slackness in Disclosure)**.** *Let (π, σ) be a DM-optimal refined h-BDE, and let j be an L-prospect with strictly positive gain from marginal disclosure. Holding π and the other disclosure probabilities fixed, at least one of the following holds:*

(i) $\sigma_j = 1$.

(ii) *An on-path individual rationality or index-ordering constraint in Proposition 5 or Theorem 1 binds and is violated by arbitrarily small positive increases in σ_j.*

The binding constraint need not be the adjacent L/H comparison. More disclosure can change an earlier choice or the value of a deviation that reorders several prospects.[32]

The search has two parts: choosing disclosure for a fixed experimentation order, and comparing the best outcomes across orders.

1. *Inner loop: the best disclosure for a given order.* Fix π and choose σ to maximize the DM's ex ante payoff subject to Proposition 5 and the global ordering in Theorem 1. Proposition 8 further narrows the choice: wherever more disclosure benefits the DM, either disclosure is full or a constraint binds.

2. *Outer loop: experimentation priorities.* Select the order with the highest payoff among those with a feasible inner problem. Because disclosure changes the indices, the ordering restriction is checked jointly with disclosure in that problem. With common μ_L and ρ_L, Proposition 6 further restricts the search: keep L-prospects in weakly decreasing prior order and vary only the order and placement of the H-prospects.

5 Comparative Statics

Should the DM prefer faster learning, a more knowledgeable Expert, or interests closer to his own? Apparently favorable changes can lower his best equilibrium payoff. Disclosure affects both the value of compliance and the value of deviations from an agreed experimentation order. A change that makes a favored prospect more attractive can tighten the constraints supporting early disclosure about alternatives. The DM may then have to accept less early information or a worse order, and the loss can exceed the direct gain from the change. Com-

[32] Appendix C.5 derives the payoff gain and proves the constraint claim.

plementary slackness identifies a binding obstruction when marginal disclosure is valuable; the obstruction need not be the adjacent boundary comparison alone.

For learning, quality and knowledge, I compare the DM's payoff in the DM-optimal refined h-BDE as one primitive varies, holding the remaining primitives and the definition of the class fixed. Section 5.1 shows that faster learning about an L-prospect and more Expert knowledge about any prospect weakly raise welfare, without common-parameter restrictions (Theorem 2). Section 5.2 presents the H-learning comparison and the L-evidence comparison under maximum tail disclosure. Section 5.3 compares optimized payoffs at zero alignment and at alignment 1/10 for the same fixed menu and refined class. The aligned optimum is strictly lower: the change in the Expert's willingness to withhold evidence eliminates the previous optimum, and the best remaining equilibrium does not recover its payoff. It then considers a higher-quality favored prospect.

5.1 When Information Helps Unambiguously

Theorem 2 (Faster L Learning and More Expert Evidence)**.** *For every $\ell \in M_L$ and $j \in M$, DM-optimal welfare is weakly increasing in λ_ℓ and μ_j, holding all other primitives fixed.*

The prospects may differ in their priors, learning rates and evidence capacities, and there may be several H-prospects. The comparisons have different economic explanations.

Faster L learning reduces the time spent on unsuccessful L experiments. The DM can consequently obtain at least his old payoff by following the old experimentation and disclosure arrangement. More H-evidence also raises the ex ante payoff from that arrangement by making immediate disclosure of a viable H-prospect more likely. Conditional on non-disclosure, an H-prospect becomes less attractive because the Expert would be more likely to disclose evidence about it if it were viable.

In both cases, following the old order need not remain the DM's best response. A new order must also preserve the Expert's willingness to reveal evidence. The proof constructs a new equilibrium in which the DM can reorder experiments, retaining disclosures whose targets are still chosen before any remaining H and removing promises no longer used. The ordering inequalities at the old optimum make these retained releases credible, including

after early H experiments for which no h-discouragement requires continued disclosure. The resulting equilibrium gives the DM at least the payoff from the old arrangement under the new parameter.

The proof that more L-evidence weakly benefits the DM relies on free disposal of information. The Expert can use the old disclosure strategy less often when she has evidence, keeping the probability and timing of revealed evidence unchanged. In particular, when capacity rises from μ_ℓ to $\mu'_\ell > 0$, she follows the old strategy with probability μ_ℓ/μ'_ℓ and otherwise withholds that prospect's evidence. This choice is made once for the prospect. The DM then faces the same information and experimentation problem. The Expert remains willing to disclose whenever the old strategy calls for it: withholding would still lead the DM to choose a viable L-prospect before any remaining H-prospect. At other states, she cannot benefit from revealing L-evidence. Thus every old payoff pair remains attainable. Appendix D.1 gives the constructions and verifies the incentives.

5.2 Information and Disclosure Incentives

Faster H learning makes a favored prospect more attractive. To keep an L-prospect ahead of it, the Expert may have to promise less good news about that alternative: her silence must leave the L-prospect attractive enough to choose first. The loss of early disclosure can exceed the direct gain from faster learning.

Example 6 (Faster H Learning Can Lower DM Welfare)**.** *Consider two L-prospects and one H-prospect with*

$$(p_{\ell_1}, p_{\ell_2}) = (0.6, 0.5), \quad \mu_{\ell_1} = \mu_{\ell_2} = \rho_{\ell_1} = \rho_{\ell_2} = 0.5, \qquad p_h \approx 0.43, \quad \mu_h = 0.1.$$

Increasing ρ_h *from* 0.4 *to* 0.6*, equivalently raising* λ_h/r *from* 2/3 *to* 1.5 *at fixed* r*, strictly lowers DM-optimal welfare.*[33]

At both learning rates, an optimum follows (ℓ_1, ℓ_2, h) absent disclosure. The Expert initially discloses L-evidence only about ℓ_2. Conditional on having this evidence, her probability of disclosure falls from ≈ 0.89 to ≈ 0.33 as h becomes faster. To keep the DM willing to choose

[33] Appendix D.2 gives the exact prior, bounds all orders and disclosure choices, and constructs equilibria attaining both optima.

ℓ_2 before h, its index after non-disclosure must remain at least as high as h's. A higher H index therefore requires a higher posterior on ℓ_2. The Expert achieves this by promising less disclosure: failure to reveal its evidence then conveys less bad news. The order is unchanged, but the DM receives less initial information. This loss outweighs the direct benefit of faster H experimentation.

Disclosure Norms and L-Evidence

Theorem 2 allows the Expert to disclose her L-evidence less often when her evidence capacity is higher. A requirement to disclose all her evidence can prevent this adjustment. Consider a norm requiring all available good news about the L-batch after H to be disclosed when H fails, while initial disclosure remains a choice.

Example 7 (More L Knowledge Under Maximum Tail Disclosure)**.** *Restrict refined h-BDE to disclose all available good news about the L-batch following H when that H fails. With three L-prospects and one H, take*

$$(p_{\ell_1}, p_{\ell_2}, p_{\ell_3}) = (0.8, 0.6, 0.2), \qquad p_h \approx 0.27, \quad \mu_h = 0.1, \quad \rho_j = 0.5 \quad (j \in M).$$

Within this maximum-tail-disclosure class, increasing the common L-evidence probability μ_L from 0.8 *to* 0.9 *strictly lowers DM-optimal welfare. At both values an optimum follows $(\ell_1, \ell_2, h, \ell_3)$ absent disclosure.*[34]

Disclosure about ℓ_3 has no extra value along this order: after h fails, it is the only remaining prospect. It nevertheless makes choosing h early more attractive, because the DM could learn about ℓ_3 before choosing ℓ_2. More evidence about ℓ_3 therefore requires less initial disclosure about ℓ_2 to preserve its priority. This loss lowers the best payoff under the norm. With unrestricted withholding, the Expert could instead retain the old disclosure distribution, as in Theorem 2.

5.3 Alignment

Payoff alignment changes the Expert's willingness to wait before revealing good news. An arrangement that relies on her indifference between disclosure and withholding may cease to

[34] Appendix D.3 gives the exact prior, the all-order upper bounds and attaining equilibria.

be sustainable.

Definition 8 (Alignment)**.** *Replace the Expert's flow payoff by* $\mathbf{1}\{j \in M_H\} + \varepsilon \mathbf{1}\{\theta_j = g\}$*, where* $\varepsilon \geq 0$*, leaving the other primitives fixed. Let* $\mathcal{C}_\varepsilon$ *be the refined h-BDE class under these payoffs, retaining full initial H-evidence disclosure and the stated disclosure restrictions, and allowing mixed experimentation strategies. Both players' incentives are evaluated under the modified payoffs.*

The alignment term can also represent a monetary payment proportional to the DM's payoff: the Expert is rewarded for successful implementation of either an H- or an L-prospect.[35]

Example 8 (Alignment can reduce DM-optimal welfare)**.** *Consider two L-prospects* ℓ_1, ℓ_2 *and one H-prospect* h*, with*

$$(p_{\ell_1}, p_{\ell_2}, p_h) = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right), \qquad (\mu_{\ell_1}, \mu_{\ell_2}, \mu_h) = (0, 1, 0), \qquad (\rho_{\ell_1}, \rho_{\ell_2}, \rho_h) = \left(\frac{2}{5}, \frac{1}{5}, \frac{3}{5}\right).$$

Holding these primitives fixed, let $V^*_{DM}(\varepsilon)$ *be the best DM payoff in* $\mathcal{C}_\varepsilon$*. Then*

$$V^*_{DM}(0) \approx 0.54, \qquad V^*_{DM}(1/10) \approx 0.51 \approx V^*_{DM}(0) - 0.03.$$

At zero alignment, an optimum initially discloses known-good ℓ_2 *with conditional probability* $1/4$ *and follows* (ℓ_1, ℓ_2, h) *absent disclosure. At alignment* $1/10$*, an optimum follows* (h, ℓ_1, ℓ_2)*, with no initial L disclosure and full disclosure of known-good* ℓ_2 *when h fails.*

At zero alignment, an Expert who knows ℓ_2 is good can be indifferent between revealing it and letting the DM choose ℓ_1 first: either continuation reaches a good L before h, giving her zero. Positive alignment changes this comparison. Immediate revelation secures the reward for the DM's success, whereas withholding may delay that reward during an unsuccessful test of ℓ_1. The old partial-disclosure strategy therefore ceases to be a best response.

The same incentive matters after an early failure of h. Both L-prospects remain, and the Expert now strictly prefers revealing known-good ℓ_2 to letting the DM experiment first on ℓ_1. Anticipating this disclosure makes choosing h early more attractive to the DM. In the example, the equilibrium at alignment $1/10$ starts with h and provides no initial L-evidence. The DM loses the early information available at zero alignment, and even his best equilibrium has a

[35] Interpret the payment as an enforceable payoff share that does not change the information observed during experimentation. If financed by the DM, a share $0 \leq \varepsilon < 1$ scales his payoff by $1 - \varepsilon$, leaving his optimal choices unchanged. The example reports his payoff before payments.

lower payoff.[36]

The same alignment arises if the DM pays the Expert a share ε of his payoff from successful implementation. In the example, a 10% share is too small to induce initial L-evidence disclosure, but changes the Expert's incentives enough to eliminate the better equilibrium available without payment. Paying too little can therefore be worse than paying nothing, even before deducting the payment from the DM's payoff.

Quality of Favored Prospects

A higher-quality favorite is more likely to serve the DM's interests, even though the Expert's payoff function is unchanged. It also raises the favorite's index. As with faster H learning, maintaining an L-prospect's priority may then require less early disclosure about it.

Keep the L primitives of Example 6 and fix $\rho_h = 1/2$ and $\mu_h = 1/10$. Increasing p_h from approximately 0.37 to 0.48 lowers DM-optimal welfare from approximately 0.77 to 0.75. An optimum follows (ℓ_1, ℓ_2, h) in both cases, but the Expert's probability of initially revealing evidence about ℓ_2, conditional on having it, falls from ≈ 0.92 to ≈ 0.36. The loss of early L-evidence outweighs both the better H-prospect and the greater chance of initial H-evidence disclosure.[37]

6 Discussion

An experiment can help a decision maker in two ways: it can establish whether the prospect being tested is viable, and it can change what an informed stakeholder is willing to reveal about the alternatives. The second channel can make a biased expert's preferred prospect worth choosing early. Once that prospect fails, evidence that she previously withheld may no longer threaten any payoff she can obtain. The DM's anticipation of this disclosure changes his experimentation priorities.

Batched disclosure equilibria can be interpreted as norms governing this exchange. The DM gives an advocate's preferred prospect a hearing; the Expert supplies evidence about al-

[36] Appendix D.4 proves the comparison of optimized payoffs at these two alignment levels.

[37] Proposition D.1 in Appendix D.2 gives the exact comparison, holding learning and evidence parameters fixed.

ternatives when that prospect fails. Both must find it worthwhile to honor the arrangement as experimentation proceeds. In a policy setting, an advocate's influence can therefore depend on knowledge about interventions outside the advocate's own agenda. Experimentation priorities govern access to existing knowledge as well as the production of new evidence. Evaluating an early trial of an advocate's preferred prospect requires considering what the trial induces the advocate to reveal about the alternatives, alongside what it teaches about the prospect itself. The DM-optimal equilibrium provides a benchmark for a decision maker who can select among sustainable arrangements. Authority over experimentation alone does not ensure that this equilibrium is selected.

The index ordering results describe how information changes the value of that authority. Non-disclosure lowers the value of an L-prospect about which the Expert was expected to disclose evidence; promised evidence at the end of an H-batch raises the value of testing that batch. The batch-endogenous index incorporates both effects. A DM optimum in the refined class must order prospects by these indices, but the indices themselves depend on disclosure. Choosing priorities thus requires considering which releases they can sustain. Common L evidence and learning parameters simplify this choice: an optimum can be selected in decreasing L prior order.

For policymakers, the comparative statics suggest investing in the capacity to learn about and experiment with ideas that industry participants may oppose, perhaps in collaboration with counterparts in similar jurisdictions. Faster learning reduces the cost of exploring these alternatives. Decision makers also stand to benefit from more knowledgeable experts, but the effect can depend on disclosure norms. Even a seemingly reasonable requirement to disclose all evidence after a favored policy fails can make greater expert knowledge reduce early disclosure and leave the decision maker worse off.

Closer alignment of interests need not improve outcomes either. A modest payment for successful implementation can change both when the Expert is willing to disclose and which prospect the DM chooses first. In Example 8, the payment is then too small to induce early disclosure, leaving the DM with less initial information than without payment. Efforts to align incentives can therefore backfire if they are insufficiently strong: paying too little can be worse than paying nothing, even before deducting the payment from the DM's payoff.

Ongoing analysis suggests that the same qualitative forces extend beyond the information structure studied here. Allowing the Expert to hold bad news need not eliminate the exchange of experimentation for favorable evidence about alternatives. It also introduces another mechanism: the Expert may willingly reveal that an L-prospect the DM is about to choose is unviable, saving time before he considers her preferred prospects. Anticipating this help can make choosing the L-prospect less costly. Balanced experimental news may also be unnecessary when "no news is good news." As an experiment continues without news, the chosen prospect becomes more promising, reinforcing the DM's willingness to continue. Failure can still make the Expert willing to disclose evidence about alternatives. I am also exploring nonconclusive evidence, where disclosure improves the DM's assessment without establishing viability for certain. The same incentive to withhold evidence until favored prospects become less promising may remain, although credibility now depends on the experimentation that follows disclosure. Full equilibrium characterizations remain to be developed.

