When Should AI Stay Silent? Advice Delivery in Turbulent Supply Chains

Wichinpong Park Sinchaisri
Haas School of Business, University of California, Berkeley, parksinchaisri@berkeley.edu

Tanish Kher
University of California, Berkeley, tkher@berkeley.edu

Abstract: Supply-chain disruptions force planners to revise orders while shipments are delayed and recovery is uncertain. AI can recommend an order at every decision, but even useful advice can be worth withholding. We study when an advisor should speak to a planner who retains ordering authority. If displays reduce the willingness to follow later advice, showing a recommendation today uses responsiveness that may be more valuable tomorrow. An optimal rule displays advice when its expected benefit covers the value of preserving that responsiveness. The benefit required for display rises with the remaining opportunities to adjust and with their expected persistence; increasing all current and future benefits proportionally leaves delivery unchanged. Information that distinguishes helpful advice from more helpful advice can be valuable early even when it has no value at the final decision. When following advice consumes capacity needed later, a more responsive planner can optimally receive less advice. In a randomized experiment with full-time managers, orders are revised on 34–43% of decisions with a display and 4–5% without one. In a four-stage supply chain calibrated to these participants, a learned local rule shows 41% as many recommendations as blanket advice (which displays every recommendation) and has lower mean cost than random timing at the same frequency in all five evaluation samples. Managing when AI speaks is part of managing supply-chain resilience.

Key words: supply chain resilience; supply chain disruption; human–AI collaboration; advice delivery; behavioral operations; bullwhip effect

1 Introduction

When a supplier halts production or a port closes, planners must revise orders before they know when deliveries will recover. Supply-chain planning systems now offer AI assistance with identifying shortages and evaluating possible responses.[1] Consider a retailer using such an advisor while awaiting shipments after a factory loses capacity. The advisor recommends a larger replenishment order to cover the retailer's growing backlog. Following that recommendation may protect the retailer, but it can also add to orders waiting at the constrained factory. A few weeks later, the planner may face a more valuable adjustment

[1] For example, SAP Integrated Business Planning supports AI-assisted responses to supplier and plant outages, including finding alternative supply and creating replenishment requisitions. See SAP's "Responding to Supply Chain Disruptions Using AI-Assisted Scenario Planning", release 2608 (2026).

as capacity returns. An advisory system therefore faces a decision before the order is placed: should the planner see the recommendation now?

An order placed to address a local shortage can affect the rest of the chain long before the resulting shipment arrives. Orders transmit information upstream, while shipments and inventories respond with delays. The bullwhip literature explains how local decisions can amplify orders and impose costs elsewhere in the chain (Lee et al. 1997, Chen et al. 2000). Supply disruption adds a physical constraint: even an adjustment that improves one position's inventory balance can compete for capacity needed by another position or at a later time. Investments in inventory, reserve capacity, and recovery information address this coordination problem (Kleindorfer and Saad 2005, Tomlin 2006, Mehrotra and Schmidt 2021). Advice enters the chain through those same ordering decisions, so its value depends on consequences beyond the planner receiving it.

Planners need not implement the orders an advisor recommends. They may follow advice, retain their intended order, or adjust by more than the suggested amount. Research on algorithmic decision support documents substantial variation in advice use across managers, tasks, and interfaces (Lehmann et al. 2022, Caro and Sáez de Tejada Cuenca 2023, Balakrishnan et al. 2026). A separate question is whether seeing a recommendation changes how the planner responds to the next one. Our baseline analysis considers the case in which each display reduces that later response. Showing a modestly useful adjustment then uses some of the influence that could be reserved for a more consequential decision. Sending another recommendation may be inexpensive, yet its effect on the planner can carry an opportunity cost throughout recovery.

We study a delivery rule that sees the planner's intended order and the advisor's recommendation, then decides whether to show that recommendation. We call this choice the delivery *gate*. The planner retains ordering authority, and the advisor's recommendation rule remains fixed. The gate seeks to reduce costs across the chain, accounting for consequences beyond the recipient's inventory position and beyond the current decision. A display can change future responsiveness even when advice is ignored; following the advice also changes the orders that the chain must fulfill. The same recommendation can therefore be worth displaying at one point in a disruption and worth withholding at another.

When each display reduces later response but leaves the operating benefits of subsequent recommendations unchanged, advice competes for the planner's remaining responsiveness. The optimal rule displays when the expected benefit exceeds the value of preserving that response and withholds when it falls below. This comparison gives a minimum benefit for display, which we call the reservation value. More remaining opportunities raise this value because waiting leaves more chances to influence a valuable adjustment. When the sequence can end at an uncertain date, a higher chance of another opportunity has the same effect. By contrast, multiplying every current and future benefit by the same factor leaves the delivery rule unchanged. Recovery expectations and the scale of operating costs thus affect advice through distinct channels.

The information needed to choose advice changes over time. Suppose every recommendation would benefit the chain if followed. At the final opportunity, both a small and a large benefit justify display, so distinguishing them has no value. Earlier, that distinction can help the advisor preserve responsiveness for the larger benefit. For two possible benefit levels, we identify how many opportunities must remain before this information becomes valuable. If the chance of another opportunity is sufficiently low, the distinction is worthless at every horizon; above that cutoff, it matters once enough opportunities remain. Recovery expectations therefore affect both which adjustment deserves advice and whether additional information can change that choice. An interface adequate for the last decision can be inadequate earlier in the disruption.

An accepted recommendation can also use capacity that would otherwise support a later adjustment. When two adjustments draw on shared capacity, display reduces responsiveness and adoption changes what the chain can do next. A planner who is more likely to follow advice can then optimally receive less of it, because the current adjustment is more likely to displace a valuable later use of capacity. When the later adjustment's benefit levels off with capacity, we identify conditions under which withholding is optimal only over an intermediate range: when little is available, there is little future benefit to protect; when capacity is abundant, both adjustments can be accommodated. We derive the range and give conditions for it to remain a single interval. Scarce capacity can therefore make withholding optimal for the planners most willing to act.

A randomized behavioral experiment with full-time managers enrolled in a part-time MBA program records intended orders before any recommendation appears and final orders afterward. Participants revise 34–43% of orders on decisions with a display, compared with 4–5% without one. Recording both orders distinguishes the decision to revise from the size of the adjustment toward advice. Randomly timed advice is also associated with higher peak orders in a rank-based comparison, with sensitivity to the statistical specification. Observations from consenting participants allow separate estimates of how managers set orders and how they respond to advice, which enter the computational supply-chain model.

The computational evaluation compares learned delivery rules and simple rules based on disagreement between the intended and recommended orders. Benchmarks include displaying every recommendation (blanket advice), displaying none (silence), and random timing at the same frequency as a learned rule. One learned rule uses the recipient's state, response history, time, and public disruption information; we call it the local gate. It displays 33.18 of 80 recommendations, or 41% of the blanket-advice total, and has a mean-cost point estimate 3.0% below blanket advice's. Paired mean and median cost differences favor it over random timing at the same frequency in each of five evaluation samples, although the magnitudes vary substantially.

A second learned rule observes information from the whole chain, including other positions' operating states. This chain-wide gate displays more often and has a higher mean cost than random timing at its own display frequency. A chain-wide policy can reproduce a local policy by ignoring its additional inputs. The

observed ranking therefore concerns the policies learned under each information set, rather than the best performance that set permits.

The paper makes three contributions.

(1) We make the timing of a fixed recommendation a decision in disruption management. Adherence-aware optimization changes what to recommend (Grand-Clément and Pauphilet 2026), and budgeted disclosure allocates a fixed allowance of support (Giacchetta et al. 2026). Here, the opportunity cost of display arises within the sequence of operating decisions: seeing advice changes future responsiveness, and adopting it can consume capacity needed later.

(2) We extend sequential allocation logic (Derman et al. 1972) to characterize when information changes advice timing and when responsiveness reverses its effect on delivery. Information can be valuable even if it never identifies a harmful recommendation; its value depends on the remaining horizon and disruption persistence. Adoption-contingent capacity use can make withholding optimal for a more responsive planner while display remains optimal for a less responsive one. We give explicit conditions on future benefits for a bounded withholding region.

(3) We distinguish the planner's intended order from the adjustment following advice, then compare targeted and random placement at matched display counts in a calibrated supply chain. The experiment measures how recommendations enter ordering decisions; the computational comparison measures the consequences of alternative delivery policies under fitted behavior. At the same display count, paired mean and median cost differences favor the learned local gate over random timing in every participant fold. Comparing targeted delivery with random timing at the same count gives firms a benchmark for selectivity beyond reducing the number of displays.

2 Related Literature

2.1 Human responses to algorithmic advice

The value of advice depends on the information held by its recipient and on how that recipient uses the recommendation. Ibrahim et al. (2021) examine how eliciting human judgment changes its value to a prediction algorithm. Boyaci et al. (2024) analyze machine input under cognitive limitations, and Balakrishnan et al. (2026) study advice weighting when the human has private information. Our delivery decision precedes that response: the system chooses whether to show a fixed recommendation while leaving the planner in control of the order.

Adherence determines how a recommendation translates into an operating action. Lehmann et al. (2022) examine the effects of complexity and transparency on advice use, and Caro and Sáez de Tejada Cuenca (2023) document managers' deviations from markdown recommendations. Prescriptive systems can respond by changing advice content (Grand-Clément and Pauphilet 2026) or by designing guidance and

automation across sequential decisions (Bastani et al. 2026, Imdahl et al. 2026). Liu and Sinchaisri (2026) study how an advisor designs its recommendation in anticipation of stochastic consultation with personal AI, distinguishing whether that advice is obtained from the weight it receives. Cunningham et al. (2026) examine human–AI collaboration in supply chains. Our gate allows delivery to vary while the recommendation rule and the planner's decision authority remain fixed.

An interface can affect both the current decision and the resources available for later decisions. Boyaci et al. (2026) show how explanations change decision accuracy and cognitive effort. Experimental evidence on AI abstention finds heterogeneous effects on acceptance across perceived-expertise groups (Wolczynski et al. 2026). In learning to defer, Zhang et al. (2026) allocate decisions between a human and an AI when workload affects human performance. Our model allocates opportunities to advise a person who makes every operating decision. Seeing advice can change modeled responsiveness even when the planner ignores it; capacity changes only when the planner acts. Advice timing consequently depends on two different events, exposure and adoption.

2.2 Information in disrupted supply chains

Supply-chain information has value through the decisions it changes. Cachon and Fisher (2000) evaluate shared demand and inventory information, while Gaur et al. (2005) study sharing and inference under serially correlated demand. With capacity constraints, Kapuściński and Parker (2022) identify conditions under which local information reproduces specified echelon policies. Our information comparison concerns the decision to display advice. Even with a fixed distribution of recommendation benefits, the information needed for this decision changes with the number of opportunities remaining to influence the planner.

Behavior can amplify the consequences of decentralized information. The bullwhip effect links local ordering decisions to upstream variability (Lee et al. 1997, Chen et al. 2000), and misperceptions of feedback can amplify orders even in simple inventory systems (Sterman 1989). Experiments examine point-of-sale information, inventory visibility, and coordination risk (Croson and Donohue 2003, 2006, Croson et al. 2014); trust and communication further shape information use (Inderfurth et al. 2013). The behavioral operations literature places these decision processes within operating models (Gino and Pisano 2008, Bendoly et al. 2010). During a shortage, protective ordering at one position can intensify pressure elsewhere (Sterman and Dogan 2015). Recording both intended orders and responses to advice makes these two sources of operating behavior distinguishable in our design.

Disruption management studies how inventory, mitigation, contingency actions, and reserve capacity support supply-chain resilience (Kleindorfer and Saad 2005, Tomlin 2006, Lücker et al. 2021). Recovery information and control-system design influence the timing of those actions (Mehrotra and Schmidt 2021, Schmidt and Raman 2022). Our model adds the timing of advice to these recovery decisions. When displays reduce later response, withholding preserves the ability to influence a subsequent adjustment. This is

analogous to reserving capacity for a later use, but the two resources are depleted by different events: display changes responsiveness, whereas adoption changes capacity. Autonomous inventory controllers choose orders directly (Oroojlooyjadid et al. 2022, Long et al. 2026). Our setting retains the intervening human response, which makes delivery and adoption distinct events.

2.3 Sequential allocation and disclosure

Sequential stochastic allocation supplies the reservation-value logic for choosing among uncertain opportunities (Derman et al. 1972). In our independent-opportunity model, the responsiveness available at successive displays is $r, \alpha r, \alpha^2 r, \ldots$; withholding leaves the next level available. Models of fatigue and rest (Baucells and Zhao 2019) and the changeover costs of collaboration (Gurvich et al. 2020) likewise recognize that using a productive resource can alter its future contribution. With fixed benefit opportunities, this allocation logic yields a reservation value for display. Allowing adoption to change the next opportunity introduces the separate cost of displaced physical capacity.

Dynamic disclosure offers the closest information-design comparison. Papanastasiou et al. (2018) use selective disclosure to improve exploration, and Giacchetta et al. (2026) derive selective-support thresholds under a disclosure budget. The opportunity cost in our setting is endogenous: displaying advice changes the response available later, and adoption can reduce the capacity available to a subsequent adjustment. These effects link the value of information to the timing of operating decisions, even when every recommendation is beneficial. The response law is a model primitive; the planner does not strategically infer information from silence. The resulting problem is to allocate opportunities to influence orders, with both human response and physical capacity carried forward.

3 A Model of Advice Delivery

A delivery policy chooses whether a planner sees an available recommendation; the planner retains authority over the order. Showing advice can affect the chain through two channels. It can change the planner's willingness to act on later recommendations, and an adopted adjustment can use capacity needed later. Holding future operating opportunities fixed isolates the first channel and yields a reservation value for advice. Allowing adjustments to compete for shared capacity brings the second channel into the decision. The objective throughout is system cost. Proofs and analytical extensions are in Section EC.1.

3.1 Feasible orders and system cost

A planner chooses whether and how far to adjust an order after seeing advice. Let a^H be the intended order and a^A the recommendation available to a gate. The gate chooses $g \in \{0,1\}$, where one displays the recommendation. A response weight W determines the planner's adjustment, and the feasible execution rule is

$$u(g) = a^H + gW(a^A - a^H), \qquad a^E(g) = \Pi_{\mathcal{A}}(u(g)). \tag{1}$$

Here $\mathcal{A}$ is the feasible order set. For nonnegative continuous orders, the map is $\Pi_{\mathcal{A}}(u) = \max\{0, u\}$. An upper bound or integer constraint requires the corresponding projection. A weight outside $[0, 1]$ can produce an infeasible adjustment even when the intended and recommended orders are nonnegative.

Our operating context is a serial retailer–wholesaler–distributor–factory chain. Orders and shipments are delayed, so an adjustment changes outstanding orders before its full inventory consequences become visible. A delivery policy must therefore value the adjustment across the chain, including the pressure it places on constrained upstream capacity. A reduction in the recipient's inventory cost alone need not justify showing the recommendation.

An upward recommendation can improve the recipient's position while increasing system cost. To see the distinction, consider a replenishment horizon with local target ℓ, shortage cost $b > 0$, overage cost $h \geq 0$, available capacity K, and incremental upstream burden $\kappa > 0$. The continuation-cost approximation is

$$C^L(q) = b(\ell - q)^+ + h(q - \ell)^+, \qquad C^S(q; K) = C^L(q) + \kappa(q - K)^+. \tag{2}$$

For feasible orders $0 \leq q_0 < q_1 \leq \ell$, let $\delta = q_1 - q_0$. Moving from the intended order q_0 to q_1 saves $b\delta$ locally. The system benefit also deducts the incremental burden beyond capacity, giving $D^S(K) = b\delta - \kappa[(q_1 - K)^+ - (q_0 - K)^+]$. When $\kappa > b$, this benefit changes sign at $K^\star = q_1 - b\delta/\kappa$. Thus the retailer in the introduction can receive a locally helpful recommendation that should be withheld because upstream capacity is too scarce. Proposition EC.2 gives the full piecewise comparison.

3.2 The display decision

Even a recommendation that reduces current system cost can affect the value of later advice. Index opportunities by k, and let the state S_k contain physical variables, responsiveness, intended and recommended orders, and any history needed to specify transitions. Inventories, backlogs, and pipelines determine the physical consequences of an adjustment; response state determines how the planner acts after a display. A local interface observes a restricted history or a function of this state. The objective remains total system cost under either observation rule.

For continuation policy π, define the cost of current action g and its expected continuation as

$$Q_k^\pi(s, g) = \mathbb{E}[c_k^g + V_{k+1}^\pi(S_{k+1}^g) \mid S_k = s], \quad \Delta_k^\pi(s) = Q_k^\pi(s, 0) - Q_k^\pi(s, 1). \tag{3}$$

The cost allocation c_k counts each physical cost once. With optimal continuation, a fully informed gate displays when $\Delta_k^\star(s) > 0$. A partially informed gate must make the corresponding comparison using its observation history. Advice can change the continuation value through both the physical state and the response state. Even an ignored recommendation may matter if the act of displaying it changes how the planner responds next time. Proposition EC.1 separates these channels under an explicit coupling.

With one opportunity left, no later recommendation competes for the planner's response, and the display boundary is zero expected benefit (Proposition EC.5). Earlier in the sequence, the gate must also value the response it gives up by displaying now. To separate that cost from changes in inventories and capacity, suppose initially that today's action leaves the distribution of future operating benefits unchanged.

3.3 Reservation values

Consider a planner with n remaining opportunities to receive advice and current responsiveness $r \in [0,1]$. At each opportunity, adopting a feasible recommendation produces benefit Y relative to retaining the intended order. The random variable Y is integrable and may be negative. The gate observes information $\mathcal{H}$, giving conditional expected benefit $M_{\mathcal{H}} = \mathbb{E}[Y \mid \mathcal{H}]$. The decision concerns the recommendation already available at that opportunity; the gate cannot replace it with another order.

If displayed, the recommendation is adopted with probability r, conditionally independently of its benefit given the signal. Its expected current benefit is therefore $rM_{\mathcal{H}}$. Display reduces the responsiveness available next time to αr, where $0 \le \alpha < 1$, whether the current recommendation is adopted or ignored. Withholding preserves r. Adoption switches between two feasible orders. The expected benefit is therefore linear in the adoption probability; costs need not be linear in an interpolated order. No display budget or direct display charge is imposed.