A separate question is how robust disclosure is when the Expert is indifferent about revealing evidence. Pessimism about her knowledge or willingness to disclose need not destroy credibility: if the DM expects less evidence, non-disclosure is less discouraging about an L-prospect's viability. This can preserve his willingness to choose that prospect before a remaining H, leaving the Expert indifferent about disclosure. Such robustness need not preserve the original experimentation path or the DM's payoff. More generally, if the DM might instead choose an H before a prospect whose viability the Expert knows, she may prefer to withhold that evidence. Requiring disclosure to remain credible under a wider range of continuation choices can therefore restrict its timing. In some cases, only tail disclosure, after all H-prospects have failed, may survive; earlier disclosure can remain credible when those choices still prioritize the known viable L-prospect.

Commitment by the DM offers a further benchmark to explore. For example, with the power to delay a known viable L-prospect, he might induce full upfront disclosure by promising the Expert a suitable amount of H implementation. In such an arrangement, the DM would bring forward the H implementation that she would otherwise obtain, skip unconfirmed L-prospects, and choose a known viable L once the promised implementation is complete. Avoiding unsuccessful L experiments could then benefit both players. A weaker com-

mitment would bind experimentation choices only while no viable solution is known. It could allow the DM to follow through on the better experimentation plans discussed in Section 3.4, but could not support promises to delay a known viable L. I am interested in quantifying the welfare loss from no commitment relative to each benchmark.

References

Daron Acemoglu. Directed technical change. *Review of Economic Studies*, 69(4):781–809, 2002. doi: 10.1111/1467-937X.00226.

Philippe Aghion and Jean Tirole. The management of innovation. *Quarterly Journal of Economics*, 109(4):1185–1209, 1994. doi: 10.2307/2118360.

Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999. doi: 10.1111/1468-0262.00022.

Alessandro Bonatti and Johannes Hörner. Collaborating. *American Economic Review*, 101(2):632–663, 2011. doi: 10.1257/aer.101.2.632.

Yeon-Koo Che and Johannes Hörner. Recommender systems as mechanisms for social learning. *Quarterly Journal of Economics*, 133(2):871–925, 2018. doi: 10.1093/qje/qjx044.

Yeon-Koo Che and Navin Kartik. Opinions as incentives. *Journal of Political Economy*, 117(5):815–860, 2009. doi: 10.1086/648432.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982. doi: 10.2307/1913390.

Mathias Dewatripont and Jean Tirole. Advocates. *Journal of Political Economy*, 107(1):1–39, 1999. doi: 10.1086/250049.

Ronald A. Dye. Disclosure of nonproprietary information. *Journal of Accounting Research*, 23(1):123–145, 1985.

John C. Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164, 1979. doi: 10.1111/j.2517-6161.1979.tb01068.x.

Sanford J. Grossman. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24(3):461–483, 1981. doi: 10.1086/466995.

Yingni Guo. Dynamic delegation of experimentation. *American Economic Review*, 106(8): 1969–2008, 2016. doi: 10.1257/aer.20141215.

Peter M. Haas. Introduction: Epistemic communities and international policy coordination. *International Organization*, 46(1):1–35, 1992. doi: 10.1017/S0020818300001442.

Marina Halac, Navin Kartik, and Qingmin Liu. Optimal contracts for experimentation. *Review of Economic Studies*, 83(3):1040–1091, 2016. doi: 10.1093/restud/rdw013.

Peter A. Hall. Policy paradigms, social learning, and the state: The case of economic policymaking in Britain. *Comparative Politics*, 25(3):275–296, 1993. doi: 10.2307/422246.

Emeric Henry and Marco Ottaviani. Research and the approval process: The organization of persuasion. *American Economic Review*, 109(3):911–955, 2019. doi: 10.1257/aer.20171919.

Johannes Hörner and Larry Samuelson. Incentives for experimenting agents. *RAND Journal of Economics*, 44(4):632–663, 2013. doi: 10.1111/1756-2171.12039.

Godfrey Keller and Sven Rady. Strategic experimentation with Poisson bandits. *Theoretical Economics*, 5(2):275–311, 2010. doi: 10.3982/TE595.

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005. doi: 10.1111/j.1468-0262.2005.00564.x.

John W. Kingdon. *Agendas, Alternatives, and Public Policies*. Little, Brown, Boston, 1984.

Jan Knoepfle and Julia Salmi. Dynamic evidence disclosure: Delay the good to accelerate the bad, 2025. URL https://arxiv.org/abs/2406.11728v3. arXiv:2406.11728v3, November 18.

Vijay Krishna and John Morgan. A model of expertise. *Quarterly Journal of Economics*, 116(2): 747–775, 2001. doi: 10.1162/00335530151144159.

Hao Li. A theory of conservatism. *Journal of Political Economy*, 109(3):617–636, 2001. doi: 10.1086/321018.

Andrew McClellan. Experimentation and approval mechanisms. *Econometrica*, 90(5):2215–2247, 2022. doi: 10.3982/ECTA17021.

Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12(2):380–391, 1981.

Michael Mintrom. Policy entrepreneurs and the diffusion of innovation. *American Journal of Political Science*, 41(3):738–770, 1997. doi: 10.2307/2111674.

Michael Mintrom and Phillipa Norman. Policy entrepreneurship and policy change. *Policy Studies Journal*, 37(4):649–667, 2009. doi: 10.1111/j.1541-0072.2009.00329.x.

Dmitry Orlov, Andrzej Skrzypacz, and Pavel Zryumov. Persuading the principal to wait. *Journal of Political Economy*, 128(7):2542–2578, 2020. doi: 10.1086/706687.

Bruno Strulovici. Learning while voting: Determinants of collective experimentation. *Econometrica*, 78(3):933–971, 2010. doi: 10.3982/ECTA8011.

Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 25A:287–298, 1988. doi: 10.2307/3214163.

A Model Details

A.1 Independent Experimentation

With posterior p_j, choosing prospect j until news and continuing with value V after failure yields

$$T_j(V) = p_j + \beta_j V, \qquad \beta_j = \rho_j(1 - p_j).$$

Indifference to a safe payoff c requires $c = T_j(c)$, whose solution is $G_j = p_j/(1 - \beta_j)$. For two prospects,

$$T_i(T_j(V)) - T_j(T_i(V)) = (1 - \beta_i)(1 - \beta_j)(G_i - G_j).$$

Adjacent exchanges therefore establish decreasing-index optimality. The resulting value satisfies

$$V(S) = \max_{j \in S} \{p_j + \beta_j V(S \setminus \{j\})\}, \qquad V(\varnothing) = 0.$$

These inequalities also exclude switching before news. Equivalently,

$$r p_j + \lambda_j [p_j + (1 - p_j) V(S \setminus \{j\}) - V(S)] - r V(S) \leq 0,$$

with equality for a maximizing action. Integrating this discounted Bellman inequality bounds any time-dependent policy by $V(S)$; boundedness and $r > 0$ remove the terminal term. The same argument applies below whenever continuation after failure includes a disclosure.

A.2 Strategies and Bayesian Updating

Let $d_j : \mathcal{K} \to \{0, 1\}$ be a deterministic prescription for prospect j. The Expert independently draws one such map for each prospect, conditional only on its evidence. Actual disclosure at κ is $\{j : d_j(\kappa) = 1, \omega_j = 1\}$. The DM independently draws an action at each continuing state, independently of the Expert's draws and the primitives. Prescriptions remain fixed until news changes the state. Optimality nevertheless allows the Expert joint disclosure deviations, including after earlier concealment, and the DM switching before news.

Fix a sequence of failures κ, and write $\eta \preceq \kappa$ for its visited prefixes. Conditional on qualities, evidence and the drawn maps, failure and non-disclosure along this sequence have likelihood proportional to

$$\prod_{j \in M} \left[\mathbf{1}\{j \notin K(\kappa) \text{ or } \theta_j = b\} \prod_{\eta \preceq \kappa} \mathbf{1}\{\omega_j = 0 \text{ or } d_j(\eta) = 0\} \right]. \tag{A.1}$$

The omitted factors, from choices and balanced news durations, do not depend on the private coordinates. Each displayed factor concerns one prospect, so multiplying by the independent prior and normalizing preserves independence across unresolved prospects. Non-disclosure conditions the existing draw; it does not cause another draw.

Let $\Phi_j(\kappa)$ be the probability, conditional on holding evidence about j, that its drawn map prescribes disclosure at some visited prefix, including κ. This probability is evaluated along

the fixed sequence, including draws that would have ended play earlier. Thus

$$\Phi_j(\emptyset) = \phi_j(\emptyset), \qquad \Phi_j(\kappa) = \Phi_j(\kappa^-) + [1 - \Phi_j(\kappa^-)]\phi_j(\kappa).$$

When the previous cumulative probability is one, the final increment is zero. The complete map still specifies behavior after private concealment. For an unresolved j, the surviving mass is $1 - p_j\mu_j\Phi_j$, of which $p_j\mu_j(1 - \Phi_j)$ carries evidence and $p_j(1 - \mu_j\Phi_j)$ is viable. Consequently

$$\varphi_j(\kappa) = \frac{p_j\mu_j[1 - \Phi_j(\kappa)]}{1 - p_j\mu_j\Phi_j(\kappa)}, \qquad \widetilde{p_j}(\kappa) = \frac{p_j[1 - \mu_j\Phi_j(\kappa)]}{1 - p_j\mu_j\Phi_j(\kappa)}.$$

At an unexpected disclosure, condition the joint evidence/map posterior on precisely the evidence revealed. At zero-probability observations, beliefs must respect verifiable evidence and observed failures.

A.3 Why the Sequence of Failures Matters

The sequence identifies every visited disclosure state, so (A.1) determines beliefs. A smaller state that retains availability cannot do so uniformly over the maintained parameters and strategies. To see this, suppose it merges distinct orders κ, κ' of a proper failed set K. Some proper prefix set S of κ is not a prefix set of κ'. Choose $j \notin K$ with $\mu_j > 0$, and prescribe disclosure about j exactly when the failed set is S. This prescription depends only on the failed set and is therefore measurable under either proposed state representation. Give every available DM action positive probability independently across states. Both failure sequences without disclosure then have positive probability, but

$$\Pr(\theta_j = g \mid \kappa, \text{non-disclosure}) = \frac{p_j(1 - \mu_j)}{1 - p_j\mu_j} < p_j = \Pr(\theta_j = g \mid \kappa', \text{non-disclosure}).$$

The merged state cannot determine the posterior. This is a uniform requirement over parameters and strategies; a particular equilibrium or parameter configuration may admit a smaller state.

B Feasible Equilibrium Outcomes

B.1 Proof of Proposition 1

At every state, disclose held evidence about surviving H-prospects and none about L-prospects. After non-disclosure, let the DM choose in decreasing index order with beliefs $q_h = p_h^-$ and $q_\ell = p_\ell$. Implement a confirmed good prospect forever, preferring H when both types are available; reveal any privately retained H-evidence before terminal L implementation. At exhaustion both payoffs are zero.

Proof. Non-disclosure rules out held H-evidence, and failures separately rule out viability of failed prospects. Independence therefore leaves the survivors' beliefs at q_j after every failure order. With no further disclosure, the DM's problem is independent experimentation, solved in Appendix A.1.

Revealing H-evidence secures the Expert's maximal payoff one, including after earlier concealment. Without it, disclosing L-evidence ends play with payoff zero. Couple any disclosure deviation with permanent withholding until its first release of L-evidence. The paths agree until then; thereafter disclosure yields zero and withholding a nonnegative payoff. Withholding is optimal at every private continuation, completing the equilibrium construction. ■

B.2 Proof of Lemma 1

Fix the Expert's mixed strategy ϕ. For the DM's independently mixed policy α, write $S(\kappa) = \{j : \alpha_j(\kappa) > 0\}$ and let $\mathcal{S}(\alpha)$ be the complete deterministic policies choosing from $S(\kappa)$ at every state. Keep terminal responses fixed as above.

Lemma B.1 (Expert best responses survive DM support selection)**.** *If ϕ is a sequential best response to α, including at private continuations after concealment, it remains so against every $a \in \mathcal{S}(\alpha)$.*

Proof. At a private continuation, unused DM prescriptions retain their product law. The Expert's payoff from any fixed continuation is a positive-weight average of numbers in $[0,1]$, one for each supported future policy. An average equal to zero or one forces every component to that endpoint. With H-evidence, optimal behavior attains one. Without it, a prescribed L

disclosure attains zero, so withholding once and resuming must also attain zero. Both comparisons survive support selection. Where disclosure is not prescribed, withholding is nonnegative and release yields zero. Backward induction over remaining failures verifies joint subsequent deviations as well. ■

Lemma B.2 (Selecting a deterministic DM policy)**.** *Every $a \in \mathcal{S}(\alpha)$ paired with ϕ is an equilibrium with the same DM payoff. Some selection weakly improves the Expert's payoff.*

Proof. The DM's action probabilities cancel from the Bayesian filter, so selection preserves beliefs. Each supported action attains the same Bellman value; backward induction preserves that value at every state. The preceding lemma preserves Expert optimality. Finally,

$$U_E(\alpha, \phi) = \sum_{a \in \mathcal{S}(\alpha)} \alpha(a) U_E(a, \phi),$$

so at least one deterministic selection weakly improves her payoff. ■

Proof of Lemma 1. Pareto efficiency makes the selected Expert payoff equal to the original one. Starting at $\kappa_0 = \emptyset$, recursively choose $j_{i+1} = a(\kappa_i)$ and append its failure to obtain $\kappa_{i+1} = (\kappa_i, j_{i+1})$. Progressing search chooses each prospect at most once; success or disclosure ends it. This is the required priority order, with unvisited prospects appended if play ends early. ■

B.3 Proofs of Lemma 2 and Propositions 2–3

Throughout, retain the initial disclosure of held H-evidence and the terminal response above. Given (π, ϕ_π), keep its on-path prescriptions. After an observed deviation, permanently cancel further L-evidence disclosure and let the DM experiment optimally at his inherited posteriors. Cancellation does not reverse earlier inferences.

For later calculations, the conditional probability of disclosure about an unresolved j at a noninitial state is

$$q_j(\kappa) = \frac{p_j \mu_j [\Phi_j(\kappa) - \Phi_j(\kappa^-)]}{1 - p_j \mu_j \Phi_j(\kappa^-)}, \qquad q_j(\emptyset) = p_j \mu_j \phi_j(\emptyset).$$

This follows by conditioning the new disclosure event on earlier non-disclosure. Independence gives the full disclosure probability $Q^+ = 1 - \prod_j (1 - q_j)$. If ℓ is chosen next, the substan-

tive probability Q excludes ℓ and satisfies $Q^+ = q_\ell + (1 - q_\ell)Q$.

Proof of Lemma 2. Revealing H-evidence secures payoff one. Without H-evidence, revealing L-evidence yields zero. If every prescribed target would be chosen before any remaining H after concealment, withholding all disclosures also ends at a known viable L with payoff zero. At other states, withholding has a nonnegative payoff. Backward induction verifies all disclosure deviations, including private continuations after earlier concealment. ■

Proof of Proposition 2. Necessity. The DM can ignore future non-disclosure and follow an optimal independent experimentation policy at his current beliefs, implementing any disclosed viable prospect for value one. His equilibrium value therefore exceeds $V_{\min}$ and hence $V_{\min}^{\text{dev}}$. For credibility, suppose prescribed L-evidence concerns a prospect after some remaining H. An Expert holding only that evidence could conceal it and obtain positive H flow if the intervening L-prospects fail, rather than the zero payoff from disclosure. Interior independent priors make this event possible. Thus a disclosed target must precede every remaining H along the continuation without further disclosure.

Sufficiency. Complete the on-path prescriptions by cancellation as above. The target-before-H property and Lemma 2 verify Expert optimality, including after private concealment. After an observed deviation, the DM optimally experiments without further disclosure. At a prescribed state with next prospect x, a distinct first choice j has value

$$\widetilde{p}_j + \rho_j(1 - \widetilde{p}_j)V_{\min}(A \setminus \{j\}; \widetilde{p}) \leq V_{\min}^{\text{dev}} \leq V_{\pi,\phi_\pi}.$$

Here A is the unresolved set and the surviving posteriors in $V_{\min}$ retain earlier information. These action inequalities, backward induction, and the Bellman argument in Appendix A.1 exclude all experimentation deviations, including switching before news. Thus the completion is an MPE. Applied to the on-path part of any MPE, the same construction preserves its outcome. ■

Proof of Proposition 3. Add promptness and interval optimality to the preceding feasibility test. Both are properties of (π, ϕ_π) and are preserved by its cancellation implementation, so the conditions are necessary and sufficient for a BDE. ■

B.4 Verification of Examples 2 and 3

These examples illustrate Proposition 2. Begin with a continuation after the last H-prospect has failed. All remaining prospects are L, so any disclosure is credible. After an observed deviation from the prescribed order, the Expert cancels further disclosure and the DM experiments optimally. Every learning parameter below is $\rho_j = 1/2$.

Delayed disclosure. Two prospects ℓ_a, ℓ_b have priors $(3/5, 2/5)$ and evidence capacities $(0, 4/5)$. Prescribe ℓ_a before ℓ_b, with all held evidence about ℓ_b revealed only after ℓ_a fails. This disclosure does not change the expected payoff from choosing ℓ_b next. At the initial choice,

$$V_{\ell_a,\ell_b} = \frac{17}{25} > \frac{29}{50} = V_{\ell_b,\ell_a}.$$

The DM therefore strictly prefers to comply, and the delayed disclosure is supported in equilibrium. Individual rationality is slack at the L disclosure point ℓ_a, so disclosure is not prompt.

Instead, reveal the same evidence before ℓ_a is chosen. Absent disclosure, ℓ_b's posterior falls to 2/17, so the DM still prefers ℓ_a first. His payoff before disclosure rises to $93/125 > 17/25$; the Expert's payoff remains zero.

An inefficient order sustained by cancellation. Four prospects have

$$(p_{\ell_a}, p_{\ell_b}, p_{\ell_c}, p_{\ell_d}) = \left(\frac{1}{2}, \frac{51}{100}, \frac{2}{5}, \frac{2}{5}\right), \qquad (\mu_{\ell_a}, \mu_{\ell_b}, \mu_{\ell_c}, \mu_{\ell_d}) = (0, 0, 1, 1).$$

Prescribe $(\ell_a, \ell_b, \ell_c, \ell_d)$ and disclose all held evidence about ℓ_c, ℓ_d after ℓ_a, ℓ_b fail in that order. If neither certificate is revealed, both remaining prospects are bad. The probability of a certificate at the promised disclosure is 16/25.

Although ℓ_b has a higher index than ℓ_a, the promised information sustains ℓ_a first. At entry, compliance is worth ≈ 0.67, exceeding the best payoff without further disclosure, ≈ 0.66, attained by $(\ell_b, \ell_a, \ell_c, \ell_d)$. After ℓ_a fails, choosing ℓ_b with the promised disclosure is worth 1667/2500, whereas choosing ℓ_c or ℓ_d first and then experimenting optimally without disclosure is worth 364/625. After both ℓ_a, ℓ_b fail without disclosure, all remaining values are zero. Thus all on-path individual rationality inequalities hold.

Now prescribe $(\ell_b, \ell_a, \ell_c, \ell_d)$, retaining disclosure after the first two prospects fail. This

raises the DM's payoff by 0.005 without changing the Expert's payoff. The new order is also supported: its entry value exceeds the best distinct-first-action deviation value ≈ 0.66, and after ℓ_b fails, compliance is worth 33/50 rather than 29/50 from either distinct first action without disclosure.

Embedding in the full game. For either example, add an initial H-prospect h with $(p_h, \mu_h, \rho_h) = (9/10, 0, 1/2)$ and begin the specified continuation after its failure. There is no initial L disclosure. The H-prospect has the highest ordinary index, and each continuation is worth at least optimal experimentation without disclosure. Hence the DM also satisfies individual rationality when choosing h. The Expert withholds L evidence before h and is indifferent about disclosure after it fails. Proposition 2 therefore supplies complete equilibrium support for all four outcomes. Each improvement strictly raises the DM's initial payoff and leaves the Expert's payoff at 19/20.

B.5 Proof of Proposition 4

Proof. Use the cancellation implementation from Proposition 2.

Interval ordering. Swap adjacent prospects u, v with $G_u < G_v$, either two L-prospects in the same batch and disclosure interval, or two H-prospects in the same batch. There is no intervening disclosure. Keep the disclosure following completion of their interval. With common post-pair continuation C and $\delta_j = \rho_j(1 - p_j)$,

$$V_{vu} - V_{uv} = (1 - \delta_u)(1 - \delta_v)(G_v - G_u) > 0.$$

The old value exceeds $V_{\min}(\{u, v\} \cup R)$, so individual rationality holds at the new first choice. The old deviation through v implies

$$V_{uv} \geq p_v + \delta_v V_{\min}(\{u\} \cup R).$$

Since $V_{vu} = p_v + \delta_v[p_u + \delta_u C] \geq V_{uv}$, division by $\delta_v > 0$ verifies the new intermediate constraint. Earlier compliance improves without changing deviation values; after both failures, beliefs and continuation coincide. Targets remain before the next H, preserving credibility. The Expert is indifferent: an L pair contributes no flow and the same failure discount, while an H pair maps continuation E to $1 - \delta_u\delta_v + \delta_u\delta_v E$ in either order. Finite adjacent exchanges

therefore sort every interval, preserving batching and support. If a disclosure point changes identity, disclosure still follows failure of every prospect in its interval.

Prompt disclosure. At a nonprompt L disclosure point $x = \pi(k)$, move a fraction ε of its following disclosures to the preceding state:

$$\Phi_j^{\varepsilon,(k-1)} = \Phi_j^{(k-1)} + \varepsilon[\Phi_j^{(k)} - \Phi_j^{(k-1)}], \qquad j \notin K_\pi^{(k)}. \tag{B.1}$$

Keep later levels and the order fixed. No H is crossed, so credibility persists. Strict individual rationality at x survives a small advance by continuity. Later cumulative disclosure and continuation values are unchanged. Earlier posteriors and cancellation values also stay fixed, while compliance improves: when x is bad, advancing a viable alternative saves the wait for its failure; when x is good, both paths yield one. The Expert is indifferent because the affected paths end on L before any H. Effective disclosure and interior failure probabilities make the DM's gain strict. Proposition 2 supports the improvement.

Process positions from last to first, advancing until the relevant individual rationality constraint binds or all disclosure has moved back. Continuity permits each advance; full advancement removes that disclosure point. Any new point lies earlier and is processed next. After finitely many positions, every remaining L disclosure point has binding individual rationality. These advances retain the order, and only remove or split intervals, so prior within-interval sorting persists. Consequently any violation of either BDE property permits a Pareto improvement. ■

B.6 Further Improvements at Binding Disclosure Points

The comparison holds the experimentation order fixed.

Proof of Lemma 3. Cancelling disclosure about the next prospect. Consider a state where the next prospect is $\ell_x \in M_L$. Condition on any unchanged disclosure at this state not occurring. Let p_{ℓ_x} be the DM's belief before disclosure about ℓ_x, and let α be the probability of that disclosure. After non-disclosure the belief is $q_{\ell_x} = (p_{\ell_x} - \alpha)/(1-\alpha)$. If W is the continuation value after ℓ_x fails, then

$$\alpha + (1-\alpha)\Big[q_{\ell_x} + \rho_{\ell_x}(1-q_{\ell_x})W\Big] = p_{\ell_x} + \rho_{\ell_x}(1-p_{\ell_x})W.$$

Thus cancelling this disclosure does not change either player's payoff along the prescribed order. The DM chooses ℓ_x immediately in either case. Disclosure therefore cannot affect the Expert's payoff, whether the prospect is viable or fails. Only this current disclosure is cancelled: earlier non-disclosure remains informative.

Advancing the following disclosure. Suppose an independent set of disclosures is scheduled after ℓ_x fails, with probability $Q > 0$ that some evidence is revealed. Its targets follow ℓ_x in the same L-batch. Write W for the DM's continuation after these disclosures do not occur, and $d_{\ell_x} = \rho_{\ell_x}(1-p_{\ell_x})$. Cancel the current disclosure about ℓ_x and move the disclosures following its failure to the current state. Keeping ℓ_x first, the values before current disclosure in the original and modified arrangements are

$$\begin{aligned} V_{\text{old}} &= p_{\ell_x} + d_{\ell_x}\Big[Q + (1-Q)W\Big], \\ V_{\text{same}} &= Q + (1-Q)(p_{\ell_x} + d_{\ell_x}W), \\ V_{\text{same}} - V_{\text{old}} &= Q(1-p_{\ell_x})(1-\rho_{\ell_x}) > 0. \end{aligned}$$

The strict inequality uses interior $p_{\ell_x}, \rho_{\ell_x}$ and effective disclosure. The Expert's payoff is unchanged. Whenever evidence is brought forward, the DM would otherwise encounter a viable L-prospect before any H; earlier disclosure only saves him the preceding experiment. This comparison starts where ℓ_x is next and retains the information from all earlier non-disclosure. ■

The gain does not establish equilibrium support for the changed order or beliefs. Earlier individual rationality constraints are preserved because compliance improves and deviation values are unchanged; after ℓ_x fails, the remaining information is unchanged. The modified current choice must still be optimal. Example 4 explains why postponing ℓ_x can also upset subsequent incentives.

B.7 Proof of Example 1

Verification of Example 1. At the priors, $G_{\ell_i} = 5/7 > G_h = 15/22$, so a DM experimenting independently tests h last, and $d_{\min}$ discloses H-evidence initially. Conditional on no initial disclosure, $\widetilde{p}_h = 27/97$ and $\widetilde{G}_h = 27/41 < 5/7$, so h remains last under minimum disclosure. Absent disclosure after h fails, each L posterior is $p^-_\ell = 3/23$, and the probability that some

L-evidence is disclosed at that point is $Q^+ = 1-(1-0.54)^2 = 1971/2500$. The conditional DM values are

$$V_{\min} = \frac{3}{5} + \frac{4}{25}\left[\frac{3}{5} + \frac{4}{25}\cdot\frac{27}{97}\right] \approx 0.70,$$
$$V_{DM}^{\text{tail}} = \frac{27}{97} + \frac{56}{97}\left[\frac{1971}{2500} + \frac{529}{2500}\left(\frac{3}{23} + \frac{8}{23}\cdot\frac{3}{23}\right)\right] \approx 0.75,$$

and adding the initial-disclosure branch of probability 3/100 gives $U_{DM}^{\min} \approx 0.71$ and $U_{DM}^{\text{tail}} \approx 0.76$.

Disclose held H-evidence initially and all surviving L-evidence whenever h has failed, also following earlier concealment. Conditional on no initial disclosure, the initial DM action values are $V_h \approx 0.75 > V_{\ell_1} = V_{\ell_2} \approx 0.70$. After a deviation testing either L first and observing failure, testing the remaining L gives $1563/2425 \approx 0.645$, versus $303/485 \approx 0.625$ from testing h; disclosure about a lone L has no additional value. After h fails and no evidence is disclosed, both surviving L posteriors are 3/23, so either order is optimal. This specifies optimal choices at every state; the balanced-clock Bellman inequalities also exclude switching before news.

The Expert obtains her maximal payoff one by disclosing H-evidence. Before h fails, withholding L-evidence preserves a nonnegative payoff, whereas disclosure gives zero. After h fails, she is indifferent because only L-prospects remain. These incentives also hold after earlier concealment, verifying the tail-disclosure equilibrium. Her ex ante payoff is ≈ 0.04 under minimum disclosure, where h is reached only after both L-prospects fail, and $11/25 = 0.44$ under tail disclosure. ■

B.8 Proof of Example 4

Use the primitives

j	ℓ_x	ℓ_a	ℓ_b	h	ℓ_d	ℓ_e
p_j	39/200	59/80	59/80	26/61	21/41	21/73
μ_j	0	20/59	20/59	1/2	0	8/15
ρ_j	20/23	1/10	1/10	3/4	3/10	4/5

For a common initial disclosure probability γ on ℓ_a, ℓ_b, let $V(\emptyset;\gamma)$ be the prescribed continuation value and $V^0(\gamma)$ the optimal independent experimentation value, calculated below. Set

$\lambda_j = r\rho_j/(1-\rho_j)$ for any $r > 0$, and define

$$\Lambda = \frac{3200}{21}[V(\emptyset;0) - V^0(0)] \approx 1.39, \qquad \gamma_* = \frac{\sqrt{1+4\Lambda}-1}{2} \approx 0.78.$$

The Expert discloses H-evidence initially. Independently for each $j \in \{\ell_a, \ell_b\}$, her schedule releases its evidence initially with probability γ_* and after ℓ_x fails otherwise. Evidence about ℓ_e is released once $\ell_x, \ell_a, \ell_b, h$ have all failed. These prescriptions persist while their target survives. There is no evidence about ℓ_x or ℓ_d. The prescribed path absent disclosure is

$$\pi = (\ell_x, \ell_a, \ell_b, h, \ell_d, \ell_e).$$

Complete policy and beliefs. Write $K = K(\kappa)$ and $R = M \setminus K$. If both ℓ_a and ℓ_b survive together with ℓ_x, choose ℓ_x, except at $K = \{\ell_d\}$ or $K = \{h, \ell_d\}$, where the choice is ℓ_a. Otherwise choose ℓ_a if it survives, then ℓ_b if it survives. When neither survives, use the table below; a singleton survivor is chosen and exhaustion gives zero. Confirmed good prospects receive the terminal response specified in the model.

R	Action	R	Action
$\{\ell_x, h, \ell_d, \ell_e\}$	ℓ_e	$\{h, \ell_d, \ell_e\}$	h
$\{\ell_x, \ell_d, \ell_e\}$	ℓ_x	$\{\ell_d, \ell_e\}$	ℓ_d
$\{\ell_x, h, \ell_e\}$	ℓ_e	$\{h, \ell_e\}$	ℓ_e
$\{\ell_x, \ell_e\}$	ℓ_e	$\{\ell_x, h, \ell_d\}$	ℓ_x
$\{h, \ell_d\}$	ℓ_d	$\{\ell_x, \ell_d\}$	ℓ_x
$\{\ell_x, h\}$	ℓ_x		

Let $\Phi_j(K)$ be cumulative eligibility under this schedule and put $\beta_j(K) = 1 - p_j\mu_j\Phi_j(K)$. Persistence makes silence equivalent to the Expert having held no eligible evidence about a surviving target, so

$$\widetilde{p}_j(\kappa) = \frac{p_j - p_j\mu_j\Phi_j(K)}{\beta_j(K)}, \qquad j \notin K.$$

For ℓ_a, ℓ_b the cumulative level is γ_* before ℓ_x fails and one thereafter; for h it is one; for ℓ_e it is one exactly when its four predecessors have failed.