Successive benefit–signal pairs are independent and identically distributed, and their distribution is unchanged by the gate's actions. This assumption isolates competition for responsiveness: using one recommendation changes the planner's response to the next, while leaving the next recommendation's operating benefit unchanged. Section 3.6 relaxes that restriction by allowing adoption to consume shared capacity. Let $F_n^{\mathcal{H}}(r)$ be the maximum expected cumulative benefit before the next signal is observed, and write $d = 1-\alpha$.

PROPOSITION 1 **(Endogenous reservation value)**. *Under the independent-opportunity model, $F_n^{\mathcal{H}}(r) = rv_n^{\mathcal{H}}$, where*

$$v_0^{\mathcal{H}} = 0, \qquad v_n^{\mathcal{H}} = v_{n-1}^{\mathcal{H}} + \mathbb{E}[(M_{\mathcal{H}} - dv_{n-1}^{\mathcal{H}})^+]. \tag{4}$$

For $r > 0$, an optimal gate displays when $M_{\mathcal{H}} > \theta_n^{\mathcal{H}}$ and withholds when $M_{\mathcal{H}} < \theta_n^{\mathcal{H}}$, where $\theta_n^{\mathcal{H}} = dv_{n-1}^{\mathcal{H}}$. Either action is optimal at equality. The reservation values are nonnegative and nondecreasing in n. At $r = 0$, every policy has zero value.

Withholding gives continuation value $rv_{n-1}^{\mathcal{H}}$. Display gives current expected benefit $rM_{\mathcal{H}}$ and leaves continuation value $\alpha rv_{n-1}^{\mathcal{H}}$. The difference is the current benefit less the value of responsiveness consumed. Dividing by $r > 0$ yields the reservation value in Proposition 1. Current responsiveness cancels because it scales both current adoption and every future adoption by the same factor.

The remaining horizon changes that balance. With one opportunity left, the continuation value is zero and every recommendation with positive expected benefit is shown. Earlier, preserving response gives access

to a larger set of future opportunities, raising the benefit required for display. If $\alpha = 1$, the behavioral opportunity cost disappears: the value becomes $nr\mathbb{E}[M_{\mathcal{H}}^+]$ and the threshold is zero at every horizon. This boundary identifies precisely the role played by response depreciation in the benchmark.

COROLLARY 1 **(A horizon-dependent capacity threshold)**. *For $r > 0$, suppose the current benefit is $D^S(K)$ from Proposition EC.2, current capacity is observed, $\kappa > b$, and the independent continuation has reservation value θ_n. If $0 \le \theta_n < b\delta$, display is strictly optimal for*

$$K > K_n = q_1 - \frac{b\delta - \theta_n}{\kappa} = K^\star + \frac{\theta_n}{\kappa}, \tag{5}$$

and withholding is strictly optimal for $K < K_n$. At equality either action is optimal. If $\theta_n > b\delta$, withhold for all K; if $\theta_n = b\delta$, display is only weakly optimal for $K \ge q_1$.

Capacity between $K^\star$ and K_n supports an adjustment that reduces current system cost, but that reduction is too small to justify its effect on later advice. As the remaining horizon grows, the gate requires a larger capacity margin before recommending the upward adjustment. For the retailer, this is the difference between having enough upstream capacity for an order to help and having enough to make that order worth prioritizing over a later recommendation. The comparison holds the distribution of those later opportunities fixed.

Implementing the reservation rule requires estimates of both current benefit and continuation value. Errors in those estimates affect performance when they change the display decision. Corollary EC.1 shows that a uniformly accurate estimate can misclassify advice only near the reservation boundary. With a bounded density of benefits near that boundary, the resulting loss is quadratic in the combined estimation error. Thus accuracy far from the delivery threshold can be less consequential than accuracy for the recommendations whose timing is in doubt.

3.4 Value of information

The reservation value also determines which information is useful. A signal can distinguish two beneficial recommendations without changing the final display decision, yet affect which one deserves a display when opportunities remain. Consider two per-opportunity information sets $\mathcal{G} \subseteq \mathcal{F}$ under the same benefit distribution and depreciation rate. Both gates know the distribution and optimize their continuation policies. Define

$$\begin{aligned} T_{\mathcal{H}}(v) &= v + \mathbb{E}[(M_{\mathcal{H}} - dv)^+], \\ J(v) &= \mathbb{E}[(M_{\mathcal{F}} - dv)^+] - \mathbb{E}[(M_{\mathcal{G}} - dv)^+]. \end{aligned} \tag{6}$$

Here $J(v)$ holds the continuation value fixed. It measures the benefit of resolving uncertainty about whether current advice exceeds the reservation value dv. The full information comparison must also allow finer information to change the continuation value.

PROPOSITION 2 **(Dynamic value of information)**. *Let* $\Delta_n = v_n^{\mathcal{F}} - v_n^{\mathcal{G}}$. *Then* $\Delta_n \geq 0$, *and there exist* $a_n \in [\alpha, 1]$ *such that*

$$\Delta_n = a_n \Delta_{n-1} + J(v_{n-1}^{\mathcal{G}}), \qquad \Delta_0 = 0. \tag{7}$$

Consequently,

$$\sum_{j=0}^{n-1} \alpha^{n-1-j} J(v_j^{\mathcal{G}}) \;\leq\; \Delta_n \;\leq\; \sum_{j=0}^{n-1} J(v_j^{\mathcal{G}}). \tag{8}$$

With $0^0 = 1$, *these bounds also apply when* $\alpha = 0$. *Moreover,*

$$J(v) = \mathbb{E}[\min\{\mathbb{E}[(M_{\mathcal{F}} - dv)^+ \mid \mathcal{G}], \mathbb{E}[(dv - M_{\mathcal{F}})^+ \mid \mathcal{G}]\}]. \tag{9}$$

Thus $J(v) > 0$ *precisely when a set of coarse-signal realizations of positive probability leaves strictly positive conditional expected mass on both sides of* dv.

Finer information contributes through the opportunities at which it changes the display decision. In (7), $J(v_{n-1}^{\mathcal{G}})$ is that gain at the current reservation value, while $a_n \Delta_{n-1}$ carries forward the gain from better informed continuation decisions. The factor a_n allows for the informed gate's changed willingness to wait. For $\alpha > 0$, the two information sets give equal values through horizon n exactly when all n terms $J(v_0^{\mathcal{G}}), \ldots, J(v_{n-1}^{\mathcal{G}})$ are zero. Agreement about which recommendations are beneficial establishes only the first of these conditions; earlier decisions can require benefits to exceed a positive reservation value.

With two positive benefit levels, the horizon at which information becomes valuable can be characterized exactly. Let Y take two positive values $y_L < y_H$, with probabilities p and $1-p$, where $0 < p < 1$. The fine signal reveals Y, the coarse signal reveals nothing, and $\mu = p y_L + (1-p) y_H$.

COROLLARY 2 **(The horizon at which information becomes valuable)**. *For* $r > 0$ *and* $0 \leq \alpha < 1$, *always displaying is optimal for the coarse gate, whose value is* $v_n^{\mathcal{G}} = \mu(1-\alpha^n)/(1-\alpha)$. *Fine information is strictly valuable at horizon n exactly when*

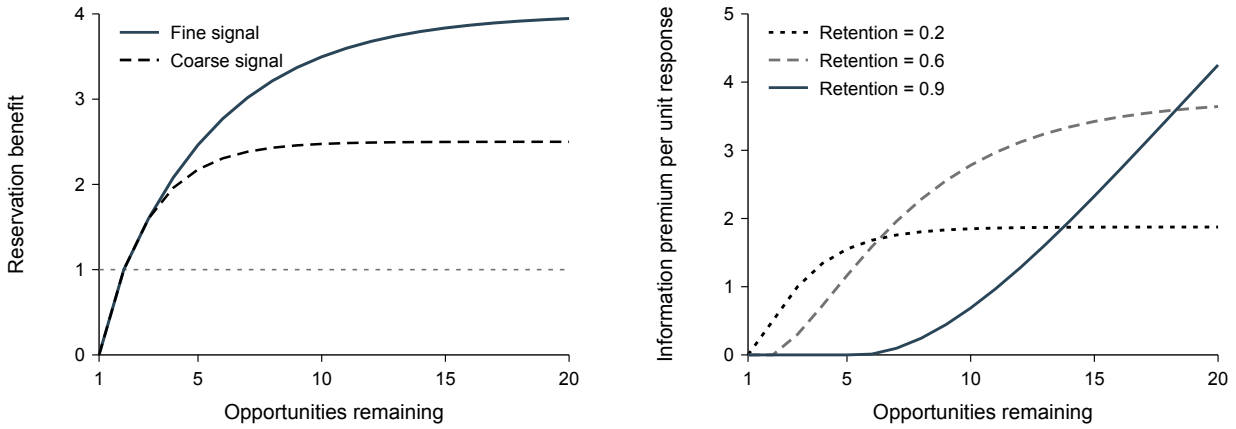
$$\mu(1 - \alpha^{n-1}) > y_L. \tag{10}$$

Otherwise the fine and coarse optimal values coincide. At every fixed finite horizon, the information premium is zero for retention rates sufficiently close to one. For each fixed $\alpha < 1$, *it is positive for all sufficiently long horizons.*

Two quantities determine when the signal becomes useful: the weaker recommendation's benefit relative to the expected benefit of a future one, and the responsiveness retained after each display. A short horizon or sufficiently persistent response makes it optimal to use both types of advice, so distinguishing them adds no value. A longer horizon creates an opportunity to wait for the stronger type. For example, suppose Y is equally likely to be 1 or 4 and $\alpha = 0.2$. At the last opportunity, both gates show every recommendation

and earn $2.5r$. With two opportunities remaining, both face reservation value $0.8(2.5) = 2$. The fine gate withholds when $Y = 1$ and displays when $Y = 4$, obtaining $3.5r$. The coarse gate displays on the conditional mean 2.5 and obtains $3r$. Information is worth $0.5r$ even though every recommendation has a positive operational benefit. No harmful recommendation, estimation error, or display budget is needed for this difference.

Figure 1 extends this example across horizons and retention rates. Finer information first becomes valuable when the weaker recommendation falls below the reservation value. With less response depreciation, that crossing occurs later. Over a long horizon, value is governed by the upper end of the benefit distribution that the signal can distinguish (Corollary EC.2).



(a) Reservation values (b) Value of the finer signal

Figure 1 Information value over the remaining horizon. Benefits are 1 and 4 with equal probabilities. The fine signal reveals the benefit; the coarse signal reveals only its distribution. Panel (a) uses retention $\alpha = 0.6$; the horizontal reference is the smaller benefit. Panel (b) gives $v_n^{\mathcal{F}} - v_n^{\mathcal{G}}$ for three retention rates. Plotted values are per unit of initial responsiveness; multiplying by r gives the corresponding expected-benefit quantities. Curves are exact evaluations of the analytical recursion.

3.5 Disruption duration

The reservation value so far depends on a known number of remaining opportunities. During an outage, that number may depend on when supply recovers. Suppose that after each opportunity the sequence continues with probability $q \in [0, 1]$, independently of its benefits and the gate's actions. There are at most n opportunities, and resolution leaves zero incremental value from this advisory task. The benefit and response assumptions of Section 3.3 otherwise remain unchanged. Holding the distribution of each opportunity fixed separates the effect of a longer disruption from a change in the value of individual adjustments.

PROPOSITION 3 **(Persistence and the reservation value)**. *Expected optimal benefit is $rv_n(q)$, with*

$$v_0(q) = 0, \qquad v_n(q) = qv_{n-1}(q) + \mathbb{E}[(M - (1-\alpha)qv_{n-1}(q))^+]. \tag{11}$$

The reservation value $\theta_n(q) = (1-\alpha)qv_{n-1}(q)$ *is nondecreasing in* q *and* n*. If a severity parameter* $s > 0$ *multiplies every current and future benefit by* s*, leaving signals, persistence, and responsiveness unchanged, then* $v_n(q;s) = sv_n(q;1)$ *and the optimal display decisions are unchanged.*

A longer expected sequence raises the value of waiting for a stronger recommendation. Increasing all benefits proportionally instead raises today's gain and the opportunity cost of using responsiveness by the same factor. This distinction matters for disruption planning: a recovery forecast changes the number of adjustments that may remain, whereas a common change in the cost scale need not change their relative priority.

REMARK 1 (A STATIONARY RESERVATION VALUE). With no deterministic horizon and $0 \le q < 1$, the reservation value is the unique nonnegative solution of

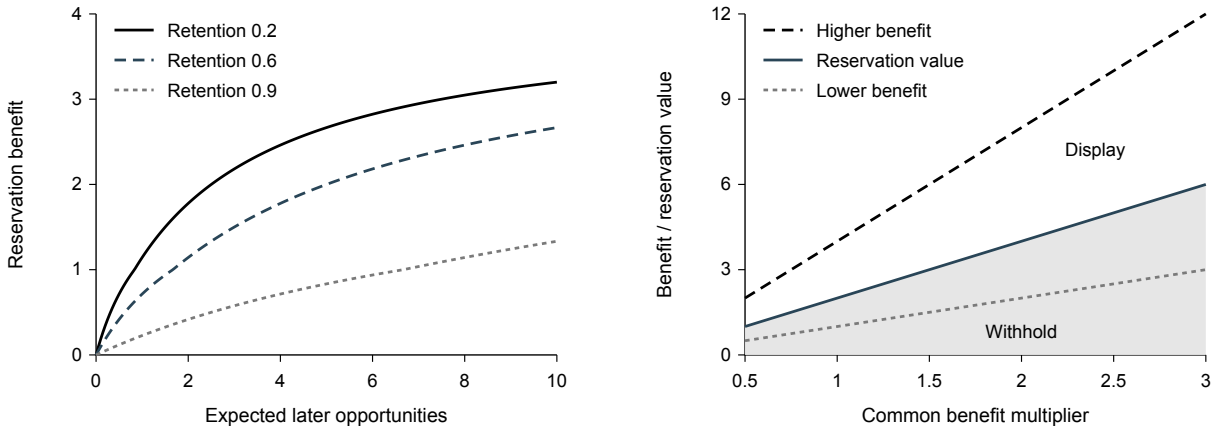
$$\theta = (1-\alpha)\frac{q}{1-q}\,\mathbb{E}[(M-\theta)^+]. \tag{12}$$

Here $q/(1-q)$ is the expected number of opportunities after the current one. Finite-horizon reservation values converge to θ.

The factor $q/(1-q)$ gives recovery information a direct role in the delivery decision: more expected later opportunities increase the value of preserving responsiveness. This complements the role of recovery information in operating plans (Mehrotra and Schmidt 2021). The interpretation depends on when useful adjustments end. If recovery itself creates further valuable adjustments, their value belongs in the continuation. A deeper capacity cut can also change the composition of benefits, so its effect need not coincide with the common scaling in Proposition 3.

Persistence also changes the value of knowing which recommendation has arrived. In the two-benefit model, a sufficiently short expected sequence never makes waiting worthwhile, even for the weaker recommendation. Finer information then has no value at any horizon. Above a persistence cutoff, enough remaining opportunities make that distinction useful (Corollary EC.3).

Figure 2 separates a longer adjustment horizon from a common increase in operating stakes. With benefits of 1 and 4 equally likely, retention $\alpha = 0.6$, and five expected later opportunities, the stationary reservation value is 2. The gate withholds a benefit of 1 and displays a benefit of 4. Doubling both benefits doubles the threshold to 4 and preserves both decisions. Extending the opportunity sequence instead changes which benefits justify display.



(a) Persistence and the reservation value

(b) Proportional changes in benefits

Figure 2 Advice timing under persistence and common benefit scaling. Benefits are 1 and 4 with equal probabilities. Panel (a) evaluates (12) at $L = q/(1-q)$ expected later opportunities for three retention rates. Panel (b) fixes $L = 5$ and $\alpha = 0.6$ and multiplies both benefits by s; the reservation value is $2s$, while the two benefit levels are s and $4s$. Shading marks benefits below the reservation value. The lower recommendation is withheld and the higher recommendation is displayed at every multiplier shown. All values are exact analytical quantities.

3.6 Competition for capacity

The independent-opportunity model preserves the distribution of later benefits after an order adjustment. In a disrupted chain, fulfillment of an adopted adjustment can instead use capacity needed for a later one. Suppose two recommendations draw on a common capacity allocation over a replenishment horizon. Display changes responsiveness whether or not the planner acts; fulfillment uses capacity only when the adjustment is adopted. The distinction creates a second opportunity cost of showing advice.

Let $K \geq \delta > 0$ be capacity available across the two opportunities. The first recommendation, if adopted, uses δ units and creates current system benefit $A > 0$; otherwise it uses no incremental capacity and creates no incremental benefit. Adoption occurs with probability $r \in (0, 1]$. Withholding preserves both K and r. A display reduces next-period responsiveness to αr, where $0 \leq \alpha \leq 1$, even if the first recommendation is ignored. At the final opportunity, adopting the feasible recommendation creates benefit $B(k) \geq 0$ at remaining capacity k. The function B is nondecreasing. Final adoption is conditionally Bernoulli with the prevailing responsiveness, independently of earlier adoption given the resulting state. There is no capacity replenishment between opportunities and no salvage value after the second. Thus both recommendations are beneficial when evaluated in isolation, and all capacity use is feasible. Benefits are measured relative to the same no-adjustment baseline and are additive across opportunities.

PROPOSITION 4 **(Response and physical opportunity costs)**. *In this two-opportunity model, the expected benefit of first displaying rather than withholding, followed by optimal final delivery, is*

$$\Delta(K, r) = r\{A - (1-\alpha)B(K) - \alpha r[B(K) - B(K-\delta)]\}. \tag{13}$$

For $r > 0$, display is optimal exactly when the term in braces is nonnegative. The reservation benefit is weakly increasing in r, and strictly increasing if $\alpha > 0$ and $B(K) > B(K-\delta)$. Hence a recommendation can be displayed for a less responsive planner and withheld for a more responsive planner, holding capacity and recommendation benefits fixed.

The first opportunity cost, $(1-\alpha)B(K)$, comes from reduced responsiveness. The second, $\alpha r[B(K) - B(K-\delta)]$, comes from capacity consumed when the first recommendation is adopted. Greater responsiveness increases both the chance of earning the current benefit and the chance of changing the physical state. Dividing by r removes the former effect but not the latter. Current responsiveness therefore affects optimal delivery here, whereas it cancels from the display comparison in Proposition 1. If $\alpha = 1$, only physical displacement remains; if $\alpha = 0$, a display eliminates the final response and its additional physical displacement has no value.