Incentives. Let $V(K)$ denote its value absent disclosure, with $V(M) = 0$. For any available

first action j, set

$$s_j(K) = \prod_{u \notin K \cup \{j\}} \frac{\beta_u(K \cup \{j\})}{\beta_u(K)}, \qquad T_j(K) = 1 - s_j(K) + s_j(K) V(K \cup \{j\}).$$

Backward induction verifies

$$V(K) \geq \widetilde{p}_j(\kappa) + \rho_j(1 - \widetilde{p}_j(\kappa)) T_j(K).$$

Substitution of the specified parameters gives nonnegative margins, with equality for prescribed actions. The identity $\gamma_*^2 + \gamma_* = \Lambda$ reduces every comparison to a linear expression in γ_*. The Bellman inequalities also exclude switching before news.

The Expert reveals H for one. Concealing L-evidence whose disclosure is prescribed gives zero: the policy never tests h while ℓ_a or ℓ_b survives, and disclosure about ℓ_e is prescribed only after h fails. Elsewhere withholding gives at least zero. This verifies every private history, including concealment.

Binding individual rationality and sorting. At absence of initial disclosure, the ordinary index order is

$$\widetilde{G}_{\ell_a} = \widetilde{G}_{\ell_b} \approx 0.70 > G_{\ell_e} = \frac{105}{157} > G_{\ell_x} = \frac{13}{20} > G_{\ell_d} = \frac{3}{5} > \widetilde{G}_h = \frac{52}{87}.$$

The alternative route $(\ell_a, \ell_b, \ell_e, \ell_x, \ell_d, h)$ receives no further useful disclosure and attains the ordinary experimentation value V^0. For a symmetric initial release probability γ, direct evaluation of the prescribed path gives

$$(1 - \gamma/4)^2 [V(\emptyset; \gamma) - V^0(\gamma)] = \frac{21}{3200}(\Lambda - \gamma - \gamma^2).$$

It vanishes at γ_*. The independent-optimal route starts with the distinct action ℓ_a, so it also attains $V_{\min}^{\text{dev}}$ and IR at ℓ_x binds. Its subsequent disclosure probability is

$$1 - \frac{9/16}{(1 - \gamma_*/4)^2} > 0.$$

The ordered disclosure intervals are $\{\ell_x\}, \{\ell_a, \ell_b, h\}, \{\ell_d, \ell_e\}$. After ℓ_x fails, both adjusted L indices are 130/193; after h fails, $G_{\ell_d} = 3/5 > \widetilde{G}_{\ell_e} = 49/101$. Thus experimentation is interval optimal, the L disclosure point has binding IR, and this is a BDE under Definition 5.

The proposed postponement. Advance the residual ℓ_a, ℓ_b evidence to initial disclosure, postpone ℓ_x until after h, and release all held ℓ_e evidence immediately after ℓ_a, ℓ_b, h have failed. The proposed path absent disclosure is $(\ell_a, \ell_b, h, \ell_x, \ell_d, \ell_e)$. If followed, it would strictly improve both payoffs:

$$U'_{DM} > U_{DM}, \qquad U'_E - U_E \approx 7.37 \times 10^{-5} > 0.$$

The payoff recursion verifies the strict DM gain.

At the original state (ℓ_x, ℓ_a, ℓ_b), testing h and then ℓ_d, ℓ_e, with full ℓ_e disclosure following h's failure, is better than testing ℓ_e, ℓ_d, h without disclosure:

$$V(h, \ell_d, \ell_e; \text{full disclosure}) - V(\ell_e, \ell_d, h; \text{no disclosure}) \approx 4.74 \times 10^{-4} > 0.$$

After postponement, at state (ℓ_a, ℓ_b), ℓ_x remains available. The corresponding comparison reverses:

$$V(h, \ell_x, \ell_d, \ell_e; \text{full disclosure}) - V(\ell_e, \ell_x, \ell_d, h; \text{no disclosure}) \approx -3.92 \times 10^{-4} < 0.$$

The surviving beliefs are $\widetilde{p}_h = 13/48$, $p_{\ell_x} = 39/200$, $p_{\ell_d} = 21/41$, and $p_{\ell_e} = 21/73$; disclosure about ℓ_e has probability 56/365 and absence of disclosure gives $\widetilde{p}_{\ell_e} = 49/309$. The deviation needs no later disclosure, so withholding evidence after deviation cannot deter it.

B.9 Proof of Example 5

There are four prospects, with $M_H = \{h\}$ and $M_L = \{\ell_x, \ell_d, \ell_e\}$. Their primitives are

j	ℓ_x	h	ℓ_d	ℓ_e
p_j	41/100	4/7	1/4	1/10
μ_j	4/5	1/2	4/5	4/5
ρ_j	1/2	1/2	1/2	1/2.

Both rules disclose all H-evidence and no evidence about ℓ_x or ℓ_d. Their prescriptions for ℓ_e differ:

$$\Phi^o_{\ell_e}(K) = \mathbf{1}\{\{\ell_x, h\} \subseteq K\}, \qquad \Phi^u_{\ell_e}(K) = \mathbf{1}\{h \in K\}. \tag{B.2}$$

Here $K = K(\kappa)$: these particular strategies depend only on the failed set. Both rules are deterministic and persistent. The old complete policy $a^o(K)$ selects the first surviving member of $(\ell_x, h, \ell_d, \ell_e)$. The new policy is

$$a^u(K) = \begin{cases} h, & h, \ell_e \notin K, \\ a^o(K), & \text{otherwise.} \end{cases} \tag{B.3}$$

Terminal responses are those of the model, including release of retained H-evidence after an L success. After absence of initial H-evidence disclosure, $\widetilde{p}_h = 2/5$; the other beliefs are their priors. Absent disclosure about ℓ_e, $\widetilde{p}_{\ell_e} = 1/46$.

Proof of Example 5. After absence of initial H-evidence disclosure, put $\delta_j = \rho_j(1 - \widetilde{p}_j)$. Then $\delta_{\ell_x} = 59/200$, $\delta_h = 3/10$, and the probability of the ℓ_e disclosure is $Q^+ = 2/25$. Thus

$$\widetilde{G}_h = \frac{4}{7} < G_{\ell_x} = \frac{82}{141} < \frac{\widetilde{p}_h + \delta_h Q^+}{1 - \delta_h(1 - Q^+)} = \frac{106}{181}. \tag{B.4}$$

Under the old rule, the same disclosure follows the failures of both ℓ_x and h in either order. The ordinary swap difference is

$$\widetilde{p}_{\ell_x}(1 - \delta_h) - \widetilde{p}_h(1 - \delta_{\ell_x}) = \frac{1}{200}.$$

Under the new rule, H first gains access to that evidence before testing ℓ_x. Its gain over the unchanged ℓ_x-first route is

$$(1 - \delta_{\ell_x})[1 - \delta_h(1 - Q^+)](G_h^+ - G_{\ell_x}) = \frac{13}{6250}.$$

For complete support, let $\Phi_j(K)$ be the relevant persistent eligibility rule and put $z_j(K) = 1 - p_j\mu_j\Phi_j(K)$. Bayes' rule gives

$$\widetilde{p}_j(K) = \frac{p_j[1 - \mu_j\Phi_j(K)]}{z_j(K)}, \qquad s_j(K) = \prod_{v \notin K \cup \{j\}} \frac{z_v(K \cup \{j\})}{z_v(K)}.$$

With zero value at exhaustion, the value of testing j is

$$B_j(K) = \widetilde{p}_j(K) + \rho_j[1 - \widetilde{p}_j(K)][1 - s_j(K) + s_j(K)V(K \cup \{j\})]. \tag{B.5}$$

Backward induction with $V(K) = B_{a(K)}(K)$ gives a strictly positive advantage over every dis-

tinct available action, for both policies at all states with at least two survivors. Thus both policies are optimal at every state. The balanced-clock Bellman inequalities also bound deviations that switch before news.

An Expert holding H-evidence reveals it for value one. Without H-evidence, she withholds L-evidence before h fails, preserving a nonnegative payoff. Every prescribed L release is after h fails, when all subsequent Expert flows are zero. The prescriptions are therefore optimal at every private history, including earlier concealment. Persistence gives the same posterior after any failure order with the same K. This verifies both complete equilibria.

Conditional on no initial disclosure, the new rule raises the DM's value by 13/6250. Initial H-evidence disclosure occurs with probability 2/7, yielding one for both players. The corresponding ex ante Expert values are

$$\frac{2}{7}+\frac{5}{7}\delta_{\ell_x}(1-\delta_h)=\frac{1213}{2800}, \qquad \frac{2}{7}+\frac{5}{7}(1-\delta_h)=\frac{11}{14}.$$

The gains are 13/8750 and 141/400. Both profiles have only H disclosure points; their L intervals are sorted, so they are BDE. The old rule violates no h-discouragement because $\Phi^o_{\ell_e}(\{h\})=0$ while $\Phi^o_{\ell_e}(\{\ell_x,h\})=1$. The new rule preserves disclosure when the preceding L is omitted. ■

B.10 Proofs of Lemma B.3 and Proposition 5

The refinement preserves promised disclosure after specified early H experiments. I define those states first, derive the protected L/H comparison, and then give the equilibrium test at every state. The final example shows why the adjacent comparison alone can miss a profitable continuation after an early H experiment.

A pair (a,ϕ) specifies the supporting equilibrium strategies. An outcome (π,ϕ_π) belongs to the refined class when some such pair induces it. The outcome notation does not specify the continuation after an observed deviation; the test at every state below verifies that continuation.

Definition B.1 (No-h-discouragement)**.** *Fix the prescribed order π. For each H-prospect h, let P_h be its prescribed predecessor set and let C_h be the ordered L-batch immediately preceding its H-batch; C_h is empty for an initial batch. A state is* protected *if, starting from the initial state $\kappa=\emptyset$,*

every tested L-prospect is the earliest unresolved prospect in π, and every tested H-prospect h has predecessor failure set $P_h \setminus E$ for a suffix E of C_h, possibly empty.

For each target j, impose the following equalities on every complete prescription with positive weight under $\phi^j(1)$. Writing its membership coordinate as $\mathbf{1}\{j \in d(\cdot)\}$, at a protected failure of h retain the same prescription on each common surviving target as at its prescribed failure:

$$\mathbf{1}\{j \in d(\kappa, h)\} = \mathbf{1}\{j \in d(\kappa_\pi^h)\}, \qquad j \notin P_h \cup \{h\},$$

where κ_π^h is the prescribed failure prefix through h. At any protected state whose failed set is a prescribed prefix set, the prescriptions on surviving targets equal those at that prescribed prefix. These are equalities of complete prescriptions, without conditioning on earlier prescribed silence. Coordinates used when evidence is not held are immaterial. Actual disclosure is $d(\kappa) \cap \{j : \omega_j = 1\}$.

There are no other prescribed L-evidence disclosures apart from the prescribed initial disclosure. After an unprotected observed deviation, the Expert cancels all remaining L-evidence disclosure. The DM retains his earlier information and can continue experimenting. With all held H-evidence disclosed initially, every surviving H-prospect belongs to the prescribed set at every entry, so all held H-evidence is disclosed.

For fixed promises under (π, ϕ), the subscript nhd denotes minimum further disclosure subject to no h-discouragement: retain the prescribed-path disclosures and their required protected copies, and cancel all other future L-evidence disclosure. Write $V_{\text{nhd}}(\kappa; \pi, \phi)$, $\tilde{v}_{\text{nhd}}(j; \kappa, \pi, \phi)$ and $V_{\text{nhd}}^{\text{dev}}(\kappa; \pi, \phi)$ for the corresponding optimal, specified-first-prospect and distinct-first-prospect values, with the prescribed-action argument included off path as above.

Protected advancement of an H may omit only a suffix of its own preceding L-batch. Within an H-batch, partial advancement and interleaving retain both type-specific orders. When skipped prospects have failed and the state has a prescribed prefix set, the displayed copy requirements retain the later prescribed disclosures. They restrict the strategy at specified states; κ remains the ordered sequence of observed failures. At an adjacent prescribed L-before-H pair (ℓ, h), call $\widetilde{G}_\ell < G_h^+(Q^+)$ a *flip*, where Q^+ counts every disclosure on the common surviving targets at h's prescribed failure. It counts the entire disclosure event on those targets.

Lemma B.3 (A flip gives a profitable protected deviation)**.** *Under no h-discouragement and H-only post-initial disclosure, a flip gives the DM a profitable deviation: test h first, then, after failure and no disclosure, test ℓ and continue in the prescribed order. Consequently equilibrium experimentation contains no flips.*

After both failures and silence, the protected-copy equalities give the same information and prescribed continuation disclosure events as after the corresponding prescribed prefix. This is a consequence of the specified protection rule, not a restriction of strategies to the unordered failed set.

Proof of Lemma B.3. Advancing h past the immediately preceding L-prospect ℓ is protected. At its failure, common surviving targets face exactly the prescriptions at its prescribed failure. There is no additional prescription on ℓ: it is not a common surviving target, and the refinement permits no additional L-evidence disclosure there. If ℓ then fails without disclosure, the failed set is the prescribed prefix through h. The copy equality at that state retains the prescribed prescriptions. Continuing in the prescribed order gives the same equality at each later prescribed prefix set. Coupling complete maps therefore gives the same accumulated disclosure events, and the same residual laws after silence, throughout this continuation. Strategies remain defined on the state κ.

Put $\delta_j = \rho_j(1-\widetilde{p}_j)$ and let W be the continuation value after both failures and silence. With Q^+ the full conditional disclosure probability on common surviving targets,

$$V_{\ell h} = \widetilde{p}_\ell + \delta_\ell\{\widetilde{p}_h + \delta_h Q^+ + \delta_h(1-Q^+)W\},$$
$$V_{h\ell} = \widetilde{p}_h + \delta_h Q^+ + \delta_h(1-Q^+)\{\widetilde{p}_\ell + \delta_\ell W\}.$$

The common-target disclosure is independent of the qualities of ℓ and h, conditional on the initial silence information, by prospect independence. Hence

$$V_{h\ell} - V_{\ell h} = (1-\delta_\ell)[1-\delta_h(1-Q^+)][G_h^+(Q^+) - \widetilde{G}_\ell] > 0$$

for a flip, contradicting sequential optimality under the unchanged Expert rule.