More available capacity need not make display more attractive. In Section 3.1, capacity determines the benefit of the current adjustment. Here it also determines what that adjustment can displace. Suppose the final recommendation serves up to $L > \delta$ units at benefit $b_2 > 0$ per unit, so $B(k) = b_2 \min\{k, L\}$. Write $d = 1-\alpha$ and consider $0 < \alpha < 1$. With scarce capacity, the final opportunity has little value to protect. Intermediate capacity makes the two recommendations compete, whereas abundant capacity removes that competition. The reservation benefit consequently rises and then falls.

COROLLARY 3 **(An intermediate-capacity withholding region)**. *Suppose*

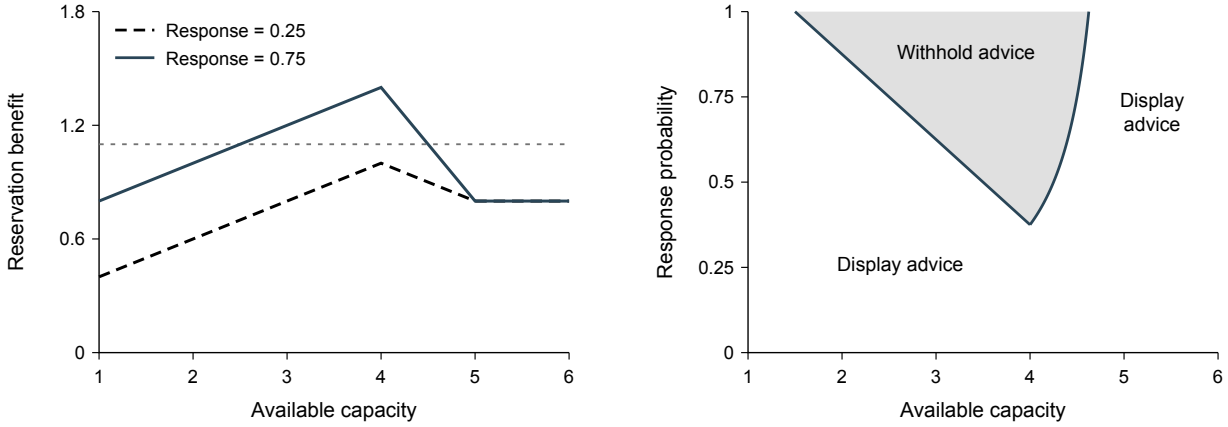
$$\max\{b_2\delta(d+\alpha r),\, db_2L\} < A < db_2L + \alpha r b_2 \delta. \tag{14}$$

Define

$$K_- = \frac{A - \alpha r b_2 \delta}{db_2}, \qquad K_+ = L + \delta - \frac{A - db_2L}{\alpha r b_2}. \tag{15}$$

Then $\delta < K_- < L < K_+ < L+\delta$. Display is strictly optimal for $\delta \le K < K_-$ and for $K > K_+$; withholding is strictly optimal for $K_- < K < K_+$. Either action is optimal at the two boundaries. Within the parameter region in (14), raising r lowers K_- and raises K_+.

For example, take $\delta = 1$, $L = 4$, $b_2 = 1$, $\alpha = 0.8$, $r = 0.75$, and $A = 1.1$. The gate displays below capacity 2.5 and above 4.5, but withholds between them. At $K = 4$, the reservation benefit is $0.8 + 0.8r$: the first recommendation is displayed when $r = 0.25$ and withheld when $r = 0.75$. The greater response probability makes the first adjustment more likely to displace the second. Figure 3 displays this comparison and the associated delivery regions.



(a) Reservation benefit (b) Optimal delivery decision

Figure 3 Shared capacity and responsiveness. The final benefit is $B(k) = \min\{k, 4\}$, first-adoption capacity use is $\delta = 1$, retention is $\alpha = 0.8$, and current benefit is $A = 1.1$. Panel (a) compares reservation benefits at two response probabilities; the horizontal reference is A. Panel (b) shows the display decision over capacity and responsiveness. For positive responsiveness, withholding is strictly optimal inside the shaded region and display is strictly optimal outside it. Either action is optimal on the boundary curves and at zero responsiveness. Values are analytical illustrations.

The withholding interval in the capped-benefit example extends to a broader class of continuation values. Let

$$R_r(K) = (1-\alpha)B(K) + \alpha r[B(K) - B(K-\delta)]$$

be the reservation benefit in Proposition 4. Its first term increases with capacity because there is more future benefit to protect. Under concavity, its second term decreases because losing δ units becomes less costly at higher capacity. The shape of their sum depends on the relative rate at which marginal benefit declines.

PROPOSITION 5 **(The shape of the withholding region)**. *Let $0 < \alpha < 1$, $r > 0$, and let B be nonnegative, increasing, concave, and continuously differentiable, with $B'(k) > 0$. For $K \geq \delta$, define*

$$\rho(K) = \frac{B'(K-\delta)}{B'(K)}, \qquad R_r'(K) = B'(K)\{(1-\alpha) + \alpha r[1-\rho(K)]\}. \tag{16}$$

If ρ is nondecreasing, the strict withholding set $\{K \geq \delta : R_r(K) > A\}$ is an interval, possibly empty or unbounded. If $B(k) \to B_\infty < \infty$ and

$$\max\{R_r(\delta), (1-\alpha)B_\infty\} < A < \sup_{K \geq \delta} R_r(K), \tag{17}$$

this interval is nonempty and bounded, with display optimal on either side. If $\rho(K) \leq 1 + (1-\alpha)/(\alpha r)$ everywhere, R_r is nondecreasing and display is optimal below a capacity cutoff and withholding above it, with either region possibly empty. For any nondecreasing B, the withholding set expands weakly as r increases.

The ratio condition requires the marginal benefit of capacity at $K-\delta$, relative to that at K, to be nondecreasing as K rises. A sufficient smoothness condition is a nondecreasing proportional decline $-B''/B'$ in marginal benefit. Under this condition the reservation benefit can rise and then fall, but cannot fall and then rise again. The two competing opportunity costs therefore produce one connected withholding region. Concavity alone does not guarantee an interior band. For example, with $B(k)=B_\infty(1-e^{-\beta k})$, $\rho(K)=e^{\beta\delta}$ is constant and the reservation benefit is monotone. Thus concavity supports competing capacity effects, while the ratio condition limits their crossings. In either case, greater responsiveness expands the withholding set.

3.7 Advice when responsiveness changes

Multiplicative depreciation makes every display reduce later responsiveness. A planner's response may instead improve after advice or recover during silence. The same comparison between current benefit and lost continuation value applies when responsiveness moves among a finite set of states.

Let z be an observable response state, let r_z be the probability of adopting current advice, and let P and Q be the response-transition matrices after display and withholding. Current conditional expected benefit M has an integrable distribution independent of z and past actions. Conditional on z and the display decision, the marginal next-state distribution is independent of M. Physical opportunities remain independent of the gate's actions, as in Section 3.3. Write $V_n(z)$ for optimal expected benefit before observing the current M; the horizon counts the recipient's remaining opportunities to receive advice.

PROPOSITION 6 **(Delivery with response transitions)**. *With $V_0(z)=0$, define the continuation cost of display by*

$$H_n(z)=\big[(Q-P)V_{n-1}\big](z). \qquad V_n(z)=(QV_{n-1})(z)+\mathbb{E}\big[(r_zM-H_n(z))^+\big]. \tag{18}$$

An optimal gate displays when $r_zM>H_n(z)$ and withholds when the inequality is reversed. Either action is optimal at equality. If $r_z>0$, the per-adoption reservation value is $H_n(z)/r_z$. If $r_z>0$ and display improves expected continuation value relative to withholding, then $H_n(z)<0$ and display is optimal for any realized M satisfying $H_n(z)/r_z<M<0$.

The sign of $H_n(z)$ determines whether showing advice consumes or builds future response value. When silence leads to a more valuable response state, current advice must compensate for that loss. When display improves continuation value, the gate can accept a current operating loss to obtain the improvement. Recovery during silence enters through Q; changes following display enter through P. If response state is unobserved, the comparison must use a belief based on the gate's available information.

With physical feedback, the continuation comparison must also account for inventories and the capacity left by an adjustment. The gate compares expected system costs under display and withholding, conditional on information available at the decision. Under the continuous response rule in (1), that expectation integrates over the feasible orders the planner may choose. When costs or constraints are nonlinear, the benefit of full adoption multiplied by mean response generally differs from the benefit of display.

4 Behavioral Experiment and Computational Design

The experiment records the order a manager intends to place and the final order after advice can appear. Assignment to advice conditions permits comparisons of ordering outcomes, while the two order entries measure adjustments following a display. These observations also supply the ordering and response models used to compare delivery policies computationally.

4.1 Experimental design

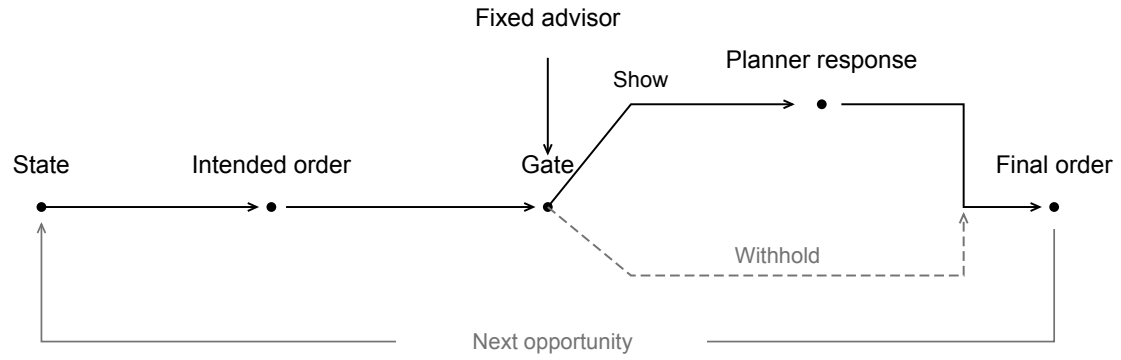
Participants were full-time managers enrolled in a part-time MBA program who completed the supply-chain task as part of an operations course. The experiment was covered by an approved umbrella institutional review board protocol for human–AI interaction research. Research consent and course grades were handled independently. Each participant managed one position in a four-echelon Beer Game for 24 weeks, while simulated policies managed the other positions. The research sample comprises 69 consenting participants among 111 eligible completers. Research analyses and behavioral calibration use only these consenting participants (sample disposition in Table EC.1).

Assignment to no advice, randomly timed advice, or a fixed selective policy was randomized within blocks defined by earlier role and team. The consenting sample contains 22, 22, and 25 participants in these conditions, respectively. The selective condition used an earlier policy with quota-triggered displays: 144 of its 169 displays followed the gate and 25 followed the quota rule. The computational evaluation uses separately trained delivery policies. Table EC.2 gives observations by condition.

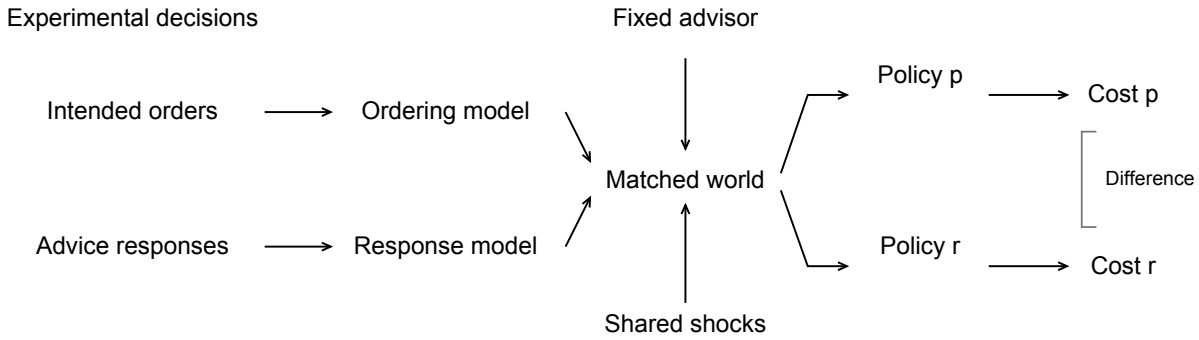
At each decision, the participant first entered an intended order. The assigned interface then displayed a recommendation or remained silent, after which the participant submitted a final order (Figure 4a). The intended order precedes that week's advice, but can reflect earlier advice and its operating consequences. The task captures repeated ordering under delayed feedback, with four predetermined disruption schedules containing 6, 10, 14, or 17 disrupted periods. Figure 4b links these observations to the fitted components and paired computational comparisons.

During each disrupted week, a required notification identified one of twelve disruption causes and stated the capacity reduction as a percentage and in units. Realized reductions ranged from 8.5 to 89.4% of nameplate capacity. Participants were not told the disruption's duration. The experimental advisor used an adaptive base-stock rule; the LLM-fitted advisor was used only in the computational evaluation.

Participants could enter orders up to 10,000 units, whereas recommendations had no upper bound. Seven displayed recommendations exceeded the entry limit, and none was adopted in full. On six eligible response occasions, the participant's intended order was already at the limit and the recommendation exceeded it, preventing any upward adjustment. The calibration retains these constrained responses (sensitivity in Section EC.2.2).



(a) A decision with optional advice



(b) From experimental observations to paired policy comparisons

Figure 4 Decision sequence and evaluation design. In panel (a), the planner states an intended order before the gate chooses whether to show the fixed advisor's recommendation. The planner submits the final order; physical and response states carry forward to the next opportunity. Panel (b) separates ordering and response models calibrated from the behavioral experiment. A fixed LLM-fitted advisor supplies recommendations to each computational policy. Comparisons use the same planner draws and exogenous shocks, with 1,500 paired worlds in five participant folds (evaluation samples with distinct held-out participants).

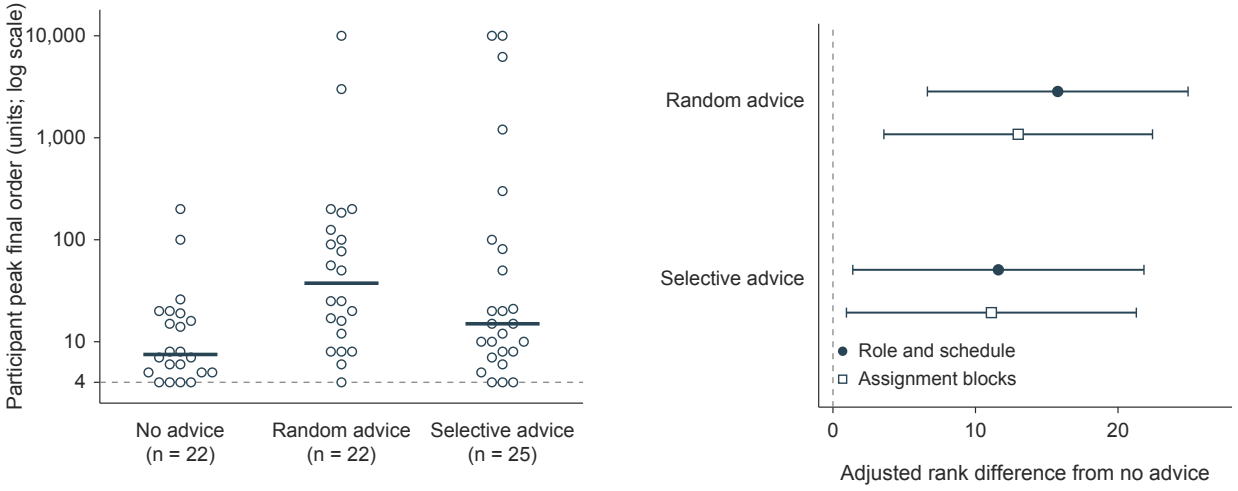
4.2 Ordering under advice

Advice can alter the size of replenishment requests as well as the frequency of order revision. Peak final orders measure each participant's largest request during weeks 5–24, after advice becomes available. The median participant peak is 7.5 units without advice, 37.5 under random advice, and 15.0 under selective advice. A rank regression adjusting for baseline ordering and role–schedule controls places the random-advice condition 15.77 ranks above no advice (pointwise 95% confidence interval $[6.63, 24.91]$; Holm-adjusted $p = 0.033$). Figure 5 shows the participant distribution and adjusted rank estimates.

The multiplicity adjustment covers 32 comparisons: four operating outcomes, two advice-versus-no-advice contrasts, and four specifications combining logarithmic or rank outcomes with role–schedule or assignment-block controls (Table EC.4). The ordering and response specifications were developed after data collection; the testing families include their reported sensitivity specifications. Ranks locate participants

within the peak-order distribution; the logarithmic specification also reflects differences in the magnitudes of their orders.

On the $\log(1 + \text{peak order})$ scale, the random-minus-no-advice estimate is 0.748 ($[0.001, 1.494]$), and the selective-minus-no-advice estimate is 0.814 ($[-0.087, 1.715]$); neither is significant after the family adjustment. The assignment-block rank estimates are positive but also fall short of the adjusted five-percent threshold. Evidence of a difference in peak-order ranks is therefore specification-dependent; the estimates establish neither a common proportional increase in orders nor an increase under both advice policies. Nor does the experiment distinguish the two advice policies on chain cost: the selective-versus-random comparison has $p = 0.67$.



(a) Participant peak orders (b) Adjusted rank comparisons

Figure 5 Peak final orders after advice becomes available. Panel (a) plots each consenting participant's largest final order in weeks 5–24 on a logarithmic scale; horizontal segments are condition medians, and the dashed line is weekly customer demand of four units. Panel (b) compares each advice condition with no advice, adjusting for pre-advice ordering. Circles use role and schedule controls ($n = 69$); squares use assignment blocks ($n = 55$). Ranks are computed within each estimation sample. Bars are pointwise 95% intervals using HC3 covariance estimates and residual-degrees-of-freedom t critical values. Of these four rank comparisons, only random advice under role and schedule controls meets the five-percent threshold after Holm adjustment across the 32-test family (Table EC.4).

The recommendations themselves sometimes call for large adjustments. Of 278 displays, 52 recommend more than 50 units against weekly customer demand of four; participants adopt 28 of these under the criterion $W \geq 0.9$ defined below. The selective condition includes 20 displayed recommendations above 1,000 units. Advice content and the states created by earlier ordering decisions can both contribute to these magnitudes. The experiment holds the advisor fixed across advice conditions and does not separately identify those channels. The order entries do, however, distinguish how far participants move from their own intended orders when a recommendation appears.

4.3 Responses to advice

Participants revise orders much more often on decisions with a display. In the random-advice condition, they revise 43.1% of orders when advice is displayed and 5.0% when it is withheld. The corresponding rates under selective advice are 34.3% and 4.2%; participants assigned no advice revise 4.7% of orders. These comparisons describe decisions with and without a display. Revision alone does not establish adoption: an order can move toward, past, or away from the recommendation.

For a displayed recommendation, the response weight $W = (a^E - a^H)/(a^A - a^H)$ measures the size and direction of that movement. Zero means retaining the intended order and one means taking the recommendation. We require $|a^A - a^H| \geq 2$ for response estimation, limiting the influence of small denominators. There are 168 eligible observations from 42 participants: 84 adoption responses ($W \geq 0.9$), 70 instances of ignoring ($W \leq 0.1$), and 14 intermediate responses. Adoption and ignoring account for most eligible responses under these thresholds. Weights outside $[0, 1]$ also occur, representing overshooting or movement away from advice.

Repeated decisions also allow responses to be related to earlier exposure and current disruption conditions. In linear probability models, the random-condition exposure coefficient for adoption is -8.6 percentage points per prior display (pointwise 95% participant-cluster interval $[-19.5, 2.2]$), whereas the selective-condition coefficient is 11.1 points ($[1.9, 20.3]$). Both control for participant effects, week, and the absolute recommendation gap. Exposure counts every earlier display. The adjusted selective-minus-random association is 23.4 points ($[-4.7, 51.5]$). The severity coefficient falls from 10.2 points per band ($[-2.3, 22.7]$) to 2.9 points ($[-16.4, 22.2]$) when participant effects are included.

None of the 17 response comparisons meets the five-percent Holm-adjusted threshold. This family includes the adoption, exact-agreement, and $W \geq 1$ definitions; alternative gap controls; the difference between exposure slopes; and the within-participant severity comparisons (Table EC.5). The estimates do not identify the proportional depreciation law in Section 3.3: they condition on eligible displays, whose occurrence and content can depend on prior advice and operating states. A computational representation of these responses must therefore be distinguished from a causal estimate of how display changes responsiveness.

4.4 Behavioral calibration

The order entries permit separate models of what a planner intends to order and how that order changes after advice. In the analytical models, adoption switches between two feasible actions and r is the probability of that switch. The observed weight W instead records the size and direction of an adjustment. The computational environment uses this continuous representation, allowing intended orders and subsequent responses to affect the chain through separate fitted components.

We fit a participant-specific ordering model and a three-state response chain. Ordering follows a five-parameter anchoring-and-adjustment rule with a shortage-augmented alternative (Sterman 1989, Sterman and Dogan 2015). The response chain describes transitions among adoption, intermediate response, and ignoring, with within-state weights winsorized to $[-1, 2]$. In computational evaluation, transitions advance whenever advice is displayed.

The fitted chain's mean response decreases from approximately 0.73 at the first modeled display to 0.44 at the twentieth. This path is an implication of the estimated transition law, not an observed twenty-display experimental trajectory. Participants received at most 16 displays and contributed at most 11 eligible observations. Moreover, 15 of the 126 consecutive eligible pairs skip an intervening display, so some estimated transitions span more than one exposure. Table EC.3 gives the transition estimates, with the eligibility gaps described in Section EC.2.1.

4.5 Supply-chain model

The fitted planners order in a serial retailer–wholesaler–distributor–factory chain with 24-week episodes. A four-week warm-up precedes 20 weeks of advice at all four positions, yielding 80 delivery opportunities. Order and shipping delays are two weeks each, and factory production takes three weeks. Demand is four units per week; each position starts with 12 units of inventory and pipeline slots of four. Weekly holding and backlog costs are 0.50 and 1.00 per unit. Factory capacity is disrupted while customer demand is held fixed, separating supply constraints from changes in final demand.

All evaluated policies execute nonnegative orders. After the response weight is applied, the adjusted order is floored at zero as in (1), before it enters the order pipeline. When a factory order reaches production, output is capped by the capacity available in that later week. This constraint matters when the fitted weight lies outside $[0, 1]$, because even nonnegative intentions and recommendations can then produce a negative blend. The learned gates were trained without the zero floor and are held fixed for evaluation. Their comparison therefore measures the performance of those fitted policies under a common feasible operating rule.

Each position independently draws a planner from the consenting participants who played that role. A world consists of these planner draws and exogenous shocks. Pairing worlds across policies holds those inputs fixed while allowing orders, inventories, backlogs, and later recommendations to follow the trajectories induced by each delivery policy. The role-specific draws preserve empirical marginal distributions; they do not reproduce dependence among planners within an actual supply chain.

The computational advisor approximates responses from Gemini 3.6 Flash (`gemini-3.6-flash`). We fit this surrogate to 19,132 queries at temperature zero under a fixed response format. Its rule anchors on received demand and applies a backlog-dependent markup with six disruption-depth regimes. This LLM-fitted rule supplies recommendations throughout the computational evaluation. It differs from the adaptive base-stock advisor used in the behavioral experiment. The surrogate permits repeated evaluation of a fixed

recommendation process; its operational scope is the elicited model and prompt. Section EC.2.6 gives the fitted function, noise specification, and discrepancies between the surrogate and the original responses.

The five evaluation samples introduced above are participant folds, each with a distinct group of held-out participants and 300 simulated episodes. Training uses the remaining participant groups, so training samples overlap across folds. Nominal factory-capacity cuts are 10, 25, 33, 50, 75, and 95%, with heterogeneous onset and duration. Disrupted periods within the advice window range from zero to 18; the 41 zero-count worlds experience their entire shock during warm-up. These paths create variation in both the constraints facing the chain and the opportunities remaining to adjust orders.

In this environment, an adjusted order changes inventories, outstanding orders, and future recommendations. The benefits of later advice therefore depend on earlier delivery decisions. The independent-opportunity model holds this channel fixed, and the shared-capacity model isolates one part of it. A computational policy's cost reflects the full trajectory produced by its displays and the fitted planners' responses.

4.6 Delivery policies

Silence withholds all advice, and blanket advice displays at all 80 opportunities. These policies bound display frequency and provide reference costs for selective delivery. The learned gates choose when to display through proximal policy optimization (PPO) (Schulman et al. 2017); they do not implement the analytical reservation rule. They share architecture, training budget, participant folds, and candidate seeds. An input mask determines which state variables each gate observes. The chain-wide gate has 64 varying inputs within a 70-slot representation. The local gate retains 26–34 inputs, depending on position, including the recipient's operating state, response history, time, and a public disruption indicator. Other-position states and chain aggregates are masked.

Two fixed rules use the disagreement between the planner and advisor. The gap filter displays according to

$$g = \mathbf{1}\{a^A - a^H \leq 4.74\}. \tag{19}$$

The hand rule adds a restriction against display at the factory. Both permit downward recommendations. Their common threshold was extracted from an earlier policy trained with an adaptive base-stock advisor in a fixed-disruption environment and is held fixed in the present evaluation.

Each learned variant uses five participant folds, three seeds, 500,000 training timesteps, discount factor $\gamma = 1$, and generalized-advantage parameter $\lambda = 0.99$. The training reward is the saving relative to a matched silence episode with the same planners, disruption path, and exogenous draws, divided by that episode's cost over the advice window. For $C_G^p(\omega) = \sum_{t=4}^{23} C_t^p(\omega)$ and positive silence cost $C_G^0(\omega)$, cumulative reward is $1 - C_G^\pi(\omega)/C_G^0(\omega)$. Maximizing this reward targets expected relative window cost. It gives a unit of saving greater weight in a world with low silence cost, so it differs from minimizing expected absolute chain cost.

We report both mean and distributional cost comparisons to make the operational consequences of that distinction visible. Training and observation details are in Section EC.2.7.

Selective delivery changes both frequency and placement. To distinguish them, two random-timing controls match the realized display count of the chain-wide and local gates, respectively, within each world. Each control redistributes that count across opportunities without the corresponding gate's targeting. Because the count comes from the gate's trajectory, these are retrospective comparisons at realized doses. Neither fixed rule has its own matched-dose control. The evaluation imposes no binding advice budget or direct display charge.

4.7 Performance measures

The operational objective is expected holding and backlog cost across all positions and weeks,

$$J(\pi) = \mathbb{E}^{\pi}\left[\sum_{t=0}^{23} C_t\right]. \tag{20}$$

For a policy p and reference r, the paired cost difference in world ω is $d_\omega = C_p(\omega) - C_r(\omega)$; a negative difference favors p. Its mean measures the average cost difference, and its median measures the central paired contrast. The latter is different from subtracting the policies' marginal median costs. Relative differences, $(C_p - C_r)/C_r$, are evaluated where reference cost is positive and answer a separate question about proportional savings.

We report variation across the five participant folds as well as the pooled comparison. For fold medians m_f, the statistic is

$$t = \frac{\bar{m}}{s_m/\sqrt{5}}, \qquad f = 1, \ldots, 5. \tag{21}$$

The corresponding mean-cost statistic uses fold means. A two-sided reference threshold of 2.776 corresponds to four degrees of freedom. Held-out participants differ across folds, but training samples overlap. The fold statistics and intervals describe variation across these overlapping fits. Intervals formed as the mean fold summary plus or minus 2.776 standard errors are labeled t_4 reference intervals; dependence across fits leaves their repeated-sampling coverage undetermined. The reference threshold is unadjusted across policy comparisons, and a value below it does not establish equivalence.

A small number of costly trajectories can have a large effect on expected cost. Alongside paired means and medians, we therefore report the 95th percentile and the mean of the 75 highest-cost worlds, the upper five percent of the evaluation sample. Tables EC.10 and EC.11 give fold medians and mean-cost comparisons. These summaries describe the fitted population and selected policies; they do not include uncertainty from re-estimating behavior or retraining delivery policies.

5 Targeted Delivery in a Disrupted Chain

Selective delivery changes how often advice appears and which order adjustments it encourages. Comparison with blanket advice measures the cost consequences of withholding some recommendations. Comparison with random timing at the same display count separates targeting from a reduction in frequency. The policies share a recommendation process and are evaluated on the same planner draws and shocks; their costs differ because delivery changes the resulting order trajectories.

5.1 Targeting versus frequency

Showing every recommendation reduces mean cost from 36,057.34 under silence to 25,141.38, a difference of 10,915.96 units (Table 1). The local gate displays 33.18 of the 80 recommendations and has mean cost 24,386.70. It uses 58.5% fewer displays than blanket advice, with a 3.0% lower mean-cost point estimate. The gap filter and hand rule have similar mean costs while displaying more often.

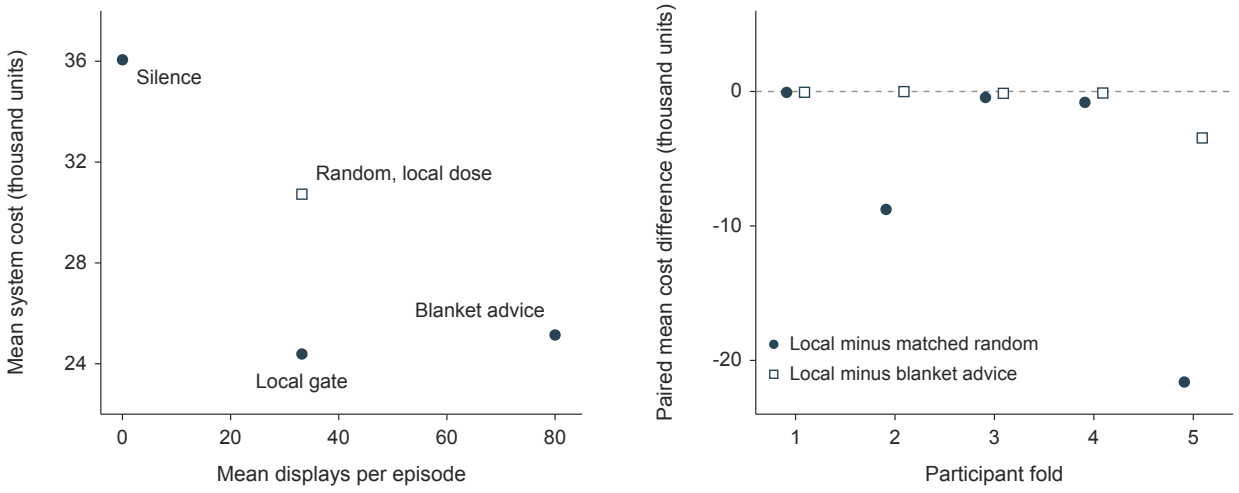
The local gate's cost difference from blanket advice is small relative to variation across fitted populations. Its paired mean difference is -754.67, with fold-mean statistic -1.12, and its pooled paired median is -27.51. These estimates do not establish cost equivalence.

Table 1 System cost and advice frequency.

Policy	Displays	Mean cost	Median cost	Paired mean vs. blanket	Paired median vs. blanket
Silence	0.00	36,057.34	1,210.78	+10,915.96	+178.32
Blanket advice	80.00	25,141.38	1,060.56	0.00	0.00
Chain-wide gate	52.11	28,041.45	1,017.82	+2,900.07	+14.60
Local gate	33.18	24,386.70	899.43	-754.67	-27.51
Gap filter	69.98	24,496.33	970.87	-645.05	-13.22
Hand rule	55.27	24,478.41	971.07	-662.97	-19.52
Random, chain-wide dose	52.11	27,849.24	1,096.09	+2,707.86	+30.48
Random, local dose	33.18	30,726.83	1,163.86	+5,585.45	+87.27

Each policy is evaluated in 1,500 paired worlds, with 300 per participant fold. Displays are means out of 80 opportunities. Cost is total holding and backlog cost across four positions and 24 weeks. Negative paired differences favor the row policy. A paired median is the median of within-world differences, not the difference between marginal medians. The random controls match the named gate's realized dose. Figure EC.1 and Table EC.10 report variation across folds.

Random placement of the same number of displays has mean cost 30,726.83. Relative to this control, the local gate lowers the paired mean by 6,340.12 units and the pooled paired median by 199.27. Every fold mean and median favors the local gate, although the magnitudes vary widely (Figure 6); the fold-median statistic is -1.64, below the nominal absolute reference threshold of 2.776. The lower cost in these fitted populations is associated with where the gate places its displays, rather than fewer displays alone. The count is taken retrospectively from the local gate's trajectory, so this comparison does not identify an optimal attention budget.



(a) Cost and display frequency (b) Paired cost differences by fold

Figure 6 Targeted delivery and matched-frequency random timing. Panel (a) shows sample mean system cost and displays for silence, blanket advice, the local gate, and its random-timing control across 1,500 paired worlds. Panel (b) shows the local gate's mean within-world cost difference from its random control (circles) and from blanket advice (squares), separately for each 300-world participant fold. Negative differences favor the local gate. Points show the observed fold summaries; Table EC.11 gives their t_4 reference intervals.

The two comparisons can be expressed relative to the operating value of blanket advice. Define the share of its mean saving as $(\bar{C}^0 - \bar{C}^\pi)/(\bar{C}^0 - \bar{C}^{\text{blanket}})$, where $\bar{C}$ is the mean cost across paired worlds and 0 denotes silence. The local gate accounts for 107% of that saving, compared with 49% for random timing at its dose. These ratios summarize sample means. The value above 100% reflects the local gate's lower mean-cost point estimate, not a resolved advantage over blanket advice. Their dependence on high-cost trajectories makes the distribution of cost relevant to this interpretation.

The chain-wide gate does not show the same mean-cost gain over its matched control. It displays 52.11 recommendations and has mean cost 28,041.45, compared with 27,849.24 for random timing at its own dose. Its paired mean difference is $+192.21$, while its pooled paired median is -15.05. Relative to the chain-wide gate, the local gate's pooled paired median is -45.00, and four of five fold medians favor it (fold-median statistic -2.28).

The distinction is between fitted delivery policies with different inputs, not between the best achievable values of those inputs. A chain-wide policy can reproduce any local policy by ignoring additional information. The learned policies also reflect estimation and optimization under the normalized training objective in Section 4.6. Their comparison therefore measures how effectively each fitted gate uses its available information; it does not establish local-information sufficiency.

5.2 The distribution of cost

Expected-cost comparisons give substantial weight to a small number of trajectories. For the local gate, mean cost is 24,386.70 and marginal median cost is 899.43; its 75 highest-cost worlds account for 45.7% of aggregate cost. Under blanket advice, the corresponding share is 45.1%. Large-cost worlds occur under every policy (Table EC.13). The mean consequently answers a different operating question from the paired median: it includes the full magnitude of rare losses, whereas the paired median describes the central within-world comparison. The favorable local-versus-random comparison appears under both summaries, but the share of blanket advice's saving is specifically a mean-cost measure.

A prolonged outage can leave fewer opportunities to clear accumulated orders before the episode ends. Among the 291 worlds with 10–18 disrupted advice-window weeks, the local-minus-chain-wide paired median is -39.99 and the fold-median statistic is -2.62, again below the nominal reference threshold. This comparison describes performance on prolonged-disruption paths. Those paths also differ in onset, capacity depth, and planner draws, so they do not isolate the persistence effect in Proposition 3 (duration comparisons in Section EC.3.2).

5.3 Behavioral fit and interpretation

The policy comparisons depend on the intended orders produced by the fitted planners. Those orders determine both the recommendation gap and the operating state in which advice is considered. The ordering rule has median held-out RMSE 6.43, exceeding the own-mean, pass-through and constant-order benchmarks (Table EC.6). Under silence, simulated factory median orders also exceed the experimental no-advice interval (Table EC.14). These discrepancies limit how closely the evaluated population represents experimental behavior. The local gate's favorable comparison with random timing is a result for that fitted population; applying the policy elsewhere requires assessing the ordering and response processes that generate its opportunities.

6 Managerial Implications

A firm deploying an AI advisor must decide which recommendations reach the planner, what information guides that decision, and how the resulting orders are evaluated. System cost provides the link between these choices. A useful current adjustment can consume responsiveness or capacity needed later, so the quality of the recommendation alone does not determine the value of showing it.

6.1 How much advice to show

Firms preparing for disruption reserve inventory and capacity for later needs. When displays reduce later responsiveness, advice creates a similar allocation problem. A modest correction while deliveries remain constrained may use response that would be more valuable for a later adjustment. The reservation value

prices that forgone opportunity. Its inputs are the benefit of the current adjustment, the response retained after a display, and the recipient's remaining opportunities to act. A fixed cap on displays does not by itself account for any of them.