For a whole H-batch B, write its no-terminal-disclosure experimentation map as $v \mapsto \tilde{P}_B + \Delta_B v$. Each advancement within the batch and the subsequent catch-up are protected;

the same map-by-map copy argument applies. If Q_B^+ is the full terminal disclosure probability on common surviving targets, the gain from advancing the batch before ℓ is

$$(1-\delta_\ell)[1-\Delta_B(1-Q_B^+)]\left[\frac{\tilde{P}_B+\Delta_B Q_B^+}{1-\Delta_B(1-Q_B^+)}-\widetilde{G}_\ell\right].$$

Thus the corresponding whole-batch inequality is necessary as well. If the next prescribed prospect is $z \in M_L$, use $Q_B^+ = q_z + (1-q_z)Q$ at the terminal entry. This protected comparison retains that full event; the H-batch ordering index has its separately defined substantive probability. ■

Proposition B.1 (Complete equilibrium support)**.** *Fix independent disclosure strategies satisfying Definition B.1, including initial H disclosure. Use the cumulative probabilities and posteriors in Appendix A.2. After a prospective failure of x, the conditional probability of non-disclosure is*

$$s_x(\kappa) = \prod_{j \notin K(\kappa)\cup\{x\}} \frac{1-p_j\mu_j\Phi_j(\kappa,x)}{1-p_j\mu_j\Phi_j(\kappa)}.$$

With value zero at exhaustion, define

$$V(\kappa) = \max_{x \notin K(\kappa)} \{\widetilde{p}_x(\kappa) + \rho_x[1-\widetilde{p}_x(\kappa)][1-s_x(\kappa)+s_x(\kappa)V(\kappa,x)]\}. \tag{B.6}$$

A complete deterministic policy a and these strategies support an MPE if and only if, at every continuing state:

1. *$a(\kappa)$ attains the maximum in* (B.6)*;*
2. *whenever a complete disclosure map of positive probability prescribes available evidence about $j \in M_L$, the continuation under a without further disclosure chooses j before any remaining H.*

The resulting equilibrium is a refined h-BDE when its on-path timing and interval ordering satisfy Definition 6.

Proof. The product for s_x follows from prospect independence and the likelihood calculation in Appendix A.2. A disclosure supplies a known viable prospect of value one; otherwise experimentation continues. Thus backward induction gives the Bellman equation, and its inequalities also exclude switching before news.

The second condition is sufficient for Expert optimality by Lemma 2. Conversely, if a prescribed target follows an H in the continuation without disclosure, a type holding only that target's evidence could conceal it until the H is chosen and obtain positive expected flow instead of zero. Interior independent priors make this event possible. This argument includes private continuations after earlier concealment: the test uses the probability of a complete prescription, even if its conditional disclosure probability after earlier prescribed non-disclosure is zero. The maintained map restrictions and on-path conditions then establish refined membership. ■

Proof of Proposition 5. Take a supporting refined h-BDE. The maintained disclosure rule retains exactly the protected promises; after any other observed deviation, the Expert cancels further disclosure. DM optimality therefore implies the on-path comparison with $V^{\text{dev}}_{\text{nhd}}$, which already permits optimal continuation after the first observed deviation. If a prescribed L disclosure concerned a prospect following a remaining H, a type holding only that evidence could benefit by withholding it until the H is chosen. This contradicts Expert optimality. Interval ordering is part of refined h-BDE membership. Lemma B.3 supplies (2) as a consequence of the on-path comparison. These are necessary conditions; complete support is characterized separately by Proposition B.1. ■

Proposition B.2 (An adjacent bound does not suffice for refined support)**.** *With heterogeneous L parameters, minimum-disclosure individual rationality, within-batch index ordering, and the last-L/augmented-H bound are insufficient for refined support, even with one H.*

Proof. There are four L-prospects $\ell_x, \ell_z, \ell_w, \ell_u$ and one H-prospect h. The raw primitives, in the displayed order, are

j	ℓ_x	h	ℓ_z	ℓ_w	ℓ_u
p_j	343/1000	14/17	7/20	1/2	3/20
ρ_j	19/20	3/4	19/20	1/5	1/20
μ_j	1/2	1/2	1/2	13/20	2/5

Disclose all H-evidence initially. Conditional on absence of disclosure, $\widetilde{p}_h = 7/10$. Consider the path $\ell_x, h, \ell_z, \ell_w, \ell_u$: never disclose ℓ_x or ℓ_z, and disclose all held evidence about ℓ_w, ℓ_u im-

mediately after h fails. The probability of a nonempty disclosure is $Q = 731/2000$; absence of disclosure leaves $(\widetilde{p}_{\ell_w}, \widetilde{p}_{\ell_u}) = (7/27, 9/94)$. Here the prescribed next L is ℓ_z and $q_{\ell_z} = 0$, so this full actual disclosure probability also equals substantive Q. The L-batches are sorted, and the preceding-L bound strictly holds:

$$G_{\ell_x} - G_h^+(Q) \approx 8.85 \times 10^{-5} > 0.$$

Let $V_{\min}(A)$ denote optimal independent experimentation on A with the current beliefs and no future disclosure. The prescribed continuation values V_{ℓ_x}, V_h satisfy

$$\begin{aligned} V_{\ell_x} - V_{\min}(\{\ell_x, h, \ell_z, \ell_w, \ell_u\}) &\approx 1.21 \times 10^{-3} > 0, \\ V_h - V_{\min}(\{h, \ell_z, \ell_w, \ell_u\}) &\approx 2.51 \times 10^{-3} > 0. \end{aligned}$$

After silence at that disclosure opportunity, descending independent indices make every remaining IR condition hold. With one H, these comparisons include the original batch-horizon IR requirements. A complete cancellation equilibrium with this path is also verified by exact finite recursion.

Under no-H-discouragement, testing h first must still release the same ℓ_w, ℓ_u evidence. Absent disclosure, the DM can test $\ell_z, \ell_x, \ell_w, \ell_u$, which is in decreasing posterior-index order. The conditional gain is

$$V_{h;\ell_z,\ell_x,\ell_w,\ell_u} - V_{\ell_x} \approx 2.15 \times 10^{-5} > 0.$$

The ex ante gain, including the initial H-evidence disclosure, is $\approx 1.26 \times 10^{-5}$. Only the prescribed boundary disclosures are used in this deviation. Any additional promised information can be ignored, so making later promises contingent on the observed order cannot remove the gain. Likewise, extra evidence on the skipped prospect ℓ_x cannot reduce the value of a feasible policy that ignores that signal. Hence no refinement-compatible completion supports the specified path and prescribed disclosures. ■

The adjacent comparison holds the continuation after ℓ_x, h fixed. The profitable deviation instead obtains the promised disclosure first and then chooses ℓ_z before the skipped ℓ_x. Cancellation can remove future disclosure, but cannot undo that information. It is included in the full $V_{\text{nhd}}^{\text{dev}}$ comparison.

C DM-Optimal Equilibrium

Throughout this appendix and the next, $W_{DM}(\mathcal{C})$ and $W_E(\mathcal{C})$ denote the suprema of the DM's and the Expert's ex ante payoffs over an equilibrium class $\mathcal{C}$; a DM-optimal equilibrium in $\mathcal{C}$ attains the first.

C.1 Proof of Theorem 1

The proof compares an H-batch with the L-prospects immediately before and after it. I first derive the batch value, then construct an equilibrium improvement when the following L has a higher index. The preceding L/H inequality follows from a protected deviation under the original strategy. Together with within-batch sorting, these comparisons give the global order.

For an L-prospect, Definition 7 uses the adjusted index after non-disclosure at its prescribed test:

$$\widehat{G}_\ell(\pi,\sigma) = \widetilde{G}_\ell\Big(\kappa_\pi^{(\pi^{-1}(\ell)-1)}\Big).$$

Fix an H-batch $B = H_\pi^{(i)} = (g_1,\dots,g_s)$, with internal order inherited from π. Suppress (π,σ) and put

$$\delta_h = \rho_h(1-\widetilde{p}_h), \qquad \tilde{P}_B = \sum_{k=1}^{s}\left(\prod_{t<k}\delta_{g_t}\right)\widetilde{p}_{g_k}, \qquad \Delta_B = \prod_{h\in B}\delta_h.$$

An empty product is one. The H beliefs are their post-initial beliefs $\widetilde{p}_h(\emptyset)$, unchanged by other prospects' failures and disclosure under independence.

Lemma C.1 (The H-Batch Representation)**.** *In an h-BDE, no effective L-evidence disclosure occurs strictly inside B. Experimentation through the batch has continuation map $v \mapsto \tilde{P}_B + \Delta_B v$. Including terminal disclosure gives*

$$v \longmapsto \tilde{P}_B + \Delta_B[Q_B + (1-Q_B)v],$$

where the first following L's own release is integrated out and that prospect is valued at its belief before the terminal disclosure. The resulting ordering index is (3)*. For fixed terminal disclosure and continuation, the batch's internal order maximizes value by decreasing ordinary H index.*

Proof. An L release strictly inside the batch would give a type holding only that prospect's evidence zero. Concealing instead reaches an H before any L test and gives positive Expert flow. Sequential incentives therefore exclude such a release.

Composing $v \mapsto \widetilde{p}_h + \delta_h v$ gives $\tilde{P}_B + \Delta_B v$. Each map sends one to a number strictly below one, so $\tilde{P}_B + \Delta_B < 1$ and the index denominator is positive. Adjacent H interchanges with a fixed terminal rule have a common continuation coefficient and a reward difference with the sign of the ordinary-index difference.

Let ℓ be the first following L, with belief p before terminal disclosure and actual release probability a. Conditional on no disclosure about other prospects, its own release and first test give

$$a + (1-a)\left[\frac{p-a}{1-a} + \frac{\rho_\ell(1-p)}{1-a}T\right] = p + \rho_\ell(1-p)T.$$

Disclosure about the other prospects has probability Q_B. This gives the stated map without changing the specified disclosure rule or the DM's individual action choices. If no L follows, $Q_B = 0$. ∎

The exclusion of the first following L's own evidence applies to valuation along the prescribed order. If an H is tested early, that evidence can spare skipped L tests. The protected-deviation comparison in Proposition 5 therefore uses the full disclosure probability Q^+ in $G_h^+(Q^+)$, including that evidence.

For a specified terminal release probability Q^+, write

$$\Gamma_B(Q^+) = \frac{\tilde{P}_B + \Delta_B Q^+}{1 - \Delta_B(1-Q^+)}.$$

This ratio is increasing in Q^+, and $\Gamma_B(Q_B) = \widehat{G}_B$. For a proper suffix B' of B used in a local construction, define $\tilde{P}_{B'}, \Delta_{B'}, \Gamma_{B'}$ by the same formulas with its inherited order. Only whole H-batches enter the global-ordering statement. For a singleton $B = (h)$ the representation reduces to the individual experiment; the separate incentive index G_h^+ continues to use the actual release probability. I use the protection and cancellation requirements of Definition B.1. In particular, a first out-of-order L experiment receives no protected disclosure.

Lemma C.2 (Permanent cancellation after a first observed deviation to L)**.** *Under these require-*

ments, every supported profile has an equivalent supported profile with unchanged prescribed play, disclosure and payoffs, in which the Expert permanently cancels future L-evidence disclosure after a first observed deviation to an L-prospect. All other states can retain their complete profiles.

Proof. After each possible first observed deviation to an L-prospect, replace every subsequent L prescription by withholding, retain immediate H-evidence disclosure, and use independent optimal experimentation at inherited conditional beliefs. These disjoint subtrees contain no protected direct-H boundary. Transforming each target's schedule separately preserves independence and all beliefs before the deviation.

The old DM could follow the same independent policy while ignoring later announcements; good news could only improve it. Thus the values of deviating actions weakly fall, while compliance and other action values stay fixed. Modified continuations are optimal by construction.

There the Expert receives at least zero by withholding L, and one by revealing H. At prescribed states, concealment leaves the unchanged public continuation and disclosure ends search, so neither creates a new private incentive. Continuations following a first observed deviation to an H-prospect, including the Expert's private continuations, remain unchanged. Prescribed payoffs are preserved. ■

Proposition C.1 (An inverted H-batch after an arbitrary prefix)**.** *Under the same refinement, let a supported H-only profile satisfy the preceding-L/H bounds and have order* (κ, B, z, R)*, where* $B = H_\pi^{(i)}$ *is an H-batch and z its first following L. The prefix and continuation may contain arbitrary prescribed initial disclosures, L intervals, H-prospects, and intermediate disclosures. If*

$$G_z > \widehat{G}_B(\pi, \sigma),$$

a supported H-only MPE strictly improves the DM payoff, retains κ*, earlier disclosure, and B's internal order, and inserts z before B is completed. All adjacent L/H bounds hold, and earlier deviation values weakly decrease. If the old profile is a refined h-BDE, so is the new one provided z is inserted inside B or its insertion before B leaves the preceding L interval sorted. The latter holds if that interval is empty or its minimum index is at least* G_z*.*

Proof. Normalize continuations after first observed deviations to L using Lemma C.2. At κ, no prescribed disclosure is effective and every earlier target of prescribed disclosure is resolved, since H is next. Until B is completed, permit its next H or the untested z, with no prescribed L-evidence disclosure; any other action permanently cancels future prescribed disclosures. At completion, copy the old terminal prescribed disclosures about survivors with the same schedule draws. Transport the complete old continuation after B if z survives, or after (B,z) if it failed. H-only timing permits merging the terminal entry and z-failure entry without losing a release.

Put $a = \tilde{P}_B + \Delta_B Q_B$, $d = \Delta_B(1-Q_B)$, $p = p_z$, and $\delta = \rho_z(1-p)$. For the actual old continuation T after B,z, testing z before B instead of afterward gains

$$[p + \delta(a + dT)] - [a + d(p + \delta T)] = (1-\delta)(1-d)[G_z - \widehat{G}_B(\pi,\sigma)] > 0.$$

Choose a deterministic optimum of this finite control problem, including cancellation alternatives. Cancelling before completion obtains no information and is bounded by the old DM's independent-experiment option, hence by his old value. Leaving z after terminal disclosure reproduces that value. The strict improvement must therefore insert z before completion without cancelling.

At the terminal announcement, failed coordinates are exactly those at old B or (B,z). The same draws give the same surviving quality, evidence, and schedule-memory law. In the latter case the old post-z prescribed disclosure only repeats the post-B prescribed disclosure almost surely; suppressing it leaves beliefs unchanged. Old complete public and private support therefore transports. Every target of prescribed disclosure still precedes the next H. Before the terminal entry and after cancellation, withholding L yields at least zero; revealing H yields one.

Protected partial advances and predecessor-L interleavings retain the same draws and relative H order. After terminal silence, append the actual continuation to its canonical old prefix, transporting every later protected boundary. The first H of B is unchanged: if z moves before it, its external direct-expediting counterparts are the old ones skipping a suffix of the preceding L interval; otherwise they are unchanged. Later boundaries retain κ and are supplied internally by transport. Keep the normalized old profile outside this subtree. Ancestor

alternative values are unchanged while compliance improves; earlier targets were already resolved and remain credible.