This distinction makes random timing a useful benchmark for a delivery interface. In the fitted chain, the local gate uses 41% of blanket advice's displays, and its paired mean and median costs are lower than those of random timing at the same count in every fold. The comparison with blanket advice does not establish cost equivalence. The matched-count comparison nevertheless shows why display frequency should be evaluated together with the adjustments selected: reducing interruptions and selecting useful opportunities are different properties of an interface.

Acceptance rates also need an operating interpretation. Advice that is easy to follow may have little system value, and a high acceptance rate can reward such recommendations. Recording intended orders, recommendations, and final orders identifies the adjustments that follow display; subsequent inventories and backlogs supply their operating context. Delivery histories should be kept by recipient because the response state belongs to the planner receiving advice. These records support estimating the current and continuation values in the model, including the response transitions in Section 3.7.

6.2 When recovery information changes advice

A supplier's revised recovery estimate can change which order adjustment deserves a display even if the recommendation itself is unchanged. A longer sequence of otherwise comparable opportunities raises the reservation value because there are more chances to use preserved responsiveness (Proposition 3). The forecast therefore informs advice timing as well as the recommended order (Mehrotra and Schmidt 2021). For the retailer awaiting shipments, the relevant horizon includes useful adjustments after production resumes, while outstanding orders and delayed shipments are still being resolved.

Persistence can also determine whether extra information about an adjustment is worth acquiring. In the two-benefit model, equally likely benefits of 1 and 4 with retention $\alpha = 0.6$ give the distinction no value at any horizon when the probability of another opportunity is at most 0.625. At persistence 0.8, the signal becomes valuable with three opportunities remaining (Corollary EC.3). The signal never identifies harmful advice. It becomes useful because the longer sequence makes deferring the smaller benefit worthwhile. This analytical example explains why the information needed to recommend an order need not be the same as the information needed to time its delivery.

The horizon is only one part of a recovery update. A forecast of a deeper capacity cut may also change which adjustments are feasible and valuable. This differs from multiplying every current and future benefit by the same factor, which leaves optimal displays unchanged. A firm revising its recovery plan should therefore update both the expected sequence of opportunities and their operating benefits. The delivery decision uses the forecast available at that time, not the recovery path subsequently realized.

6.3 Coordinating advice with upstream capacity

The value of an accepted recommendation depends on the capacity it leaves for subsequent decisions. In the shared-capacity model, greater responsiveness makes the current adjustment more likely to succeed, but also makes it more likely to displace the next use of capacity. Proposition 4 shows that this second effect can justify less advice to a more responsive planner. An interface that rewards acceptance alone can therefore prioritize the wrong adjustment. Delivery should use the same system-cost objective and capacity constraints as the operating plan.

The bounded withholding region in Corollary 3 makes the competition for capacity explicit. When little capacity remains, the later adjustment has little benefit to protect; when capacity is abundant, both adjustments fit. At intermediate capacity, the current adjustment can displace a more valuable later use. Applying this comparison requires the usable capacity over the replenishment horizon and the value of the competing adjustments. The announced percentage loss of factory capacity alone supplies neither.

For a retailer facing a constrained supplier, the operational question is whether the advised order improves fulfillment over the relevant horizon or mainly adds to outstanding requests. The bullwhip literature likewise directs attention to the upstream consequences of local ordering (Lee et al. 1997, Chen et al. 2000). Our capacity result concerns the displacement of later adjustments. It gives a reason to evaluate advice against the chain's recovery needs even when the recipient is likely to follow it and its local inventory benefit is positive.

6.4 Choosing information for the delivery decision

The value-of-information results also clarify the role of upstream visibility. A signal can leave today's adjustment beneficial while revealing that its benefit is too small to justify a display now. At the final opportunity, distinguishing two beneficial recommendations has no value because both should be shown. Earlier, preserving response for the larger benefit can make the same signal valuable. An interface's information requirements therefore depend on the opportunities that remain, even when recommendation quality is unchanged.

A firm can assess an information investment by comparing delivery decisions with and without the signal and evaluating the system-cost consequences where they disagree. Aggregate prediction accuracy can obscure information that matters in a small number of costly states. In the independent-opportunity model, uniformly bounded errors can change delivery only near the reservation boundary, and a bounded density of benefits there makes the resulting loss quadratic in estimation error (Corollary EC.1). This directs estimation effort toward decisions where the current benefit and the value of waiting are difficult to distinguish.

Simple rules based on the planner's state and disagreement with the advisor provide a practical comparison for a richer interface. The local gate's favorable performance in the fitted chain does not show that

upstream information is dispensable: the chain-wide gate also reflects the way that information was represented and learned. Additional visibility is valuable to the firm when its delivery policy uses it to improve consequential adjustments. That assessment can change as the disruption progresses and capacity becomes available again.

6.5 Scope

The behavioral evidence comes from one manager population in a controlled ordering task. The experiment distinguishes intended orders from subsequent adjustments, but does not identify a structural response-depreciation rate. The computational comparisons use the fitted ordering and response processes described in Section 5.3. Deterministic demand and factory-capacity shocks define their operating scope. Variation across five overlapping fits does not capture all uncertainty in calibration or policy training. The analytical results give conditional prescriptions for allocating advice; their implementation requires estimates of response and operating consequences in the firm's setting.

7 Conclusion

During a supply-chain disruption, planners make a sequence of connected order adjustments. An AI advisor can supply a recommendation at each decision, but the value of showing it depends on what happens next. We study this delivery decision while holding the recommendation rule fixed and leaving ordering authority with the planner. The central tradeoff is between influencing the current adjustment and preserving the ability to influence a later one. When adoption changes available capacity, displaying advice can also alter the adjustments the chain will be able to make.

The analysis identifies conditions under which useful advice is worth withholding. If displays reduce subsequent responsiveness, the optimal rule compares the current benefit with an endogenous reservation value. A longer sequence of opportunities and greater disruption persistence raise that value; a common increase in current and future benefits preserves the delivery decisions. This allocation problem also changes the value of information. A signal that only distinguishes among beneficial recommendations can be worthless at the final decision yet valuable earlier. With shared capacity, a more responsive planner can optimally receive less advice because acting now is more likely to displace a valuable later adjustment. The bounded withholding region makes the competing uses of capacity explicit.

Recording intended and final orders separates a planner's initial choice from the adjustment following a recommendation. Participants revise orders much more often on decisions with a display, and the peak-order rank comparison indicates that advice can accompany larger requests during disruption, with sensitivity to specification and multiple-testing scope. The separately fitted ordering and response processes populate a chain in which delivery policies change both the frequency and placement of advice. The local gate shows 41% as many recommendations as blanket advice; paired mean and median cost differences favor it over

random timing at the same count in every participant fold. Fewer displays alone do not reproduce the local gate's performance in these fitted populations. Cost estimates remain sensitive to high-cost trajectories and variation across fits.

For managers, the results favor evaluating advice delivery against the sequence of operating decisions it is intended to improve. Acceptance alone misses the upstream consequences of the order; display frequency alone misses which adjustments were encouraged or forgone. A useful evaluation records the intended order, recommendation, final order, and subsequent system cost, and compares targeting with both blanket advice and random timing at the same frequency. Recovery expectations and available capacity enter this assessment because they determine which future adjustments a current display may displace.

Two extensions would deepen these prescriptions. Response dynamics that depend on the quality of past advice would connect recommendation content to the reservation value. Allocating advice among several planners sharing capacity would connect delivery timing to the choice of recipient. Those decisions belong in the operating model because advice can change human responsiveness and redirect scarce supply. Designing when AI speaks is therefore part of coordinating a supply chain through disruption and recovery.

References

Abel, David, D. Ellis Hershkowitz, Michael L. Littman. 2016. Near optimal behavior via approximate state abstraction. *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48. 2915-2923.

Balakrishnan, Maya, Kris Johnson Ferreira, Jordan Tong. 2026. Human-algorithm collaboration with private information: Naive advice-weighting behavior and mitigation. *Management Science*, 72 (1), 265-284.

Bastani, Hamsa, Osbert Bastani, Wichinpong Park Sinchaisri. 2026. Improving human sequential decision making with reinforcement learning. *Management Science*, 72 (1), 733-755.

Baucells, Manel, Lin Zhao. 2019. It is time to get some rest. *Management Science*, 65 (4), 1717-1734.

Bendoly, Elliot, Rachel Croson, Paulo Goncalves, Kenneth Schultz. 2010. Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, 19 (4), 434-452.

Boyaci, Tamer, Caner Canyakmaz, Francis de Vericourt. 2024. Human and machine: The impact of machine input on decision making under cognitive limitations. *Management Science*, 70 (2), 1258-1275.

Boyaci, Tamer, Caner Canyakmaz, Francis de Vericourt. 2026. Beyond the black box: Unraveling the role of explainability in human-artificial intelligence collaboration. *Management Science*, Articles in Advance.

Cachon, Gérard P., Marshall Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science*, 46 (8), 1032-1048.

Caro, Felipe, Anna Sáez de Tejada Cuenca. 2023. Believing in analytics: Managers' adherence to price recommendations from a DSS. *Manufacturing & Service Operations Management*, 25 (2), 524-542.

Chen, Frank, Zvi Drezner, Jennifer K. Ryan, David Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, 46 (3), 436-443.

Croson, Rachel, Karen Donohue. 2003. Impact of POS data sharing on supply chain management: An experimental study. *Production and Operations Management*, 12 (1), 1-11.

Croson, Rachel, Karen Donohue. 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science*, 52 (3), 323-336.

Croson, Rachel, Karen Donohue, Elena Katok, John Sterman. 2014. Order stability in supply chains: Coordination risk and the role of coordination stock. *Production and Operations Management*, 23 (2), 176-196.

Cunningham, Caitlin, Rebekah Baik, Leela Nageswaran, Leonard Boussioux. 2026. Decision making on draft: Human–AI collaboration in supply chains. SSRN Working Paper 6744822.

Derman, Cyrus, Gerald J. Lieberman, Sheldon M. Ross. 1972. A sequential stochastic assignment problem. *Management Science*, 18 (7), 349-355.

Gaur, Vishal, Avi Giloni, Sridhar Seshadri. 2005. Information sharing in a supply chain under ARMA demand. *Management Science*, 51 (6), 961-969.

Giacchetta, Carlotta, Alessandro Bogani, Cesare Barbera, Giovanni De Toni, Michele Caprio, Andrea Pugnana, Andrea Passerini. 2026. Should I stay or should I show? Learning to selectively disclose information. arXiv:2609.39818.

Gino, Francesca, Gary Pisano. 2008. Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10 (4), 676-691.

Grand-Clément, Julien, Jean Pauphilet. 2026. The best decisions are not the best advice: Making adherence-aware recommendations. *Management Science*, 72 (1), 667-692.

Gurvich, Itai, Kevin J. O'Leary, Lu Wang, Jan A. Van Mieghem. 2020. Collaboration, interruptions, and changeover times: Workflow model and empirical study of hospitalist charting. *Manufacturing & Service Operations Management*, 22 (4), 754-774.

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2), 65-70.

Ibrahim, Rouba, Song-Hee Kim, Jordan Tong. 2021. Eliciting human judgment for prediction algorithms. *Management Science*, 67 (4), 2314-2325.

Imdahl, Christina, William Schmidt, Kai Hoberg. 2026. Targeted automation and sustaining human–AI learning. *Production and Operations Management*, 35 (4), 1520-1538.

Inderfurth, Karl, Abdolkarim Sadrieh, Guido Voigt. 2013. The impact of information sharing on supply chain performance under asymmetric information. *Production and Operations Management*, 22 (2), 410-425.

Kakade, Sham, John Langford. 2002. Approximately optimal approximate reinforcement learning. *Proceedings of the Nineteenth International Conference on Machine Learning*. 267-274.

Kapuściński, Roman, Rodney P. Parker. 2022. Conveying demand information in serial supply chains with capacity limits. *Operations Research*, 70 (3), 1485-1505.

Kleindorfer, Paul R., Germaine H. Saad. 2005. Managing disruption risks in supply chains. *Production and Operations Management*, 14 (1), 53-68.

Lee, Hau L., V. Padmanabhan, Seungjin Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science*, 43 (4), 546-558.

Lehmann, Cedric A., Christiane B. Haubitz, Andreas Fügener, Ulrich W. Thonemann. 2022. The risk of algorithm transparency: How algorithm complexity drives the effects on the use of advice. *Production and Operations Management*, 31 (9), 3419-3434.

Liu, Yueyang, Wichinpong Park Sinchaisri. 2026. Strategic advice in the age of personal AI. arXiv:2603.02055.

Long, Carol Xuan, David Simchi-Levi, Feng Zhu, Huangyuan Su, Andre P. Calmon, Flavio P. Calmon. 2026. Reliability and effectiveness of autonomous AI agents in supply chain management. arXiv:2605.17036.

Lücker, Florian, Sunil Chopra, Ralf W. Seifert. 2021. Mitigating product shortage due to disruptions in multi-stage supply chains. *Production and Operations Management*, 30 (4), 941-964.

MacKinnon, James G., Halbert White. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29 (3), 305-325.

Mehrotra, Mili, William Schmidt. 2021. The value of supply chain disruption duration information. *Production and Operations Management*, 30 (9), 3015-3035.

Oroojlooyjadid, Afshin, MohammadReza Nazari, Lawrence V. Snyder, Martin Takac. 2022. A deep Q-network for the beer game: Deep reinforcement learning for inventory optimization. *Manufacturing & Service Operations Management*, 24 (1), 285-304.

Papanastasiou, Yiangos, Kostas Bimpikis, Nicos Savva. 2018. Crowdsourcing exploration. *Management Science*, 64 (4), 1727-1746.

Schmidt, William, Ananth Raman. 2022. Operational disruptions, firm risk, and control systems. *Manufacturing & Service Operations Management*, 24 (1), 411-429.

Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov. 2017. Proximal policy optimization algorithms. arXiv:1707.06347.

Sterman, John D. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35 (3), 321-339.

Sterman, John D., Gokhan Dogan. 2015. I'm not hoarding, I'm just stocking up before the hoarders get here: Behavioral causes of phantom ordering in supply chains. *Journal of Operations Management*, 39–40 6-22.

Tomlin, Brian. 2006. On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, 52 (5), 639-657.

Wolczynski, Nicholas, Yotam Liel, Maytal Saar-Tsechansky, Lior Zalmanson. 2026. The differential effects of AI abstention on advice acceptance. arXiv:2610.27258.

Zhang, Zheng, Cuong C. Nguyen, David Rosewarne, Kevin Wells, Gustavo Carneiro. 2026. Fatigue-aware learning to defer via constrained optimisation. arXiv:2604.00904.

Electronic Companion for "When Should AI Stay Silent? Advice Delivery in Turbulent Supply Chains"

EC.1 Proofs and Analytical Extensions

All expectations in this section are finite. Results about one intervention opportunity hold the continuation rule and joint distribution of state and signals fixed. They are not assertions that decision values remain unchanged after a sequence of interventions.

EC.1.1 Proof of Proposition 1

The induction starts from $F_0^{\mathcal{H}}(r) = 0$. Suppose $F_{n-1}^{\mathcal{H}}(r) = rv_{n-1}^{\mathcal{H}}$ for every $r \in [0,1]$. Conditional on the current signal, withholding yields $rv_{n-1}^{\mathcal{H}}$. Displaying yields $rM_{\mathcal{H}} + \alpha rv_{n-1}^{\mathcal{H}}$. The independence assumptions ensure that the same deterministic continuation coefficient applies after either action and after every signal. Maximizing, and then taking expectations before the current signal, gives

$$F_n^{\mathcal{H}}(r) = \mathbb{E}\left[\max\{rv_{n-1}^{\mathcal{H}}, rM_{\mathcal{H}} + \alpha rv_{n-1}^{\mathcal{H}}\}\right] = r\left(v_{n-1}^{\mathcal{H}} + \mathbb{E}[(M_{\mathcal{H}} - dv_{n-1}^{\mathcal{H}})^+]\right).$$

Integrability follows from $\mathbb{E}|M_{\mathcal{H}}| \le \mathbb{E}|Y| < \infty$ and induction. For $r > 0$, the conditional difference has the sign of $M_{\mathcal{H}} - dv_{n-1}^{\mathcal{H}}$, proving the policy statement including ties. At $r = 0$, both actions and all continuations yield zero. Each increment of $v_n^{\mathcal{H}}$ is nonnegative, so both $v_n^{\mathcal{H}}$ and $\theta_n^{\mathcal{H}} = dv_{n-1}^{\mathcal{H}}$ are nondecreasing in n. When $\alpha = 1$, the same maximization has $d = 0$ and yields $v_n^{\mathcal{H}} = n\mathbb{E}[M_{\mathcal{H}}^+]$. □

EC.1.2 An estimated reservation rule

The independent-opportunity model also permits a bound on the cost of estimating the delivery rule. Use the undiscounted, fixed horizon of Proposition 1, with n opportunities and initial response r_0. At opportunity k, let $h_k = n - k + 1$ and let a gate display when $\widehat{M}_k > \widehat{\theta}_k$. Its estimates may depend on observed history and the current signal. Let r_k be responsiveness just before opportunity k, so $r_1 = r_0$. Write $F_n^{\widehat{\pi}}(r_0)$ for its expected benefit.

COROLLARY EC.1 **(Loss from an estimated reservation rule)**. *Suppose deterministic bounds satisfy* $|\widehat{M}_k - M_k| \le \varepsilon_{M,k}$ *and* $|\widehat{\theta}_k - \theta_{h_k}| \le \varepsilon_{\theta,k}$ *almost surely. Put* $\varepsilon_k = \varepsilon_{M,k} + \varepsilon_{\theta,k}$. *Then*

$$0 \le F_n^{\star}(r_0) - F_n^{\widehat{\pi}}(r_0) \le \mathbb{E}^{\widehat{\pi}}\left[\sum_{k=1}^{n} r_k |M_k - \theta_{h_k}| \mathbf{1}\{|M_k - \theta_{h_k}| \le \varepsilon_k\}\right] \le r_0 \sum_{k=1}^{n} \varepsilon_k. \tag{EC.1}$$

If the benefit signal has a density bounded by L_k *on* $[\theta_{h_k} - \varepsilon_k, \theta_{h_k} + \varepsilon_k]$, *the last bound can be replaced by* $r_0 \sum_{k=1}^{n} L_k \varepsilon_k^2$.