Normalize these continuations again relative to the selected path, removing any off-path early-z prescribed disclosures without changing its outcome or protected H boundaries. If z precedes a nonfinal H, the feasible reverse interchange with fixed terminal continuation gives $G_z \geq G_h$ by optimality. Before the final H, the augmented interchange gives $G_z \geq G_h^+$; an additional prescribed disclosure about z is neutral when it is tested first after silence. All other bounds are inherited. Internal insertion creates a singleton L interval; insertion before B requires exactly the stated sorting condition for the enlarged interval. ■

Corollary C.1 (Ordering Necessity by BDE Improvement)**.** *Every DM-optimal refined h-BDE satisfies*

$$\widehat{G}_B(\pi, \sigma) \geq G_\ell$$

at each boundary between an H-batch $B = H_\pi^{(i)}$ *and its first following L-prospect* ℓ*.*

Proof. If the inequality fails, Proposition C.1 supplies a complete H-only MPE e' with $U_{DM}(e') > U_{DM}(e)$. Apply the BDE improvement argument to this supported local replacement. Here H-only timing already gives promptness; the possible remaining change is ordering the enlarged preceding L interval. The local refinement-preserving selection retains current-batch disclosure on the protected interleavings and after subsequent catch-up. The complete continuation is transported at each rejoined failure set; equality of protected H action values suffices for the required weak deviation comparisons. After unprotected observed deviations, the Expert still permanently cancels further disclosure, and the individual L/H bounds continue to hold. It therefore yields a refined h-BDE $\widehat{e}$ with

$$U_{DM}(\widehat{e}) \geq U_{DM}(e') > U_{DM}(e),$$

contradicting optimality. The selection is applied to the complete replacement equilibrium, including the ancestor and private-incentive checks. The strict gain survives the subsequent weak improvement. ■

The interval-sorting argument in the proof of Proposition 4 is applied here with the protected continuations retained. Current-batch prescribed disclosures use the same eligibility draws under partial advancement and predecessor-L interleaving, and later promises remain available after catch-up. An omission from an earlier L-batch must be resolved before a different H-batch can be advanced with protection.

Proof of Theorem 1. For an H-batch $B = H_\pi^{(i)}$, write $J = \widehat{G}_B(\pi, \sigma)$ for a batch's index. If its first following L-prospect ℓ has $G_\ell > J$, set

$$a = \tilde{P}_B + \Delta_B Q_B, \quad d = \Delta_B\big(1 - Q_B\big), \quad p = p_\ell, \quad \delta = \rho_\ell(1-p).$$

With T the common candidate continuation after the batch, ℓ, and absence of disclosure, the interchange gain is

$$p + \delta(a + dT) - [a + d(p + \delta T)] = (1-\delta)(1-d)(G_\ell - J) > 0.$$

Proposition C.1 supports a strict improvement at this boundary, including earlier states and later disclosure. Corollary C.1 supplies a weakly better refined h-BDE and thus the contradiction. Consequently $J \geq G_\ell \geq \widetilde{G}_\ell$.

At a preceding-L boundary, Lemma B.3 and its whole-batch comparison give

$$\widetilde{G}_x \geq \frac{\tilde{P}_B + \Delta_B Q^+}{1 - \Delta_B(1 - Q^+)} \geq J,$$

where Q^+ is the full actual terminal disclosure probability. If ℓ follows the batch, $Q^+ = q_\ell + (1 - q_\ell)Q_B$ at that entry. The second inequality follows because Q^+ includes the first following L-prospect's own disclosure, which is excluded from the normalized batch index. A violation of the first inequality would be a profitable deviation under the unchanged rule: after the batch and x fail, the DM resumes the same protected continuation.

Within each L interval, interval optimality gives decreasing adjusted indices. Concatenating these inequalities with both boundary comparisons establishes weak descent of the entire sequence, including any initial and terminal L intervals. Every comparison retains individual experiments and interruptions; grouping changes only the continuation-value calculation. ■

C.2 Proof of Proposition 6

Lemma C.3 (Sorting homogeneous L labels within a batch)**.** *Consider a batch of L-prospects with common evidence probability $\mu \in [0,1]$ and failure discount $\rho \in (0,1)$. Before the DM chooses the first prospect in the batch, evidence about prospect j is disclosed with probability $\mu\sigma_j$ conditional on that prospect being good, independently across prospects. Write*

$$\beta_j = 1 - p_j\mu\sigma_j, \qquad \widetilde{p}_j = \frac{p_j(1-\mu\sigma_j)}{1-p_j\mu\sigma_j}.$$

Assume $0 < p_j < 1$ and that the prescribed order decreases in $\widetilde{p}_j$. There is a redistribution of disclosure and a relabelling of the order that puts these prospects in decreasing prior p_j, while preserving the probability of disclosure before experimentation on the batch and the entire chronological sequence of posteriors absent disclosure. Consequently it preserves both players' on-path payoffs, for any unchanged continuation after the batch.

Proof. If $\mu = 0$, adjusted and original priors coincide and there is nothing to change. Otherwise consider an adjacent inversion: b precedes a, where $p_a > p_b$ but $\widetilde{p}_b \geq \widetilde{p}_a$. Set

$$\widetilde{p}'_a = \widetilde{p}_b, \qquad \widetilde{p}'_b = \widetilde{p}_a, \qquad \sigma'_j = \frac{p_j - \widetilde{p}'_j}{\mu p_j(1-\widetilde{p}'_j)} \quad (j = a, b),$$

and exchange the two labels in the order. A feasible posterior for prior p ranges over $[m(p), p]$, where

$$m(p) = \frac{p(1-\mu)}{1-p\mu}.$$

The function m is nondecreasing. Thus $m(p_a) \leq \widetilde{p}_a \leq \widetilde{p}_b \leq p_b < p_a$ and $m(p_b) \leq m(p_a) \leq \widetilde{p}_a \leq p_b$; both new disclosure probabilities belong to $[0,1]$. Moreover,

$$\beta'_a\beta'_b = \frac{(1-p_a)(1-p_b)}{(1-\widetilde{p}'_a)(1-\widetilde{p}'_b)} = \frac{(1-p_a)(1-p_b)}{(1-\widetilde{p}_a)(1-\widetilde{p}_b)} = \beta_a\beta_b.$$

The probability of no disclosure before experimentation on the batch is therefore unchanged. After silence, the chronological posterior sequence remains $(\widetilde{p}_b, \widetilde{p}_a)$; the common ρ makes its discounted transition law unchanged as well. Both prospects are L, so the Expert's flow payoff and the event of reaching subsequent H-prospects also remain unchanged. Iterating adjacent inversions terminates with decreasing priors. ■

Here Q^+ denotes the probability of any actual disclosure at the entry under discussion; if its prescribed action is $\ell \in M_L$, then $Q^+ = q_\ell + (1 - q_\ell)Q$, with all probabilities evaluated at that entry. The exchange preserves $Q^+ = 1 - \prod_j \beta_j$, not the sum of individual release probabilities. Proposition C.2 verifies complete support. The normalized block index excludes the first following L's own disclosure and can change when that target changes; apply global ordering to the resulting supported optimum.

Corollary C.2 (The cross-batch homogeneous inference)**.** *Suppose a homogeneous-L refined h-BDE has been sorted within each batch as in Lemma C.3. Suppose also that its normalized block-endogenous sequence decreases and that each H-batch index is at least the ordinary index of the first L in its succeeding nonempty batch. Then its L-prospects are in decreasing ordinary index along the whole order.*

Proof. The first L in each batch has its highest ordinary index. The hypothesized H-batch-to-first-L bound therefore bounds every ordinary index in that batch. For an L-prospect ℓ in an earlier batch and j in a later one, the intervening decreasing block-endogenous sequence gives

$$G_\ell \geq \widetilde{G}_\ell \geq \widehat{G}_{H_\pi^{(i)}}(\pi, \sigma) \geq G_j,$$

where $H_\pi^{(i)}$ is the H-batch preceding the later batch. The block grouping accounts for consecutive H-prospects without an intervening L-batch. ■

Retain each target's eligibility event for disclosure before the DM chooses the first prospect in its batch, and at every protected copy under Definition B.1. The Expert retains her later promises once the DM has caught up on the skipped prospects. After an unprotected observed deviation, she permanently cancels further disclosure. Earlier non-disclosure remains informative.

Proposition C.2 (Homogeneous redistribution with complete support)**.** *Let a refined h-BDE have common L parameters μ_L, ρ_L. If every member of every L-batch following an H-batch has an adjusted index at least as high as the ordinary index of every H remaining at the point when the DM would choose the first prospect in that L-batch, the rule above implements its prescribed outcome. Lemma C.3's exchanges preserve complete support and both payoffs under this implementation.*

Proof. A target of prescribed disclosure is chosen before the next H, so its evidence can be disclosed at only one state along the prescribed order. Copying its eligibility event preserves the required prescribed disclosures. Removing optional prescribed disclosures cannot raise a complete deviation value: follow that deviation under the old rule until extra good news supplies value one. Compliance is unchanged; choose Bellman maximizers elsewhere.

After silence at a protected disclosure opportunity, catch-up rejoins a supported prefix before another H-batch. After an unprotected observed deviation, the Expert cancels all future prescribed disclosures. The assumed bounds permit independent sorting with every target of prescribed disclosure before H, so a forbidden first H cannot improve that order. An optimal selection therefore tests targets of prescribed disclosure before H, verifying credibility by Proposition B.1, including after concealment.

Exchange adjacent inverted L labels a, b in the same batch using Lemma C.3. It preserves Q^+ and the chronological posterior sequence. Before the prescribed disclosure, both survive at every protected state; a premature test is forbidden, with unchanged cancellation value when matched by physical target and its prior. Following non-disclosure, match their chronological positions. Posteriors, failure discounts, and all other prescribed disclosures agree. This matching preserves predecessor suffixes, partial H advances, catch-up, and first forbidden moves, including restored later promises. After both fail, all remaining data coincide. Backward induction gives identical complete Bellman problems, then extends backward to states before that disclosure using unchanged Q^+ and cancellation values. Thus every prescribed DM constraint is preserved.

Chronological indices, including dominance for targets of prescribed disclosure, are unchanged, so the same credibility argument gives complete Expert support. The H order, interval sorting, protection equalities, and individual bounds are preserved; the exchange lemma gives equality of both payoffs. ■

Proof of Proposition 6. Start from a DM optimum with internally sorted H-batches. For j in an L-batch following an H-batch, a later H-batch B with predecessor ℓ and first member h_1 satisfies, by global ordering and Lemma B.3,

$$\widetilde{G}_j \geq \widetilde{G}_\ell \geq G^+_{h_1}(Q^+) \geq G_{h_1} \geq G_h \qquad (h \in B).$$

Here immediate $Q^+ = 0$ for a nonsingleton H-batch. Thus the preceding proposition applies. Normalize and remove every within-batch prior inversion by finitely many exact exchanges. For an initial batch the same Q^+ and chronological posteriors preserve payoffs and target-before-H credibility; subsequent states use the same transposition argument. The resulting profile remains a supported optimum.

Apply Theorem 1 and Corollary C.1 to it. Each batch's first L now has its largest prior, so Corollary C.2 orders priors across batches as well. ■

C.3 Proof of Proposition 7

Proof. For each L-prospect and H-batch, construct an auxiliary arm with phases $k = 0, \dots, m$ and absorbing state †. Activation pays flow $u(k)$ and moves to † at rate $\lambda^A(k)$; passivity pays subsidy s and advances to $k+1$ at rate $\lambda^P(k)$, with zero passive rate at m. At †, the available flows are zero and s. Put $\rho^A = \lambda^A/(r + \lambda^A)$ and similarly for ρ^P. The value is

$$V(k;s) = \max\{(1-\rho^A(k))u(k) + \rho^A(k)\max(0,s), (1-\rho^P(k))s + \rho^P(k)V(k+1;s)\}.$$

At the last phase, the second alternative is s. Finite backward recursion and the discounted Bellman inequality verify this value with switching allowed at every time.

Indexability. Suppose $u(k)$ is nonincreasing and $\rho^A(k)$ nondecreasing. For $s \geq 0$, unrolling the recursion yields

$$V(k;s) - s = \max\left\{0, \max_{t \geq k}\left(\prod_{v=k}^{t-1}\rho^P(v)\right)(1-\rho^A(t))(u(t)-s)\right\}.$$

If $s \geq u(k)$ every term is nonpositive. If $0 \leq s < u(k)$, monotonicity and $\rho^P(v) < 1$ make immediate activation strictly better than every delayed positive term. For $s < 0$, backward induction also selects activation: passive flow is negative and later active values $(1-\rho^A(t))u(t)$ cannot increase. Thus

$$V(k;s) = \max\{s, (1-\rho^A(k))u(k) + \rho^A(k)\max(0,s)\}.$$

Passivity is optimal exactly when $s \geq u(k)$, so passive-optimal states expand with the subsidy and the Whittle index is $u(k)$.

Correspondence with experimentation. Phases track successive states without disclosure before the prospect or batch is chosen. For an L with posterior $\widetilde{p}_j(k)$, set

$$\delta_j(k) = \rho_j[1 - \widetilde{p}_j(k)], \qquad u_j(k) = \frac{\widetilde{p}_j(k)}{1 - \delta_j(k)}, \qquad \lambda_j^A(k) = \frac{r\delta_j(k)}{1 - \delta_j(k)}.$$

Non-disclosure lowers the posterior, so u_j falls and $\rho_j^A = \delta_j$ rises. For an H-batch B, use constant parameters

$$u_B = \frac{\widetilde{P}_B + \Delta_B Q_B}{1 - \Delta_B(1 - Q_B)}, \qquad \lambda_B^A = \frac{r\Delta_B(1 - Q_B)}{1 - \Delta_B(1 - Q_B)}.$$

These active values reproduce the affine continuation maps for the L-prospect and the batch. The batch calculation in the proof of Theorem 1 integrates out the first following L's own disclosure. The representation matches discounted values at the fixed r, not the original distribution of experiment durations.

If a passive phase corresponds to experiment x, followed by silence with conditional probability s_x, choose

$$\rho_j^P(k) = \delta_x s_x, \qquad \lambda_j^P(k) = \frac{r\delta_x s_x}{1 - \delta_x s_x}.$$

The factor s_x includes non-disclosure on every target, including the next chosen L. These rates are finite and reproduce the discounted weight of moving to the next silent phase. Use zero passive rate at the last phase. The preceding indexability argument therefore identifies $\widetilde{G}_j$ and $\widehat{G}_B$ as the corresponding Whittle indices.

Replicating the order. Global ordering bounds the implementation indices of all later units by the current unit's index. For a later L-batch that has not yet received its disclosure, its members still have their ordinary indices. Let ℓ_1 be its first prospect and B the preceding H-batch. The stronger boundary inequality in the proof of Theorem 1 and the homogeneous prior order imply

$$\widehat{G}_B \geq G_{\ell_1} \geq G_\ell \quad \text{for every } \ell \text{ in that later batch.}$$

Common ρ_L makes the ordinary index increasing in the prior. The current unit has index at least $\widehat{G}_B$, so it also dominates these not-yet-adjusted indices. Choosing a highest-index unit, resolving ties by π, therefore reproduces the prescribed unit order. Expanding the H-batches

recovers the prospect order. This last comparison is where homogeneity is needed. ■

C.4 Implementing the Candidates Used in Optimization

A candidate consists of an order π and probabilities σ_ℓ of disclosing evidence about each L-prospect just before the DM chooses the first prospect in its batch. All H-evidence is disclosed initially. For each prospect, draw eligibility once, independently across prospects, and use the same draw at every copy required by Definition B.1. After any other observed deviation, the Expert cancels further L-evidence disclosure. This specifies the disclosure rule used to evaluate $V^{\text{dev}}_{\text{nhd}}$ before a supporting DM strategy has been chosen.

Proposition C.3 (Implementation of Globally Ordered Candidates)**.** *Suppose a candidate satisfies the three conditions in Proposition 5 and the global batch-endogenous index ordering in Theorem 1. Then it is induced by a refined h-BDE. Consequently, if that equilibrium class is nonempty, maximizing the DM's payoff over these candidates attains its DM-optimal payoff.*

Proof. On-path optimality. Keep the protected disclosure rule fixed. Backward induction and the on-path individual rationality comparisons make the prescribed action optimal at every prescribed prefix: the prescribed suffix is optimal once its later actions are, and its value dominates every distinct first action followed by optimal continuation under the rule. Thus we can select π at those states, including at ties.

Each disclosure target weakly outranks every remaining H. Consider a target j following non-disclosure, just before the DM chooses the first prospect in its batch. For any later H-batch, let ℓ be the last L immediately before that batch, h its first H, and k any member of it. On-path credibility places j before or at ℓ. Global ordering, the protected adjacent comparison, and interval ordering within the H-batch imply

$$\widetilde{G}_j \geq \widetilde{G}_\ell \geq G^+_h(Q^+) \geq \widetilde{G}_h \geq \widetilde{G}_k.$$

The middle inequality follows from the full on-path individual rationality comparison by Lemma B.3. Every later H-batch has such a preceding L-batch. Hence the adjusted index of every current disclosure target is at least that of every remaining H-prospect. The target's belief stays fixed until it is chosen: disclosure about it is scheduled at only one state along the

prescribed order, and the protected copies repeat the same eligibility event. The H beliefs stay fixed after initial non-disclosure.

Optimal continuation after a protected deviation. A positive L disclosure can occur only after the last H in its batch fails. A protected early failure leaves a suffix E of the preceding L-batch unchosen. The current targets have the same beliefs as when the DM would choose the first prospect in their batch under π. Until E is exhausted, the only action that retains protection is its earliest remaining prospect. No new disclosure occurs during this catch-up. At its end, the failed set is the prescribed prefix set; copying the same eligibility events ensures the same surviving beliefs and future disclosures. Select the prescribed optimal suffix there.

At an intermediate catch-up state, let N be the optimal value without further disclosure and V the optimal value with the retained promises. Then $V \geq N$: the DM can follow an optimal independent policy, using any additional evidence only to implement a confirmed viable prospect. Any action after which the Expert cancels disclosure has value at most N. If $V > N$, an optimal action must therefore continue catch-up. If $V = N$, choose a prospect with the highest independent index, preferring L to H in a tie. Such an L exists because every current disclosure target weakly outranks every remaining H. If this choice retains protection, its value is at least N, so it is optimal. Otherwise the Expert cancels further disclosure after observing the deviation; the DM then continues in independent index order with the same tie-breaking.

Repeating this selection either completes catch-up and resumes π, or leads to cancellation, after which the DM chooses all current disclosure targets before any H. These choices are consistent across states: use π at prescribed prefixes and whenever catch-up restores a prescribed prefix set, the selection just described while skipped prospects remain, and independent index order after cancellation. At other states with no positive L disclosure, select a Bellman maximizer.

Expert incentives and optimization. The selected continuation is optimal for the DM and reaches each disclosed L-prospect before any remaining H if the Expert withholds its evidence. She therefore has nothing to gain by withholding a prescribed disclosure. An additional L disclosure cannot benefit her, and revealing H-evidence attains her preferred outcome. Proposition B.1 verifies complete incentives, including earlier concealment, remembered random-

ization and joint disclosure deviations. The prescribed timing and interval ordering make the resulting equilibrium a refined h-BDE.

Conversely, Proposition C.4 supplies a DM-optimal refined h-BDE whenever the class is nonempty. Its order and disclosure probabilities satisfy Proposition 5 and Theorem 1, so it belongs to the candidate class. Since every candidate has just been implemented, maximizing over that class attains exactly the equilibrium optimum. ■

C.5 Attainment and Proof of Proposition 8

Path payoffs. In the maintained initial-H class, let $V_{DM}(i)$ and $V_E(i)$ be the prescribed continuation values at $\kappa_\pi^{(i)}$ after observing no disclosure. Put $j = \pi(i+1)$ and $\delta_j = \rho_j[1 - \widetilde{p}_j(\kappa_\pi^{(i)})]$. Let s_{i+1} be the conditional probability of no disclosure at the next entry $\kappa_\pi^{(i+1)}$, following j's failure. If that entry prescribes $\ell \in M_L$, then

$$s_{i+1} = [1 - q_\ell(\kappa_\pi^{(i+1)})][1 - Q(\kappa_\pi^{(i+1)})].$$

If it prescribes H, use $s_{i+1} = 1 - Q(\kappa_\pi^{(i+1)})$; at exhaustion put $s_n = 1$. The recursions are

$$V_{DM}(i) = \widetilde{p}_j(\kappa_\pi^{(i)}) + \delta_j[1 - s_{i+1} + s_{i+1} V_{DM}(i+1)],$$
$$V_E(i) = \mathbf{1}\{j \in M_H\}(1 - \delta_j) + \delta_j s_{i+1} V_E(i+1).$$

Both exhaustion values are zero. The Expert's value includes her flow during H experiments that eventually fail. At subsequent disclosure opportunities the only held evidence that can be newly revealed is L-evidence, which gives her zero. Initial disclosures are valued separately: revealing held H-evidence gives the Expert one, while an L-only release gives zero. Any initial disclosure confirms a good prospect and gives the DM one.

Proposition C.4 (Attainment)**.** *If the refined h-BDE class is nonempty, a DM-optimal equilibrium exists.*

Proof. By the deterministic selection argument, fix a complete deterministic DM policy and its induced order. There are finitely many such policies and finitely many complete binary disclosure maps per target. Their mixed probabilities lie in finite simplexes. The refinement imposes fixed support restrictions; credibility excludes maps revealing a target where

the chosen continuation reaches H first. Timing imposes linear equalities. The DM's Bellman inequalities and interval ordering are closed conditions: cumulative disclosure is linear in the weights, and every Bayesian denominator is at least $1 - p_j > 0$, so values and indices are continuous. Each policy's feasible set is therefore compact. Ex ante DM payoff is continuous on it, and maximization over finitely many policies establishes attainment whenever some feasible policy exists. ■

Lemma C.4 (The gain from disclosure before experimentation on a batch)**.** *Fix a maintained H-only disclosure order and all disclosure probabilities except σ_j for an L-prospect j in an L-batch $\mathcal{B}$. Let E be the ordered L prefix before j in that batch. Conditional on silence about the other batch targets, write*

$$V(E \text{ followed by a continuation worth } v) = \tilde{P}_E + \Delta_E v.$$

If V_B is the DM payoff before the Expert can disclose evidence about the batch and before its first prospect is chosen, then

$$\frac{\partial V_B}{\partial \sigma_j} = p_j \mu_j \prod_{k \in \mathcal{B} \setminus \{j\}} (1 - p_k \mu_k \sigma_k)(1 - \tilde{P}_E - \Delta_E).$$

Under the maintained interior priors and $0 < \rho_k < 1$, this derivative is strictly positive if and only if $\mu_j > 0$ and E is nonempty.

Proof. Couple primitive states and independent target eligibility draws. An added prescription changes the outcome only if j is good, its evidence is available, its draw lies in the added interval, and no other batch target is disclosed. Immediate disclosure gives one. Without it, testing E before the known-good j gives $\tilde{P}_E + \Delta_E$. Independence across prospects leaves the prefix posteriors unchanged by conditioning on j's event. Multiplying this payoff difference by the event probability gives the derivative.

For $E = (x_1, \ldots, x_m)$, the usual fixed-order recursion gives

$$1 - \tilde{P}_E - \Delta_E = \sum_{k=1}^{m} \left[\prod_{i<k} \rho_{x_i} (1 - \widetilde{p}_{x_i}) \right] (1 - \widetilde{p}_{x_k})(1 - \rho_{x_k}).$$

Each silence factor is at least $1 - p_k > 0$ and $\widetilde{p}_{x_k} \leq p_{x_k} < 1$. The sum is strictly positive exactly

when the prefix is nonempty. The other factor is positive exactly when $\mu_j > 0$. For an empty prefix, $\tilde{P}_\varnothing = 0$ and $\Delta_\varnothing = 1$. ■

Proof of Proposition 8. A DM-optimal refined h-BDE satisfies global ordering by Theorem 1. Fix its order and all probabilities except $\sigma_j < 1$. Increasing σ_j preserves the prescribed timing and on-path credibility: the Expert still discloses evidence about j just before the DM chooses the first prospect in its L-batch. The same draw is used at every required protected copy.

The remaining candidate restrictions are finitely many on-path individual rationality and index-ordering inequalities. They are continuous in σ_j. In particular, the optimized deviation values are continuous by finite Bellman recursion, and the posterior denominators are at least $1 - p_k > 0$. If no binding inequality were violated by arbitrarily small increases, all inequalities would remain satisfied on some common positive interval. Proposition C.3 would then implement each nearby candidate, allowing its optimal continuation to change. The strictly positive payoff derivative in Lemma C.4, multiplied by positive probability and discounting of reaching the batch, would improve the DM's ex ante payoff. This contradicts optimality. ■

D Comparative Statics

D.1 Proof of Theorem 2

The L-evidence comparison preserves the old disclosure distribution and checks the Expert's incentives with her additional information. For L learning and H evidence, the old order need not remain optimal. I first derive bounds ensuring that the targets of retained disclosure can be tested before every remaining H. I then let the DM reorder experiments in a temporary problem, remove unused disclosures, and restore the refinement's cancellation rule without reducing his payoff. The last step verifies credibility after protected deviations.

More L-evidence. Fix an equilibrium and increase only μ_ℓ to $\mu'_\ell > 0$. Conditional on holding ℓ's evidence, use its old disclosure strategy with probability μ_ℓ/μ'_ℓ and otherwise withhold throughout. Draw once for this prospect, independently of all others. If $\delta_\varnothing$ denotes the always-withhold strategy, then

$$\mu'_\ell(\phi')^\ell(1) + (1-\mu'_\ell)\delta_\varnothing = \mu_\ell \phi^\ell(1) + (1-\mu_\ell)\delta_\varnothing.$$

Thus the joint distribution of qualities and disclosure prescriptions is unchanged, including at off-path states. The same DM policy has the same beliefs and continuation values at every state and remains optimal. If both evidence probabilities are zero, no change is needed.

The additional private information does not upset the Expert's incentives. At every retained positive prescription, concealing the known-good L-target leads to its implementation before any H, as required by the old equilibrium's credibility condition, so disclosure and concealment both give zero. At a prescription to withhold, withholding yields at least zero, including for a newly informed type assigned the always-withhold strategy. Held H-evidence is still disclosed for payoff one. These comparisons apply after earlier concealment and cover joint private deviations. Each retained strategy and the always-withhold strategy obey the protected-copy restrictions. The old payoff pair is therefore attainable in the refined class at μ'_ℓ.

Bounds at an old optimum. For the other comparisons, take a DM optimum, which exists by Proposition C.4. Credibility and H-only timing give each L-target at most one effective disclosure opportunity: initially or at the end of the H-batch preceding its L-batch. Represent it by one independent draw determining whether held evidence is disclosed, with the same draw used at protected states. Let g_t be target t's adjusted index after non-disclosure, and let G_h be an H-prospect's index after initial non-disclosure. For an H-batch B, define its full experiment-and-disclosure map and fixed point by

$$F_B(v) = P_B + \Delta_B\{Q_B^+ + (1 - Q_B^+)v\}, \qquad J_B = \frac{P_B + \Delta_B Q_B^+}{1 - \Delta_B(1 - Q_B^+)}.$$

Here Q_B^+ is the full conditional probability of disclosure at the end of the batch, including evidence about the first following L-prospect. Global endogenous ordering and the protected individual and whole-batch comparisons imply

$$g_t \geq G_h \quad \text{for each remaining } H\text{-prospect } h, \qquad g_t \geq J_B \quad \text{for each later } H\text{-batch } B. \tag{D.1}$$

For the batch bound, compare t with the last L before B, then use the full protected batch-advance inequality. For the individual bound, use the first H of each later batch and its internal index ordering. These arguments use no prior-ordering or common-parameter assumption.

Increasing one ρ_ℓ leaves all H maps fixed and can only raise an L-target's index. Increasing one μ_h lowers its silence belief and ordinary index. For each fixed internal H order, its batch map weakly falls on $[0,1]$. Since sorting the ordinary H indices maximizes the map for every continuation,

$$\max_{\eta} F'_{B,\eta}(v) \leq \max_{\eta} F_{B,\eta}(v) = F_B(v).$$

The newly sorted batch's fixed point therefore cannot rise. In either comparison, the bounds in (D.1) survive.

Allowing the experimentation order to change. Consider a temporary DM control problem that allows more changes to the prescribed order while retaining disclosure. Keep the H-batch sequence and allow any L test while preserving the old disclosures at the ends of those batches, provided their targets survive. Within each batch fix the H order; testing an H outside that sequence cancels all later disclosure. This relaxation lets the DM adjust the L order. The next lemma shows how to choose an optimal continuation that keeps disclosure targets ahead of remaining H-prospects.

Lemma D.1. *At the initial disclosure or a disclosure at the end of an H-batch, condition on non-disclosure. If a target's adjusted index dominates every remaining H index G_h and every remaining batch index J_B, an optimal continuation can test it before every remaining H. The continuation can do so for all targets of that disclosure simultaneously.*

Proof. A sequence with continuation value $A+Dv$ is worth at most g when followed by continuation g if its index is at most g: then $A + Dg - g \leq 0$. The assumed bounds give this property to each complete H-batch and each of its proper prefixes. Inserting an ordinary experiment of index at most g preserves the property. Indeed, write the existing segment as $P \circ S$. If $S(g) > g$, the inserted experiment weakly lowers $S(g)$; if $S(g) \leq g$, its output is at most g and $P(g) \leq g$ gives the bound. Apply this also to each resulting prefix. Deleting promised disclosures lowers the map at $g \in [0,1]$. Concatenation therefore preserves the bound, including ordinary experiments after cancellation.

Take an optimal experimentation order after non-disclosure that tests a target x of the preceding disclosure after an H. Among the L tests through x, choose y with the highest index

at its test. This index is at least g_x. Let q_y denote its quality belief at that test. Moving y before its preceding segment $P(v) = A + Dv$ is permitted and gains

$$q_y(1-D) - A(1-\delta_y) = (1-\delta_y)\{g_y(1-D) - A\} \geq 0, \qquad \delta_y = \rho_y(1-q_y).$$

If that segment promises disclosure about y, condition on whether the Expert holds its evidence and the strategy calls for revealing it. On this event, y is good, so testing it first gives one. Otherwise its belief is its posterior after non-disclosure at the old test. Independence leaves all other prospects unchanged, removes y from the disclosure, and gives the displayed comparison. Thus an optimum can begin with an L test. Repeat after failure until all targets of the preceding disclosure precede any H; repeat after each later batch-end disclosure. ■

Optimize the temporary problem using this selection. Remove promised disclosures whose targets are tested before the corresponding H-batch is completed, as well as disclosures unused after a cancellation. The selected path never used them, so its payoff is unchanged. Deleting information cannot increase the optimal control value: the old rule could reproduce a reduced-rule policy, implementing value one whenever extra good news arrived. Hence the selected path remains optimal.

Merge each H group with no remaining disclosure into the next group with a positive probability of disclosure, and combine trailing groups with no remaining disclosure. Sort the H indices in each merged group. This cannot lower the control value: before its terminal announcement, ordinary index sorting improves a fixed set of experiments; if sorting completes the H group sooner, earlier information can only help. Paths that cancel before the announcement lose no used information. The target bounds also persist: the merged order inserts ordinary H tests into the last positive group's sorted order, so the preceding comparison of continuation maps applies.