Proof. With optimal continuation, the advantage of display is $r_k(M_k-\theta_{h_k})$. An incorrect action loses $r_k|M_k-\theta_{h_k}|$ relative to that continuation; an action chosen at a true tie loses zero. Because the estimated advantage differs in its per-unit-response component by at most ε_k, an incorrect action requires $|M_k-\theta_{h_k}|\le\varepsilon_k$. Summing the one-step losses along the estimated policy's own state distribution telescopes the optimal continuation values, leaving $F_n^\star-F_n^{\hat\pi}$. The terminal continuation is zero. This specializes the performance-difference argument in Proposition EC.6. Since responsiveness never increases, $r_k\le r_0$, which gives the linear bound. Under the density condition, r_k is determined before the independent current signal and

$$\mathbb{E}[|M_k-\theta_{h_k}|\mathbf{1}\{|M_k-\theta_{h_k}|\le\varepsilon_k\}]\le L_k\int_{-\varepsilon_k}^{\varepsilon_k}|u|\,du=L_k\varepsilon_k^2.$$

Taking expectations and summing proves the final claim. □

The result bounds implementation loss conditional on benefit and threshold accuracy. It does not supply those accuracy guarantees from fitted data, and it applies to the independent-opportunity model. Physical feedback changes the appropriate continuation comparison. In particular, an expected saving that already averages over adoption must be compared with $r_k\theta_{h_k}$, not with the per-adoption threshold θ_{h_k}.

EC.1.3 Proof of Proposition 6

The finite horizon, finite state space, and integrability of M imply finite values by induction. Conditional on z and the observed benefit m, withholding gives $(QV_{n-1})(z)$. Display gives $r_z m+(PV_{n-1})(z)$. Their difference is $r_z m-H_n(z)$, so taking the larger of the two values and then integrating over M proves (18). Division by r_z gives the threshold only when $r_z>0$. At $r_z=0$, display remains valuable when it improves continuation value, so the undivided comparison is required. If $H_n(z)<0$ and $r_z>0$, every m strictly between $H_n(z)/r_z$ and zero is negative and favors display. For example, with two states having $r_0=0$ and $r_1=1$, deterministic benefit one, Q equal to the identity, and display moving state zero to state one, $V_1=(0,1)$ and $H_2(0)=-1$. A display that has no current adoption can then improve the next decision. This is an analytical example of a response transition, not an estimate of the experimental response process. □

The scalar model is recovered by allowing responsiveness itself to be the state, with deterministic transitions to r after withholding and αr after display. Because $V_{n-1}(r)=rv_{n-1}$, the continuation difference is $r(1-\alpha)v_{n-1}$ and the display comparison reduces to $M>\theta_n$. The same Bellman argument applies on this bounded continuous state space. For a finite observable-state implementation, row-stochastic transition matrices must be estimated on the same display clock used by the gate. If the transition depends on the current benefit, replacing P by $P(m)$ yields a benefit-dependent continuation difference; a single state-specific threshold then requires additional restrictions.

EC.1.4 Proof of Corollary 1

For $r > 0$, Proposition 1 compares the observed current benefit $D^S(K)$ with θ_n. On $q_0 < K < q_1$, their difference is $b\delta - \kappa(q_1 - K) - \theta_n$. If $0 \le \theta_n < b\delta$, this difference is strictly negative at q_0 and strictly positive at q_1, because $\kappa > b$. It has the unique zero in (5). The outer segments of D^S have values $(b-\kappa)\delta$ and $b\delta$, giving the global comparison and the two boundary cases. The derivative of K_n with respect to θ_n is $1/\kappa$. Monotonicity in the remaining horizon follows when the continuation distribution is fixed. At $r = 0$, both actions remain equivalent regardless of capacity. □

EC.1.5 Proof of Proposition 2

The tower property gives $M_{\mathcal{G}} = \mathbb{E}[M_{\mathcal{F}} \mid \mathcal{G}]$. Conditional Jensen's inequality for the convex function $x \mapsto (x - dv)^+$ therefore implies $J(v) \ge 0$ for each fixed v. We also need monotonicity of the Bellman map in its continuation argument. For $u \ge v$ and each real m, the positive-part function satisfies

$$-d(u-v) \le (m-du)^+ - (m-dv)^+ \le 0.$$

Taking expectations and adding $u - v$ shows

$$\alpha(u-v) \le T_{\mathcal{H}}(u) - T_{\mathcal{H}}(v) \le u - v. \tag{EC.2}$$

Thus $T_{\mathcal{H}}$ is nondecreasing and has secant slopes between α and one, without requiring a density or differentiability.

Assume inductively that $\Delta_{n-1} \ge 0$. Adding and subtracting $T_{\mathcal{F}}(v_{n-1}^{\mathcal{G}})$ yields

$$\begin{aligned}\Delta_n &= T_{\mathcal{F}}(v_{n-1}^{\mathcal{F}}) - T_{\mathcal{G}}(v_{n-1}^{\mathcal{G}}) \\ &= \left[T_{\mathcal{F}}(v_{n-1}^{\mathcal{F}}) - T_{\mathcal{F}}(v_{n-1}^{\mathcal{G}})\right] + J(v_{n-1}^{\mathcal{G}}).\end{aligned}$$

If $\Delta_{n-1} > 0$, define a_n as the bracketed difference divided by Δ_{n-1}; (EC.2) places it in $[\alpha, 1]$. If $\Delta_{n-1} = 0$, choose any a_n in that interval. The recursion follows, and its terms are nonnegative. Since $\Delta_0 = 0$, induction establishes $\Delta_n \ge 0$ for all n. Iterating the recursion expresses each $J(v_j^{\mathcal{G}})$ with a coefficient equal to a product of $n-1-j$ such secant slopes. Bounding each factor between α and one proves (8), including the empty product equal to one.

For (9), put $A_v = \mathbb{E}[(M_{\mathcal{F}} - dv)^+ \mid \mathcal{G}]$ and $B_v = \mathbb{E}[(dv - M_{\mathcal{F}})^+ \mid \mathcal{G}]$. Then $M_{\mathcal{G}} - dv = A_v - B_v$. Hence

$$J(v) = \mathbb{E}[A_v - (A_v - B_v)^+] = \mathbb{E}[\min\{A_v, B_v\}].$$

The nonnegative minimum has positive expectation if and only if it is positive on a set of positive probability. This requires both strictly positive conditional parts there. For $\alpha > 0$, every lower-bound coefficient in (8) is strictly positive. Consequently $\Delta_n = 0$ exactly when $J(v_j^{\mathcal{G}}) = 0$ for $j = 0, \dots, n-1$. This last equivalence need not follow when $\alpha = 0$, because earlier gains can have zero propagation coefficients. □

EC.1.6 Proof of Corollary 2

Under the coarse signal, $M_{\mathcal{G}} = \mu > 0$. Induction in (4) gives $v_n^{\mathcal{G}} = \mu(1-\alpha^n)/(1-\alpha)$: its reservation value is $\mu(1-\alpha^{n-1}) < \mu$ when $\alpha > 0$. At $\alpha = 0$, the coarse gate is indifferent after the last opportunity, so always displaying remains optimal. The fine-signal gain at the coarse continuation is

$$J(v_j^{\mathcal{G}}) = p\big(\mu(1-\alpha^j) - y_L\big)^+,$$

because $\mu(1-\alpha^j) \le \mu < y_H$. These terms are nondecreasing in j. For $\alpha > 0$, Proposition 2 makes $\Delta_n > 0$ equivalent to positivity of at least one of the first n terms, hence to positivity of the last one. For $\alpha = 0$ and $n \ge 2$, the last term is strictly positive and enters the lower bound with coefficient one; for $n = 1$ it is zero. This proves (10) including ties and $0^0 = 1$. At fixed finite n, its left side tends to zero as $\alpha \uparrow 1$. At fixed $\alpha < 1$, it tends to $\mu > y_L$ as $n \to \infty$. These limits yield the last two statements. □

EC.1.7 Proof of Corollary EC.2

COROLLARY EC.2 **(The long-horizon limit)**. *If $M_{\mathcal{H}}$ has a finite essential supremum $\overline{m}_{\mathcal{H}}$, then*

$$\lim_{n\to\infty} v_n^{\mathcal{H}} = \frac{\max\{0, \overline{m}_{\mathcal{H}}\}}{1-\alpha}. \tag{EC.3}$$

The limiting difference between fine and coarse values, per unit of initial responsiveness, is therefore $(\max\{0, \overline{m}_{\mathcal{F}}\} - \max\{0, \overline{m}_{\mathcal{G}}\})/(1-\alpha)$.

Proof. Write $b_{\mathcal{H}} = \max\{0, \overline{m}_{\mathcal{H}}\}$ and $d = 1-\alpha > 0$. If $b_{\mathcal{H}} = 0$, $M_{\mathcal{H}} \le 0$ almost surely and the recursion remains at zero. Otherwise, $T_{\mathcal{H}}(b_{\mathcal{H}}/d) = b_{\mathcal{H}}/d$, since $M_{\mathcal{H}} \le b_{\mathcal{H}}$ almost surely. Monotonicity of $T_{\mathcal{H}}$ and $v_0 = 0$ bound every v_n between zero and $b_{\mathcal{H}}/d$. The sequence is nondecreasing and hence converges to a finite v_∞. The Lipschitz property (EC.2) permits passage to the limit in $v_n = T_{\mathcal{H}}(v_{n-1})$, giving $\mathbb{E}[(M_{\mathcal{H}} - dv_\infty)^+] = 0$. Therefore $M_{\mathcal{H}} \le dv_\infty$ almost surely, so $b_{\mathcal{H}} \le dv_\infty$. Combining with the upper bound proves the limit. Applying the result separately to both information sets and subtracting gives the stated limiting information premium. □

EC.1.8 A capacity instance with no last-opportunity information value

The example in Section 3.4 can be realized within the capacity benchmark, without a free-standing benefit distribution. Set $q_0 = 4$, $q_1 = 8$, $b = 1$, $\kappa = 2$, and $\ell \ge 8$. Let capacity independently equal 6.5 or 8 with equal probability. Equation (EC.7) then gives benefits 1 and 4, both strictly positive. A fine signal observes current capacity and a coarse signal does not. With $\alpha = 0.2$, $v_1 = 2.5$ under both signals and the second-opportunity reservation value is 2. The corresponding capacity threshold is 7. The fine signal distinguishes the two sides of this threshold even though both capacities exceed the operational threshold $K^\star = 6$. This delivers $v_2^{\mathcal{F}} = 2.5 + \frac{1}{2}(4-2) = 3.5$ and $v_2^{\mathcal{G}} = 2.5 + (2.5-2) = 3$. These parameters illustrate an analytical possibility and are not calibrated to the computational experiment.

EC.1.9 Proof of Proposition 3 and Remark 1

Suppose the value with $n-1$ opportunities is $rv_{n-1}(q)$. Withholding gives $qrv_{n-1}(q)$, while displaying gives $rM + q\alpha rv_{n-1}(q)$. Taking the maximum and averaging over the current signal yields (11). This also proves linearity in r and the threshold rule, including indifference at the boundary. For $v \geq 0$, the map $T_q(v) = \mathbb{E}[\max\{qv, M + \alpha qv\}]$ is nondecreasing in both q and v. Starting at $v_0 = 0$, induction gives nonnegativity, monotonicity in q, and monotonicity in n. Multiplying by $(1-\alpha)q$ yields the stated reservation comparisons. Finally $T_q(sv; sM) = sT_q(v; M)$ for $s > 0$. Induction gives the scaling identity; both sides of the display comparison are multiplied by the same positive number. □

For the stationary characterization, $T_q(v) = \mathbb{E}[\max\{qv, M + \alpha qv\}]$ maps $[0,\infty)$ into itself. Both affine terms have slopes in $[0,q]$, so $|T_q(v) - T_q(w)| \leq q|v-w|$. Integrability of M makes the map finite. For $q < 1$ it is a contraction and has a unique fixed point v_∞, to which $v_n = T_q^n(0)$ converges. Moreover, $0 \leq v_\infty \leq \mathbb{E}[M^+]/(1-q)$. Its fixed-point equation gives $(1-q)v_\infty = \mathbb{E}[(M-(1-\alpha)qv_\infty)^+]$. Substituting $\theta = (1-\alpha)qv_\infty$ yields (12). When $q > 0$, its left side is strictly increasing and its right side is continuous and nonincreasing, with the right side bounded above by $(1-\alpha)q\mathbb{E}[M^+]/(1-q)$; hence the nonnegative solution is unique. At $q = 0$ the solution is $\theta = 0$. The same fixed point is the infinite-horizon optimum: the expected benefit omitted after opportunity n is bounded uniformly over policies by $q^n\mathbb{E}[M^+]/(1-q)$, which tends to zero. Thus finite-horizon optimization converges in value and the stationary Bellman selector is optimal. □

EC.1.10 Persistence and the value of information

Use the independent-opportunity model of Proposition 3, with $0 \leq \alpha < 1$ and continuation probability $q \in [0,1]$. Let benefits take the two positive values $y_L < y_H$ with probabilities p and $1-p$, where $0 < p < 1$. The fine signal reveals the benefit, the coarse signal reveals nothing, and $\mu = py_L + (1-p)y_H$.

COROLLARY EC.3 **(A persistence threshold for information)**. *For positive initial responsiveness, fine information is strictly valuable with at most n opportunities exactly when*

$$(1-\alpha)q\mu\frac{1-(\alpha q)^{n-1}}{1-\alpha q} > y_L. \tag{EC.4}$$

With $0^0 = 1$, this condition also covers $n = 1$ and $\alpha q = 0$. Define

$$q_c = \frac{y_L}{(1-\alpha)\mu + \alpha y_L}. \tag{EC.5}$$

If $q \leq q_c$, fine and coarse values coincide at every finite horizon. If $q > q_c$, fine information is strictly valuable at all sufficiently long horizons. The cutoff q_c increases strictly with response retention α.

Proof. The coarse gate's value coefficient is

$$v_n^{\mathcal{G}}(q) = \mu\frac{1-(\alpha q)^n}{1-\alpha q}.$$

To verify this expression, its implied reservation value is at most μ, since $(1-\alpha)q \le 1-\alpha q$. Display is therefore weakly optimal, and its Bellman recursion reduces to $v_n^{\mathcal{G}} = \mu + \alpha q v_{n-1}^{\mathcal{G}}$. At a common continuation coefficient $v_j^{\mathcal{G}}$, the fine-signal gain is

$$J_q(v_j^{\mathcal{G}}) = p\big((1-\alpha)qv_j^{\mathcal{G}} - y_L\big)^+.$$

These gains are nondecreasing in j. The fine Bellman map $T_{\mathcal{F},q}(v) = qv + \mathbb{E}[(Y-(1-\alpha)qv)^+]$ is nondecreasing, with secant slopes in $[\alpha q, q]$. Hence a positive final gain $J_q(v_{n-1}^{\mathcal{G}})$ implies a positive value difference. If this gain is zero, all earlier gains are zero; induction then makes the fine and coarse values coincide through horizon n. Substituting the coarse value proves (EC.4).

The left side converges to $(1-\alpha)q\mu/(1-\alpha q)$ and never exceeds that limit. The limit is strictly above y_L exactly when $q > q_c$. This proves both horizon statements, including equality at the cutoff. Finally,

$$\frac{\partial q_c}{\partial \alpha} = \frac{y_L(\mu - y_L)}{\{\mu - \alpha(\mu - y_L)\}^2} > 0.$$

□

EC.1.11 Proof of Proposition 4

At the final opportunity, displaying is weakly optimal because $B(k) \ge 0$ and there is no subsequent response or capacity value. Withholding initially therefore yields $rB(K)$. Displaying initially yields

$$rA + \alpha r\{(1-r)B(K) + rB(K-\delta)\}.$$

The two capacity states correspond to ignoring and adopting the first recommendation, respectively. Subtracting $rB(K)$ gives (13). Since $r > 0$, its sign is the sign of the term in braces. The derivative of the reservation benefit with respect to r is $\alpha[B(K) - B(K-\delta)] \ge 0$, with the asserted strictness condition. At $r = 0$, both policies have zero value; this degenerate case does not support a strict preference. The formula remains valid at $\alpha = 0$ and $\alpha = 1$. □

EC.1.12 Proof of Corollary 3

For $K \ge \delta$, the reservation benefit in Proposition 4 is

$$R(K) = \begin{cases} db_2 K + \alpha r b_2 \delta, & \delta \le K \le L, \\ db_2 L + \alpha r b_2 (L + \delta - K), & L < K < L + \delta, \\ db_2 L, & K \ge L + \delta. \end{cases}$$

It is continuous, strictly increasing on the first segment, strictly decreasing on the second, and constant on the third. Condition (14) places A strictly above $R(\delta)$ and the final plateau, and strictly below $R(L)$. There

are exactly two crossings. Solving $R(K) = A$ on the first and second segments gives K_- and K_+, respectively, with the asserted locations and actions. For parameters satisfying the strict inequalities,

$$\frac{\partial K_-}{\partial r} = -\frac{\alpha\delta}{d} < 0, \qquad \frac{\partial K_+}{\partial r} = \frac{A - db_2 L}{\alpha b_2 r^2} > 0.$$

The comparative static is local to this parameter region; at its boundaries a display region can disappear or a crossing can coincide with a segment endpoint. □

EC.1.13 Proof of Proposition 5

Differentiating the reservation benefit gives (16). Since $B'(K) > 0$, its sign is the sign of $c - \rho(K)$, where $c = 1 + (1-\alpha)/(\alpha r)$. If ρ is nondecreasing, this sign can change only from positive to negative. The function R_r is therefore nondecreasing up to a possibly empty plateau and nonincreasing afterward. Its strict superlevel sets are intervals. This conclusion includes the cases in which it is monotone over the entire domain.