Reoptimize and repeat if further promised disclosures are unused. Each repetition removes a prospect's positive disclosure probability at a batch end, so the process terminates. Each retained disclosure is then followed by the tests of all its targets before another H, by Lemma D.1. Initial targets likewise precede every H. Consecutive L tests are index sorted by optimality; H groups are sorted by construction, and each is followed by disclosure with positive probability. The continuation after cancellation is independently sorted. An inversion

where that continuation begins could be swapped without losing a disclosure, contradicting control optimality. The resulting experimentation order is interval optimal.

Restoring equilibrium support. Retain the disclosure promised along this order, using the same randomization at the protected early-H states, and cancel later L disclosure after every other observed deviation. Evidence about a target now tested before the relevant H-batch ends is no longer promised there, so an expedited H creates no new promise. Protected states with nonzero future disclosure preserve the auxiliary H order; any reordering after terminal cancellation has no remaining disclosure to affect. Repeated copies give no new information after silence. The restricted rule removes only off-path information and preserves the chosen path and payoff, so the path remains DM-optimal. Select optimal actions elsewhere.

Credibility must also hold after protected early H experiments. After an H-batch with positive disclosure probability, condition on non-disclosure. The omitted suffix of the preceding L-batch must be caught up before another H can be tested without cancellation. Catch-up rejoins the prescribed continuation, where targets precede H. Cancellation instead permits independent index sorting, which also puts every current target before every remaining H by (D.1). Formally, if the protected value exceeds the independent optimum, an optimal action must continue catch-up. If the values coincide, choose a highest-index L first, breaking ties before H; it either continues catch-up or starts an optimal cancellation policy. Backward induction gives compatible optimal selections at every protected entry, retaining earlier silence.

Releasing held L-evidence and concealing it thus both give the Expert zero at each retained prescription, including after private concealment. Elsewhere withholding yields at least zero, and held H-evidence is disclosed for payoff one. Proposition B.1 supplies the complete private-deviation and between-news verification. The resulting equilibrium satisfies the refinement, H-only timing, and interval ordering.

Welfare. The constructed equilibrium attains at least the old strategy pair's payoff under the new primitive. A higher ρ_ℓ weakly raises that payoff by backward induction in $q + \rho(1-q)T$, with $T \geq 0$. For higher μ_h, couple evidence so old good news remains available. Additional initial H good news gives one; otherwise the payoff is unchanged. Ex ante value weakly rises; conditional-on-silence welfare may fall. Reoptimization proves both comparisons. □

D.2 Proof of Example 6 and the Quality Comparison

The examples require bounds on the best equilibrium payoff, not merely comparisons between particular strategies. With the two L-prospects in Example 6, put

$$q = \frac{p_h(1-\mu_h)}{1-p_h\mu_h}, \qquad \eta = \rho_h(1-q), \qquad g = \frac{q}{1-\eta}, \qquad z = \frac{g}{2-g}.$$

At all parameter values used below,

$$\frac{1}{3} < z < \frac{1}{2}, \qquad q + \frac{3}{4}\eta < \frac{7}{10} + \frac{q}{20}.$$

The optimum, conditional on initial H non-disclosure, is

$$U^* = \frac{9}{10} - \frac{1}{10(1-z)} + \frac{q}{20}. \tag{D.2}$$

Proof. If h is first, full disclosure of both L-prospects after its failure bounds the continuation by 3/4. Thus the h-first payoff is at most $q + 3\eta/4$, below the independent order (ℓ_1, ℓ_2, h). If h is second, the last L retains its prior: earlier disclosure about it would not be credible. Its own disclosure when it is the sole survivor is payoff-neutral. Since its ordinary index exceeds g, moving it before h is profitable. Therefore both L-prospects precede h.

Only initial L disclosure can matter. For initial disclosure probability σ_i conditional on evidence, put $\beta_i = 1 - p_i\mu_i\sigma_i$ and $x_i = 1 - (1-p_i)/\beta_i$. Individual rationality requires $x_i \geq z$, hence $\beta_i \geq (1-p_i)/(1-z)$. Including initial disclosure, the two orders yield

$$U_{12} = \frac{9}{10} - \frac{\beta_2}{5} + \frac{q}{20}, \qquad U_{21} = \frac{9}{10} - \frac{\beta_1}{4} + \frac{q}{20}.$$

The lower bounds on β_i bound both by (D.2). Set $\sigma_1 = 0$ and $\sigma_2 = 4[1 - 1/(2(1-z))] \in (0,1)$ to attain it. After non-disclosure, $(x_1, x_2) = (3/5, z)$; independent index sorting chooses (ℓ_1, ℓ_2, h), breaking the ℓ_2/h tie toward ℓ_2. Use this independent policy at every state and disclose no further L evidence. Concealing prescribed evidence about ℓ_2 still leads to that viable L before h, so the Expert is indifferent. Revealing held H-evidence is optimal, and all protection requirements are satisfied. This establishes the bound and its equilibrium attainment. ■

For Example 6, the exact prior is $p_h = 20/47$. The posterior is $q = 2/5$ and initial H disclo-

sure has probability 2/47 at both rates. Substitution gives

ρ_h	g	z	σ_2	W_{DM}
2/5	10/19	5/14	8/9	182/235
3/5	5/8	5/11	1/3	703/940

The upper-bound conditions hold at both rates, and the difference in optimal payoffs is $5/188 > 0$.

Proposition D.1 (Higher H Quality Can Lower DM Welfare)**.** *With the same L primitives,* $\rho_h = 1/2$ *and* $\mu_h = 1/10$*, raising* p_h *from* $70/187$ *to* $10/21$ *lowers the refined-class DM optimum from approximately* 0.77 *to* 0.75*.*

Proof. The corresponding posteriors are $q = 7/20, 9/20$, and $z = q$. The h-first bounds are $19/32 < 287/400$ and $21/32 < 289/400$, so the preceding proof applies. Initial disclosure probabilities $\sigma_2 = 12/13, 4/11$ attain the bounds. Restoring the initial H branch in (D.2) yields the stated optima, a decrease of $142/7293$. ■

D.3 Proof of Example 7

Proof. Write $m = \mu_L \in \{4/5, 9/10\}$. The exact prior $p_h = 10/37$ implies initial H non-disclosure probability $36/37$, posterior $1/4$ and failure discount $3/8$.

Both high-prior L-prospects must precede h: if either remained after it, choosing h yields at most $1/4 + 3/8 = 5/8$, whereas choosing that L and then h yields at least $13/20$. Prospect ℓ_3 must follow h, since its index is at most $1/3 < G_h = 2/5$. Thus only the orders $(\ell_1, \ell_2, h, \ell_3)$ and its first-two reversal remain. For initial disclosure about $i = 1, 2$, put $\beta_i = 1 - p_i m \sigma_i$. Full disclosure about ℓ_3 after h has probability $Q^+ = m/5$. The protected boundary comparison and interval ordering require

$$g(m) = \frac{10 + 3m}{25 + 3m}, \qquad z(m) = \frac{10 + 3m}{40 + 3m}, \qquad \beta_i \geq \frac{1 - p_i}{1 - z(m)}.$$

Here substantive $Q = 0$: disclosure about the sole tail prospect does not change the prescribed h, ℓ_3 continuation value $13/40$, but matters when h is chosen early. The two conditional initial

payoffs are

$$U_{12} = \frac{1933}{2000} - \frac{\beta_2}{10}, \qquad U_{21} = \frac{1933}{2000} - \frac{\beta_1}{5}.$$

Both are bounded by $1933/2000 - (40 + 3m)/750$.

To attain the bound, disclose initially only about ℓ_2, with $\sigma_2 = (35 - 3m)/(45m)$, so its posterior is $z(m)$. Reveal held evidence about ℓ_3 at the protected failures of h and required catch-up states, using the same eligibility event; cancel after other observed deviations. Select Bellman-maximizing DM actions, breaking ties toward $(\ell_1, \ell_2, h, \ell_3)$. Substitution in (B.6) verifies this path at both parameter values; ℓ_2 and h tie after ℓ_1 fails. Initial disclosure is credible because the known viable ℓ_2 precedes h; all later disclosure occurs after h fails, when the Expert is indifferent. These comparisons also cover concealment and retained information after cancellation. Restoring the initial H branch,

$$W_{DM}^{\text{tail}}(m) = \frac{1}{37} + \frac{36}{37}\left[\frac{1933}{2000} - \frac{40 + 3m}{750}\right].$$

The DM-optimal payoff falls by approximately 3.89×10^{-4}. ■

D.4 Proof of Example 8

Proof. Only ℓ_2 can be the subject of private evidence. Write t for its initial disclosure probability conditional on evidence and $q = (2/5)(1-t)/(1-(2/5)t)$ for its posterior after non-disclosure.

The optimum at zero alignment. If $t > 0$, credibility requires every positive-probability continuation to choose ℓ_2 before h; otherwise concealment could yield positive H flow. A branch beginning with ℓ_1 therefore continues with ℓ_2, whose incentive constraint is $q \geq 1/3$, or $t \leq 1/4$. The unconditional branch payoffs are

$$F_{\ell_1}(t) = \frac{1568}{3125} + \frac{18}{125}t \leq \frac{3361}{6250}, \qquad F_{\ell_2}(t) = \frac{1418}{3125}.$$

If $t = 0$, these two branches yield at most their values at zero; an h-first branch, even with full disclosure after failure, yields at most $317/625 < 3361/6250$. Mixtures are bounded by the same comparisons. The optimum is attained at $t = 1/4$ with no later disclosure and independent index sorting at every state, favoring ℓ_2 over h at their tie. The posterior indices are $10/19, 5/13, 5/13$, so the order is (ℓ_1, ℓ_2, h). The Expert is indifferent to revealing the known

viable ℓ_2 before h and optimal withholding applies elsewhere. Thus $W_{DM}(0) = 3361/6250$.

Every baseline optimizer has this initial disclosure and order. At an initial prescription to withhold, an informed Expert receives $(16/25)\varepsilon$ by waiting through ℓ_1, rather than ε from immediate revelation. Any positive alignment therefore upsets that prescription. This observation alone does not establish a fall in the optimal payoff; the following bound does.

The optimum at $\varepsilon = 1/10$. At state (h), the DM strictly prefers ℓ_1 after non-disclosure: its index is 10/19, whereas ℓ_2's is at most 5/11. Any later disclosure about the sole survivor is payoff-neutral. An Expert holding ℓ_2 therefore strictly prefers immediate revelation, worth ε, to waiting, worth $(16/25)\varepsilon$. Full disclosure at (h) is necessary, including after earlier concealment. With $q \in [0, 2/5]$ the initial posterior, the DM's initial action values must consequently be

$$\begin{aligned} W_h(q) &= \frac{49}{125} + \frac{36}{125}q, \\ W_{\ell_2}(q) &= \frac{56}{625} + \frac{569}{625}q, \\ W_{\ell_1}(q) &= \frac{2}{5} + \frac{6}{25}\max\left\{\frac{1}{25} + \frac{24}{25}q, \frac{1}{5} + \frac{12}{25}q\right\}. \end{aligned}$$

The first continuation in the maximum chooses ℓ_2 before h and is preferred exactly when $q \geq 1/3$. But

$$W_h - W_{\ell_2} = \frac{189 - 389q}{625} > 0, \qquad W_h - W_{\ell_1} = \frac{36q - 11}{625} > 0 \quad (q \geq 1/3).$$

Thus ℓ_2 is never first; whenever ℓ_1 is first, its failure is followed by h. Every mixed branch therefore reaches h before ℓ_2. An informed Expert strictly prefers initial withholding on either branch: waiting through h and revealing at its failure pays $13/25 + (17/25)\varepsilon$ on the h-first branch and $78/625 + (352/625)\varepsilon$ on the ℓ_1, h branch, both greater than ε at 1/10. Hence $t = 0$ in every equilibrium in the class. Substitution of $q = 2/5$ gives action values $317/625, 1568/3125, 1418/3125$ for h, ℓ_1, ℓ_2 respectively. Every equilibrium must choose h first and attain at most 317/625.

To attain this bound, prescribe (h, ℓ_1, ℓ_2) with full ℓ_2 disclosure only at (h) and cancellation

after any other observed deviation. Bellman maximization gives the following action values:

κ	ℓ_1	ℓ_2	h
$\emptyset$	1568/3125	1418/3125	317/625
(ℓ_1)	$--$	53/125	49/125
(ℓ_2)	56/125	$--$	49/125
(h)	2/5	2/25	$--$

After two failures choose the survivor. Initial withholding is optimal by the preceding comparison; revealing at (h) saves the informed Expert the delay of ℓ_1. At (ℓ_1) the DM immediately chooses ℓ_2, so withholding is optimal even though disclosure is cancelled. When ℓ_2 is the sole survivor, revealing and concealing are equivalent. These cases cover every private continuation; no evidence can concern a failed ℓ_2. Balanced news makes a delay in revelation a combination of immediate revelation and waiting for news, while the public Bellman inequalities exclude switching before news. Complete support therefore holds at the modified payoff. The optimal-payoff loss is

$$\frac{3361}{6250} - \frac{317}{625} = \frac{191}{6250} > 0.$$

■

E Latent Information and Independent Pairs

An unobserved common factor can affect evidence generation while inducing exactly the independent quality–evidence pairs assumed in the model. The example below constructs such a factor. Its purpose is to distinguish the observed information structure from a possible underlying source of evidence; arbitrary latent dependence need not have this property. Let X be a latent variable with distribution F_θ conditional on the quality profile. Given (θ, X), draw the evidence indicators independently with

$$\Pr(\omega_j = 1 \mid \theta, X) = \mathbf{1}\{\theta_j = g\} m_j(X), \qquad m_j(X) \in [0,1].$$

Neither player observes X, which has no further effects on payoffs, actions, or experimentation. For two prospects with independent qualities and arbitrary $p_1, p_2 \in (0,1)$, let $X \in \{-2,-1,0,1,2\}$ and use the following rates and conditional probabilities. The law of X depends on θ_1 but not on θ_2.

x	$m_1(x)$	$m_2(x)$	$\Pr(X = x \mid \theta_1 = b)$	$\Pr(X = x \mid \theta_1 = g)$
-2	1/4	5/8	11/60	13/60
-1	3/8	7/16	4/15	2/15
0	1/2	3/8	1/10	3/10
1	5/8	7/16	4/15	2/15
2	3/4	5/8	11/60	13/60

All latent probabilities are positive, and both rates are interior. Although X is correlated with quality and affects both evidence rates, direct calculation gives, for every θ,

$$\mathbb{E}[m_1(X) \mid \theta] = \mathbb{E}[m_2(X) \mid \theta] = \tfrac{1}{2}, \qquad \mathbb{E}[m_1(X) m_2(X) \mid \theta] = \tfrac{1}{4}.$$

If both prospects are good, each evidence profile has probability 1/4. If only one is good, its good news is available with probability 1/2; if neither is good, no evidence exists. Together with the quality prior, this is the joint law of (θ, ω) for independent prospect-specific pairs with $\mu_1 = \mu_2 = 1/2$. The structures therefore give the same histories and payoffs under the same strategies and experiments, including nonseparable disclosure.