When $B(k)$ converges to B_∞, the increment $B(K) - B(K-\delta)$ converges to zero, so $R_r(K) \to (1-\alpha)B_\infty$. Condition (17) places the initial value and the tail strictly below A, and some interior value strictly above it. Continuity and the single-peaked shape imply a nonempty bounded strict superlevel interval. Advice is displayed outside it, with either action optimal at equality. The stated upper bound on ρ makes the derivative nonnegative everywhere. Finally,

$$\frac{\partial R_r(K)}{\partial r} = \alpha[B(K) - B(K-\delta)] \geq 0,$$

so the strict withholding sets are nested as r increases, without requiring concavity or the ratio condition. □

If B is twice differentiable and $h(k) = -B''(k)/B'(k)$ is nondecreasing, then

$$\log \rho(K) = \int_{K-\delta}^{K} h(u)\, du$$

is nondecreasing, establishing the stated sufficient condition. With exponential saturation, $B(k) = B_\infty(1 - e^{-\beta k})$, direct substitution instead gives

$$R_r'(K) = \beta B_\infty e^{-\beta K}\{(1-\alpha) - \alpha r(e^{\beta\delta} - 1)\}.$$

Its sign is constant. A strictly concave continuation benefit can therefore produce increasing, decreasing, or constant reservation benefits; concavity by itself is insufficient for the interior-band conclusion. A smooth example is $B(k) = \int_0^k e^{-u^2}\, du$, for which $\rho(K) = e^{2\delta K - \delta^2}$ increases strictly. With $\delta = 0.2$, $\alpha = 0.8$, and $r = 0.75$, the reservation benefit is approximately 0.158 at $K = \delta$, peaks at 0.205, and tends to 0.177. A current benefit of $A = 0.19$ therefore produces an interior withholding interval. The reversal does not require a kink or a hard cap in future benefit.

EC.1.14 Proof of Proposition EC.1

PROPOSITION EC.1 **(Operational and future-response effects)**. *Couple show and withhold transitions using common exogenous innovations. Let their successor states be* (X^0, Z^0) *and* (X^1, Z^1)*, suppressing the index* $k+1$*. Assume the continuation value is defined and integrable also at the hybrid state* (X^1, Z^0)*. Then*

$$\begin{aligned} \Delta_k^\pi(s) &= O_k^\pi(s) + \mathcal{R}_k^\pi(s), \\ O_k^\pi(s) &= \mathbb{E}[c_k^0 - c_k^1 + V_{k+1}^\pi(X^0, Z^0) - V_{k+1}^\pi(X^1, Z^0) \mid s], \\ \mathcal{R}_k^\pi(s) &= \mathbb{E}[V_{k+1}^\pi(X^1, Z^0) - V_{k+1}^\pi(X^1, Z^1) \mid s]. \end{aligned} \tag{EC.6}$$

If immediate physical costs and physical successor states coincide, $O_k^\pi = 0$*. Display can still have nonzero value through* $\mathcal{R}_k^\pi$*.*

Proof. By definition,

$$\Delta_k^\pi(s) = \mathbb{E}[c_k^0 - c_k^1 + V_{k+1}^\pi(X^0, Z^0) - V_{k+1}^\pi(X^1, Z^1) \mid s].$$

Adding and subtracting $V_{k+1}^\pi(X^1, Z^0)$ yields (EC.6). Integrability and the stipulated availability of the hybrid state make both terms well defined. If $c_k^0 = c_k^1$ and $X^0 = X^1$ under the coupling, the operational term vanishes. The remaining term can be positive, zero, or negative. Its attribution depends on the hybrid-state convention; reversing the order in which the two components are changed can give another decomposition with the same total. □

EC.1.15 Proof of Proposition EC.2

The capacity-cost functions and feasible orders are defined in Section 3.1.

PROPOSITION EC.2 **(Capacity-dependent reversal)**. *The order increase reduces focal cost by* $b\delta$*. Its system benefit is*

$$D^S(K) = b\delta - \kappa[(q_1 - K)^+ - (q_0 - K)^+]. \tag{EC.7}$$

This benefit is nondecreasing in capacity. If $\kappa > b$*, its sign changes at*

$$K^\star = q_1 - b\delta/\kappa \in (q_0, q_1). \tag{EC.8}$$

The recommendation raises system cost below $K^\star$ *and reduces it above* $K^\star$*. If* $\kappa \le b$*, its system benefit is nonnegative for every* K*.*

Proof. Because $q_0 < q_1 \le \ell$, $C^L(q_0) - C^L(q_1) = b\delta$. The incremental upstream cost is $\kappa[(q_1 - K)^+ - (q_0 - K)^+]$. Thus

$$D^S(K) = \begin{cases} (b-\kappa)\delta, & K \le q_0, \\ b\delta - \kappa(q_1 - K), & q_0 < K < q_1, \\ b\delta, & K \ge q_1. \end{cases}$$

The function is continuous and nondecreasing. If $\kappa > b$, its middle segment crosses zero exactly once at $q_1 - b\delta/\kappa$, strictly between q_0 and q_1. If $\kappa \le b$, its minimum is nonnegative. □

EC.1.16 Proof of Proposition EC.3

Now suppose there are two remaining advice opportunities. The first has expected operational benefit A. At the second, the realized benefit is W_jB, where $B \geq 0$ is measurable with respect to the information available at that decision and W_j is the response weight following first action j. This linear response-to-benefit relation is an assumption of the reduced-form two-opportunity model. Assume the first display leaves the distribution of B unchanged and affects second-opportunity benefits only through responsiveness. Conditional on all information available for the second display choice, let the response weight have mean $\mu_j \geq 0$ under first action j, with $0 \leq \mu_1 < \mu_0$, and assume these means do not depend on B. Thus the second recommendation is weakly beneficial under either first action. This assumption excludes a later gate exploiting additional signals about the response weight.

PROPOSITION EC.3 **(Preserving future responsiveness)**. *Under these two-opportunity assumptions, define $L = (\mu_0 - \mu_1)\mathbb{E}[B]$. Showing the first recommendation is weakly optimal exactly when $A \geq L$. Positive current benefit alone need not justify display.*

Proof. At the second opportunity the conditional expected benefit from display is $\mu_j B \geq 0$ under first action j. Because no subsequent opportunities remain, display is weakly optimal. Withholding at the first opportunity yields expected second-opportunity savings $\mu_0\mathbb{E}[B]$. Displaying yields $A + \mu_1\mathbb{E}[B]$. Comparing the two gives $A \geq (\mu_0 - \mu_1)\mathbb{E}[B]$. The condition on the second decision's information is essential: marginal nonnegativity of a response mean would not suffice if a gate observed an additional signal separating harmful and helpful responses. □

EC.1.17 Proof of Proposition EC.4

PROPOSITION EC.4 **(Capacity threshold with future-response cost)**. *Combine the capacity benchmark with the two-opportunity model, taking $A = D^S(K)$ and holding $L \geq 0$ fixed as capacity varies. If $\kappa > b$ and $0 \leq L < b\delta$, the display threshold is*

$$K^\dagger = K^\star + L/\kappa = q_1 - (b\delta - L)/\kappa. \tag{EC.9}$$

Display is strictly optimal for $K > K^\dagger$, withholding for $K < K^\dagger$, and both are optimal at equality. If $L > b\delta$, withhold at every capacity; if $L = b\delta$, display is only weakly optimal for $K \geq q_1$.

Proof. Proposition EC.3 gives the first display's net benefit $D^S(K) - L$. If $0 \leq L < b\delta$ and $\kappa > b$, the net benefit is negative at q_0, positive at q_1, and strictly increasing between them. Solving $b\delta - \kappa(q_1 - K) - L = 0$ gives $K^\dagger = q_1 - (b\delta - L)/\kappa$. Its derivative with respect to L, with other primitives fixed, is $1/\kappa > 0$. The maximum operational benefit is $b\delta$, which yields the two boundary cases in the statement. No result is asserted when L itself changes endogenously with K without further restrictions. □

EC.1.18 Proof of Proposition EC.5

For a fixed decision opportunity and continuation rule, let D be an integrable intervention benefit with a fixed joint distribution with the information available to the gate. It can represent the dynamic difference in (3). A narrower order-mediated benchmark for the quadratic examples below holds the future-response convention fixed and defines

$$D(s) = Q^S(a^H \mid s) - Q^S(a^I \mid s), \tag{EC.10}$$

for feasible a^I. These two benefit definitions coincide only when displaying advice has no separate continuation effect on response.

PROPOSITION EC.5 **(Value of information at a fixed opportunity)**. *For $\mathcal{F}$-measurable binary decisions, let $m_{\mathcal{F}} = \mathbb{E}[D \mid \mathcal{F}]$ and $d_{\mathcal{F}} = \mathbf{1}\{m_{\mathcal{F}} > 0\}$. Then*

$$V(\mathcal{F}) = \sup_g \mathbb{E}[gD] = \mathbb{E}[(m_{\mathcal{F}})^+]. \tag{EC.11}$$

If $\mathcal{G} \subseteq \mathcal{F}$, then

$$V(\mathcal{F}) - V(\mathcal{G}) = \mathbb{E}[|m_{\mathcal{F}}| \mathbf{1}\{d_{\mathcal{F}} \neq d_{\mathcal{G}}\}] \geq 0. \tag{EC.12}$$

Also $V(\mathcal{F}) \leq \mathbb{E}[D^+]$, with equality exactly when

$$\min\{\mathbb{E}[D^+ \mid \mathcal{F}], \mathbb{E}[D^- \mid \mathcal{F}]\} = 0 \quad \textit{almost surely}, \tag{EC.13}$$

where $D^- = (-D)^+$. Zero-value states need not be distinguishable from positive or negative states.

Proof. For an $\mathcal{F}$-measurable binary decision, the tower property gives $\mathbb{E}[gD] = \mathbb{E}[gm_{\mathcal{F}}]$. Pointwise maximization yields $d_{\mathcal{F}} = \mathbf{1}\{m_{\mathcal{F}} > 0\}$ and (EC.11). If $\mathcal{G} \subseteq \mathcal{F}$, $d_{\mathcal{G}}$ is also $\mathcal{F}$-measurable. Hence

$$V(\mathcal{F}) - V(\mathcal{G}) = \mathbb{E}[m_{\mathcal{F}}(d_{\mathcal{F}} - d_{\mathcal{G}})].$$

Where the decisions agree the integrand is zero; where they differ it is $|m_{\mathcal{F}}|$, including a zero contribution when $m_{\mathcal{F}} = 0$. This proves (EC.12).

For the ceiling, write $A = \mathbb{E}[D^+ \mid \mathcal{F}]$ and $B = \mathbb{E}[D^- \mid \mathcal{F}]$. Then $m_{\mathcal{F}} = A - B$ and

$$\mathbb{E}[D^+] - V(\mathcal{F}) = \mathbb{E}[A - (A - B)^+] = \mathbb{E}[\min(A, B)].$$

The integrand is nonnegative, so equality holds exactly when $\min(A, B) = 0$ almost surely. Full measurability of the three-valued sign of D is sufficient but unnecessary. For example, when D takes only zero and one and $\mathcal{F}$ is trivial, always displaying attains the ceiling without distinguishing zero from one. □

EC.1.19 Proof of Corollary EC.4

COROLLARY EC.4 **(Capacity information can be strictly valuable)**. *In Proposition EC.2, suppose* $\kappa > b$ *and the gate observes all primitives except* K*. If capacity has positive conditional probability on each side of* $K^\star$*, observing capacity strictly increases expected optimal display benefit, in the absence of a separate exposure cost.*

Proof. Conditional on the gate's original information, the capacity benchmark gives strictly positive and strictly negative values of $D^S(K)$ with positive probability. Therefore $A = \mathbb{E}[(D^S)^+]$ and $B = \mathbb{E}[(D^S)^-]$ are both strictly positive. Observing capacity yields A, whereas without it the best benefit is $(A-B)^+$. Their difference is $\min(A,B) > 0$. The statement is conditional and requires sign crossing, not merely variation in capacity. □

REMARK EC.1 (NESTED INFORMATION IN THE SEQUENTIAL PROBLEM). Consider the same controlled environment, objective, action set, and feasible interventions under two observation systems. Suppose the richer history can reconstruct the restricted history at every decision, including actions and any randomization used by the restricted policy. Every restricted policy can then be implemented by ignoring the additional information. If $\Pi_L \subseteq \Pi_G$ are the resulting admissible classes,

$$\inf_{\pi\in\Pi_G} J(\pi) \le \inf_{\pi\in\Pi_L} J(\pi). \tag{EC.14}$$

The argument also holds for the specified relative-cost objective when it is common to both classes. It requires common dynamics, objectives, and admissible policies. It says nothing about the ordering of two fitted policies.

REMARK EC.2 (ESTIMATED POLICIES AND INCREMENTAL INFORMATION). A comparison of fitted local and chain-wide gates combines differences in representation, approximation, optimization, and estimation. Shared architecture and training budget do not ensure that either policy attains the optimum within its observation class. The sign of a fitted-policy cost contrast therefore does not measure the value of optimal use of the additional information.

EC.1.20 Information loss with endogenous physical transitions

Consider the fully observed model in Section 3.2 with horizon T, binary actions, finite state spaces, finite costs, and a common initial-state distribution. Finite spaces ensure existence of optimal policies and avoid measurable-selection issues. A restricted policy π can use its entire observation history and private randomization. Physical and response transitions can depend on the action. Define Q_k^* using fully observed optimal continuation, $V_k^*(s) = \min_g Q_k^*(s,g)$, and

$$\Delta_k^*(s) = Q_k^*(s,0) - Q_k^*(s,1), \qquad g_k^*(s) = \mathbf{1}\{\Delta_k^*(s) > 0\}.$$

The value J^* is expected cost under the fully observed optimum. All expectations under π below use states actually reached under π, not states sampled under an unrelated reference policy.

PROPOSITION EC.6 **(Decision-relevant information under feedback)**. *For every admissible restricted policy,*

$$J(\pi) - J^* = \mathbb{E}^{\pi}\left[\sum_{k=1}^{T} |\Delta_k^*(S_k)|\, \mathbf{1}\{g_k \neq g_k^*(S_k)\}\right]. \tag{EC.15}$$

Thus a restricted policy attains J^* exactly when it takes an optimal action almost surely at every opportunity along its own trajectory, allowing either action at ties. If restricted observations O_k admit functions $\widehat{\Delta}_k(O_k)$ satisfying

$$|\widehat{\Delta}_k(O_k) - \Delta_k^*(S_k)| \leq \varepsilon_k \tag{EC.16}$$

almost surely under the induced gate $g_k = \mathbf{1}\{\widehat{\Delta}_k(O_k) > 0\}$, then

$$0 \leq J(\pi) - J^* \leq \sum_{k=1}^{T} \varepsilon_k\, \mathbb{P}^{\pi}\{|\Delta_k^*(S_k)| \leq \varepsilon_k\}. \tag{EC.17}$$

Proof. Conditional on the state and chosen action, $Q_k^*(S_k, g_k)$ equals the expected current cost plus V_{k+1}^* at the successor state. Summing and taking expectations under π telescopes the continuation terms, yielding

$$J(\pi) - J^* = \mathbb{E}^{\pi} \sum_{k=1}^{T} \{Q_k^*(S_k, g_k) - V_k^*(S_k)\}.$$

Any common terminal cost is included in V_{T+1}^* and cancels in the same way. With two actions, each bracket equals the summand in (EC.15). All brackets are nonnegative, proving the equality condition. When the estimated and optimal signs disagree, (EC.16) implies $|\Delta_k^*| \leq \varepsilon_k$; this includes a zero estimated advantage when the true advantage is positive. A zero true advantage has zero loss regardless of the tie choice. Bounding each disagreement cost by $\varepsilon_k \mathbf{1}\{|\Delta_k^*| \leq \varepsilon_k\}$ proves (EC.17). □

The identity specializes the performance-difference principle to a finite-horizon binary-action problem (Kakade and Langford 2002). The approximation requirement is also stronger than predicting observed orders or matching a learned gate. It concerns the optimal dynamic advantage, including omitted physical and behavioral state, on the restricted policy's reachable support. In particular, a favorable local-versus-chain-wide fitted-policy contrast does not verify (EC.16). The bound clarifies what a proposed information reduction would need to preserve, in the spirit of approximate state abstraction (Abel et al. 2016).

EC.1.21 A restricted adherence-only gate

PROPOSITION EC.7 **(Absorption under adherence-only stationary gating)**. *Suppose a statistic ϕ depends only on the prefix of response outcomes observed at previous displays and changes only when a new display occurs. A deterministic stationary policy $g = \mathbf{1}\{\phi \in \mathcal{S}\}$, with no dependence on time or other changing state, withholds forever after ϕ first leaves $\mathcal{S}$.*

EC.1.22 Proof of Proposition EC.7

Let N_t count previous displays and $\phi_t = \phi(w_1, \ldots, w_{N_t})$. If $\phi_t \notin \mathcal{S}$, the specified deterministic gate withholds. Then $N_{t+1} = N_t$ and $\phi_{t+1} = \phi_t \notin \mathcal{S}$, so it withholds again. Induction proves permanent withholding. Time dependence, an independently evolving state, randomized probing, or exogenous response recovery can break the induction and lie outside the proposition. □

COROLLARY EC.5 **(Prefix consumption under the restricted gate)**. *Under Proposition EC.7, a planner's realized response draws form a prefix of the exposure-indexed queue. The first withholding time determines its length. That stopping time may depend on observed adherence; the prefix property does not establish that response information has zero value or that all useful policies depend only on a population mean.*

Proof. Each display consumes the next queue entry, withholding consumes none, and the first withholding is permanent. The final number of consumed entries can therefore be a history-dependent stopping index. This establishes the prefix statement without restricting how that index is chosen. □

EC.1.23 Focal and system objectives

Consider a quadratic benchmark over a common feasible interval containing the displayed actions:

$$Q^S(a \mid s) = K_s + H_s(a - a_s^\star)^2, \quad H_s > 0, \tag{EC.18}$$

$$Q^L(a \mid s) = k_s + h_s(a - a_{L,s}^\star)^2, \quad h_s > 0. \tag{EC.19}$$

Let $a^I = a^H + w\delta_s$ be feasible, with $\delta_s = a^A - a^H$.

PROPOSITION EC.8 **(Conflicting focal and system benefits)**. *Under* (EC.18)–(EC.19),

$$D(s) = H_s w\delta_s[2(a_s^\star - a^H) - w\delta_s], \tag{EC.20}$$

$$D^L(s) = h_s w\delta_s[2(a_{L,s}^\star - a^H) - w\delta_s].$$

If the distinct focal and system optima are feasible, there exist feasible interventions with opposite signs of focal and system benefit, in either direction.

Proof. Expand the two differences of squares. Set $a^H = a_s^\star$, $a^A = a_{L,s}^\star$, and $w = 1$. Moving to the focal optimum strictly lowers focal cost and strictly raises system cost. Interchanging the optima reverses the signs. □

COROLLARY EC.6 **(Objective and observation are distinct)**. *Restriction to a local observation does not imply optimization of focal cost. A locally measurable gate can be evaluated by $\mathbb{E}[gD]$, while a globally informed gate can optimize the wrong objective. Whether a particular local observation attains the richer-information optimum requires additional conditions.*

Proof. Measurability restricts the admissible decisions; the functional $\mathbb{E}[gD]$ specifies their objective. Neither definition determines the other. □

PROPOSITION EC.9 **(The magnitude of disagreement does not determine benefit)**. *Fix $d > 0$ and $w \in (0,1]$. There are feasible quadratic benchmarks with identical intended and recommended orders and $|a^A - a^H| = d$, but opposite signs of D. Moreover, with $w > 0$, benefit requires the adjustment to move toward the system optimum and have magnitude less than twice the initial distance to that optimum.*

Proof. Set $a^A = a^H + d$ and choose a^H large enough to keep all selected quantities nonnegative. Choose two feasible system optima on opposite sides of $a^H + wd/2$. Equation (EC.20) gives opposite benefit signs. For a nonzero adjustment, that equation is positive precisely when δ_s and $a_s^\star - a^H$ have the same sign and $w|\delta_s| < 2|a_s^\star - a^H|$. The signed gap is therefore relevant in the benchmark, but it is not itself sufficient for benefit. □

EC.1.24 Selection at a fixed opportunity

PROPOSITION EC.10 **(Selection premium)**. *Let D be integrable with a fixed distribution. The better constant action earns $\max\{0, \mathbb{E}[D]\}$ relative to always withholding. A gate observing D earns $\mathbb{E}[D^+]$. Its premium is*

$$P = \mathbb{E}[D^+] - \max\{0, \mathbb{E}[D]\} = \min\{\mathbb{E}[D^+], \mathbb{E}[D^-]\}. \tag{EC.21}$$

The premium is strictly positive exactly when D has positive probability of each strict sign.

Proof. Pointwise maximization gives $g = \mathbf{1}\{D > 0\}$. Set $A = \mathbb{E}[D^+]$ and $B = \mathbb{E}[D^-]$; the difference is $A - (A - B)^+ = \min(A, B)$. Strict positivity is equivalent to $A, B > 0$. □

PROPOSITION EC.11 **(Dispersion and selection)**. *If integrable D' is a mean-preserving spread of D, then its selection premium is weakly greater. Strict improvement requires $\mathbb{E}[(D')^+] > \mathbb{E}[D^+]$; sign variation alone does not establish that strict inequality.*

Proof. The constant-action term is unchanged because the means coincide. Convexity of x^+ and the mean-preserving-spread order imply $\mathbb{E}[(D')^+] \geq \mathbb{E}[D^+]$. Subtract the shared constant-action term. □ The proposition concerns the distribution of a recommendation's benefit at a fixed opportunity.

COROLLARY EC.7 **(The episode oracle is not a decision-level bound)**. *For a fixed realized world, the ex post minimum over all binary display profiles is no larger than the minimum over the two constant profiles. Thus an oracle selecting blanket advice or silence is not a lower bound on the cost of every decision-level policy.*

Proof. The two constant profiles form a subset of $\{0,1\}^{80}$. Minimization over the larger set cannot increase the optimum. Equality can hold; strict improvement is not guaranteed. Both ex post optimizations have access to hindsight, so this inclusion does not establish attainability by an ex ante policy. □

REMARK EC.3 (BUDGETS AND DISPLAY COSTS). In a static problem with a weak expected-display cap and known D, an optimal policy thresholds positive benefit at a nonnegative shadow price and may randomize at a tie. It need not exhaust the cap when too few opportunities have positive benefit. The computational evaluation imposes no binding cap or direct display charge. The opportunity cost in the dynamic model instead comes from the stipulated response transitions. Any dynamic threshold must incorporate continuation effects, rather than applying a one-shot dose rule to an endogenous sequence.

EC.2 Experimental Analysis and Calibration

EC.2.1 Experimental sample and response observations

The experimental sample comprises 69 consenting completers and 1,656 weekly decisions. Of 111 eligible completers, 42 lacked recorded research consent and are excluded from all research analyses and behavioral calibration. Table EC.1 reports the study's recruitment disposition and the response-eligible subsample. Assignment to no advice ($n = 22$), random advice ($n = 22$), and selective advice ($n = 25$) was blocked by earlier role and team. The selective condition used an earlier fixed policy, with 144 gate-triggered and 25 quota-triggered displays; its chain-cost contrast with random advice has $p = 0.67$. The computational policies evaluated in the main text differ from that intervention, so their comparisons do not estimate its experimental treatment effect.

Table EC.1 Experimental sample flow.

Step	n	Record
Initial sessions	120	
less off-roster	115	5 removed
less incomplete	112	3 removed
less duplicate	111	1 removed
less without recorded research consent	69	42 removed
Analyzed completers	69	24 decisions per participant
Decisions	1,656	69×24
Contributing an eligible response	42	At least one eligible display
Eligible responses	168	Displayed advice with $\lvert a^A - a^H \rvert \geq 2$

Counts distinguish recruitment, completed decisions, and observations eligible for response estimation; these are different units of analysis.

Response weights are ratios $(a^E - a^H)/(a^A - a^H)$ when the denominator has magnitude at least two. Using $W \geq 0.9$ for adoption and $W \leq 0.1$ for ignoring, the 168 eligible observations comprise 84 adoption, 70 ignoring, and 14 intermediate responses. Random advice contributes 82 observations (31 adoption, 38 ignoring, 13 intermediate), and selective advice contributes 86 (53, 32, and 1). The response composition therefore differs across display mechanisms. Pooling imposes a common conditional response model despite those differences; observations within a participant are not independent behavioral replications.

Table EC.2 Experimental observations used for behavioral calibration.

Condition	Participants	Decisions	Displays	Eligible responses (participants)
No advice	22	528	0	0 (0)
Random advice	22	528	109	82 (22)
Selective advice	25	600	169	86 (20)
Total	69	1,656	278	168 (42)

A response observation requires a displayed recommendation with $|a^A - a^H| \geq 2$. Participants make repeated decisions; the last column gives both the number of eligible decisions and the number of people contributing them.

Table EC.3 gives initial probabilities, transition probabilities, and within-state means. Rounding can make rows differ slightly from one. Transitions are indexed by eligible response observations. Of 126 consecutive eligible pairs, 15 skip at least one intervening display. Interpreting all pairs as one-display transitions approximates these eligibility gaps as a single transition. The resulting calibration uncertainty is distinct from Monte Carlo error: standard errors near 0.0035 for means based on 20,000 simulated paths quantify numerical averaging under fitted parameters, not uncertainty in the transition estimates.

Table EC.3 Fitted exposure-indexed response chain.

State	Initial	$\rightarrow$ adopt	$\rightarrow$ middle	$\rightarrow$ ignore	Mean w
adopt	0.6429	0.6875	0.0313	0.2813	$+1.0115$
middle	0.1429	0.1667	0.4167	0.4167	$+0.5231$
ignore	0.2143	0.2200	0.0200	0.7600	-0.0001

Stationary probabilities (adopt, middle, ignore): 0.4091, 0.0408, 0.5502. Transition estimates include eligibility gaps; see text.

Response paths and within-state innovations are drawn by position and exposure index, enabling paired policies to encounter the same jth response innovation at their jth display. This convention couples exposures, not calendar dates. Fold-specific fitting excludes the held-out participants. The maximum observed display count is sixteen, and the maximum number of eligible responses per participant is eleven. Predictions through twenty modeled displays extend beyond both observed sequences. The stationary intermediate-state probability of 0.0408 is not the observed intermediate-response proportion of $14/168$.

EC.2.2 Order limits and measured response

The experimental advisor was an adaptive base-stock rule. Its recommendations were nonnegative but had no upper bound, while the order-entry interface rejected quantities above 10,000. Seven displayed recommendations exceeded that limit. The largest, 72,244 units, occurred at a factory position with 69,394 units of backlog. A control for accepting the recommendation could populate the order field, but the entry limit still applied when the order was submitted. None of these seven recommendations was adopted in full.

Six of the 168 eligible response observations combine an intended order at the entry limit with a recommendation above it. For those observations, a zero adjustment is mechanically consistent with the interface constraint and cannot be interpreted solely as unwillingness to follow advice. Excluding them changes the

median of participant-level median response weights from 0.5455 to 0.5476 and reduces the contributor count from 42 to 41. This comparison concerns that summary statistic; it does not establish invariance of the fitted response transitions or policy costs. The baseline calibration retains the observed adjustments.

EC.2.3 Ordering and cost comparisons

The unit of analysis is a consenting participant. Outcomes over weeks 5–24 are maximum final order, maximum intended order, the sample standard deviation of final orders, and total chain cost. All 69 participants are retained in the role–schedule specifications, including the six whose intended orders reach 1,000 units or more. The outcome and response specifications were developed after data collection. They supplement the original selective-versus-random chain-cost comparison.

For each outcome Y_i, the logarithmic specification regresses $\log(1+Y_i)$ on advice-condition indicators, role and disruption-schedule indicators, and $\log(1+\text{peak final order})$ and $\log(1+\text{final-order s.d.})$ from weeks 1–4. No recommendation is displayed during those first four weeks. The rank specification uses average ranks of the outcome and its corresponding pre-advice measure, together with condition, role, and schedule. Peak intended orders use pre-advice intended orders as their baseline. We use HC3 covariance estimates (MacKinnon and White 1985) and residual-degrees-of-freedom Student t references. Rank regressions estimate differences in adjusted outcome ranks; they are not exact randomization tests.

An assignment-block specification replaces role and schedule indicators with the original team–role block effects. Fourteen participants are the sole consenting observation in their block and contribute no within-block treatment comparison; these specifications use 55 participants in 23 blocks. Ranks are computed within each estimation sample. The four outcomes, two advice-versus-no-advice contrasts, and four specifications form a family of 32 tests. Holm's sequential procedure adjusts the two-sided p-values across this family (Holm 1979). Pointwise confidence intervals describe individual estimates and are not simultaneous intervals. The family includes both scales and the assignment-block specifications so that a claim is not selected by comparing their unadjusted significance levels.

The role–schedule rank comparison for random advice and peak final orders has adjusted $p = 0.033$. Its assignment-block counterpart has adjusted $p = 0.256$; the corresponding selective-advice comparisons have $p = 0.719$ and $p = 0.798$. No other comparison in this family meets the adjusted five-percent threshold. Expanding the family to include HC1 covariance with normal critical values and specifications with untransformed baseline controls gives 80 tests; the same HC3 peak-order rank comparison then has adjusted $p = 0.080$. This broader family changes the inferential conclusion without changing the estimate. Intended-order changes can reflect operating states altered by earlier advice as well as changes in the planner's decision process; the experiment does not separately identify those channels.

Table EC.4 Ordering and cost after advice becomes available.

Outcome	Advice	Log estimate [95% CI]	Rank estimate [95% CI]	Holm p log / rank
A. Role and schedule controls; 69 participants				
Peak final order	Random	0.748 [0.001, 1.494]	15.77 [6.63, 24.91]	1.000 / 0.033
	Selective	0.814 [−0.087, 1.715]	11.61 [1.39, 21.82]	1.000 / 0.719
Peak intended order	Random	0.476 [−0.259, 1.210]	10.58 [0.98, 20.19]	1.000 / 0.783
	Selective	0.731 [−0.146, 1.607]	9.54 [−1.28, 20.36]	1.000 / 1.000
Final-order s.d.	Random	0.557 [−0.157, 1.272]	12.40 [3.64, 21.15]	1.000 / 0.195
	Selective	0.649 [−0.207, 1.506]	6.72 [−2.75, 16.19]	1.000 / 1.000
Chain cost	Random	0.539 [−0.324, 1.402]	13.28 [0.90, 25.66]	1.000 / 0.828
	Selective	0.510 [−0.491, 1.512]	8.00 [−5.05, 21.06]	1.000 / 1.000
B. Assignment-block controls; 55 participants				
Peak final order	Random	1.016 [−0.126, 2.157]	13.00 [3.57, 22.42]	1.000 / 0.256
	Selective	0.890 [−0.884, 2.665]	11.12 [0.95, 21.28]	1.000 / 0.798
Peak intended order	Random	0.838 [−0.285, 1.961]	9.94 [1.21, 18.68]	1.000 / 0.719
	Selective	0.816 [−0.931, 2.563]	8.01 [−0.94, 16.96]	1.000 / 1.000
Final-order s.d.	Random	0.808 [−0.343, 1.958]	10.73 [1.48, 19.97]	1.000 / 0.686
	Selective	0.755 [−1.015, 2.524]	7.31 [−1.85, 16.47]	1.000 / 1.000
Chain cost	Random	0.939 [−0.469, 2.346]	17.57 [4.57, 30.57]	1.000 / 0.285
	Selective	0.631 [−1.382, 2.644]	14.13 [0.13, 28.13]	1.000 / 1.000

All effects compare the named advice condition with no advice over weeks 5–24. Confidence intervals are pointwise 95%, using HC3 covariance estimates and residual-degrees-of-freedom Student t critical values. Holm-adjusted p-values cover all 32 tests in both panels, across outcomes, conditions, and scales. Panel A includes 22 no-advice, 22 random-advice, and 25 selective-advice participants. Panel B excludes the 14 singleton assignment blocks. Log estimates concern $\log(1+Y)$; rank estimates concern average ranks within the estimation sample. Neither is a proportional effect on arithmetic mean orders.

EC.2.4 Response regressions

The response sample contains 168 displayed recommendations with $|a^A - a^H| \geq 2$ from 42 consenting participants. Adoption is $A_{it} = \mathbf{1}\{W_{it} \geq 0.9\}$. Two alternative definitions are exact agreement, $E_{it} = \mathbf{1}\{a^E_{it} = a^A_{it}\}$, and $\mathbf{1}\{W_{it} \geq 1\}$. The latter includes overshooting and is distinct from exact agreement. No outcome from a nonconsenting participant enters these analyses.

Within each advice condition, the exposure model is

$$Y_{it} = \eta_i + \beta N_{it} + \gamma t + \xi |a^A_{it} - a^H_{it}| + u_{it}, \tag{EC.22}$$

where Y is a response indicator and N_{it} counts every display strictly before week t. Participant effects absorb stable response differences. A sensitivity uses $\log(1 + |a^A - a^H|)$ in place of the absolute gap. The between-condition exposure-slope comparison includes participant effects and condition interactions with prior displays, week, and absolute gap. The adjusted condition comparison includes a selective-advice indicator, absolute gap, upward-recommendation indicator, week, and the current disruption indicator.

The severity model uses 81 eligible disrupted decisions from 35 participants, codes mild, moderate, and severe as 1, 2, and 3, and includes condition, absolute gap, direction, and week. Its coefficient imposes a

common increment between adjacent bands. The within-participant specification replaces condition with participant effects while retaining gap, direction, and week; 18 participants contribute within-person severity variation. Table EC.5 reports all 17 exposure, condition, and severity comparisons across the response definitions and sensitivity specifications. They form one response family, distinct from the participant-level operating outcomes, and use Holm adjustment across all 17 tests.

Table EC.5 Responses to displayed advice.

Comparison	Coefficient [95% CI]	p	Holm p	Decisions/ participants
Adoption: $W \geq 0.9$				
Prior displays, random	-8.6 $[-19.5, 2.2]$	0.113	0.604	82/22
Prior displays, selective	11.1 $[1.9, 20.3]$	0.021	0.269	86/20
Prior displays, random; log gap	-6.3 $[-18.1, 5.4]$	0.277	0.831	82/22
Prior displays, selective; log gap	10.8 $[2.6, 19.1]$	0.012	0.179	86/20
Selective minus random	23.4 $[-4.7, 51.5]$	0.101	0.604	168/42
Exposure slope: selective minus random	19.7 $[6.1, 33.4]$	0.006	0.095	168/42
Severity, pooled	10.2 $[-2.3, 22.7]$	0.105	0.604	81/35
Severity, participant effects	2.9 $[-16.4, 22.2]$	0.761	1.000	81/35
Exact agreement				
Prior displays, random	-12.9 $[-22.6, -3.2]$	0.012	0.179	82/22
Prior displays, selective	12.3 $[3.2, 21.5]$	0.011	0.174	86/20
Selective minus random	27.0 $[-0.4, 54.5]$	0.053	0.425	168/42
Severity, pooled	13.3 $[1.4, 25.2]$	0.030	0.304	81/35
Severity, participant effects	4.8 $[-13.4, 23.0]$	0.597	1.000	81/35
Response at least as large as full adoption: $W \geq 1$				
Prior displays, random	-9.9 $[-19.3, -0.5]$	0.040	0.358	82/22
Prior displays, selective	11.1 $[1.9, 20.3]$	0.021	0.269	86/20
Selective minus random	26.0 $[-2.4, 54.3]$	0.072	0.504	168/42
Severity, pooled	13.4 $[1.6, 25.3]$	0.028	0.304	81/35

Coefficients and intervals are in percentage points. Pointwise 95% confidence intervals use participant-cluster CR1 covariance, with correction $G/(G-1) \times (N-1)/(N-K)$ and Student t_{G-1} critical values. G counts participants and K includes the intercept and participant effects when present. Holm adjustment covers all 17 tests shown. Each exposure coefficient is per earlier display, each severity coefficient is per severity band, and the slope contrast is per earlier display. The estimates condition on an eligible displayed recommendation.

None of the response tests rejects at the adjusted five-percent level. For the primary adoption definition, the random-condition exposure coefficient is -0.0863 (cluster standard error 0.0522; adjusted $p = 0.604$). Exact agreement gives -0.1290 (0.0469; adjusted $p = 0.179$). The difference between condition-specific exposure slopes has adjusted $p = 0.095$. Within-participant severity estimates are 0.0291 for adoption and 0.0479 for exact agreement, with adjusted $p = 1.000$ in both cases.

Random assignment governs the advice condition, but these regressions condition on an eligible display. Earlier advice can change intended orders, gaps, and subsequent states; selective displays are chosen by

the gate or quota. The coefficients describe conditional response associations and do not identify treatment effects on total cost or a structural retention parameter. The adjustment addresses the stated testing family; it does not convert these conditional comparisons into randomized exposure or severity interventions.

EC.2.5 Ordering model and predictive checks

The ordering model uses participant-specific anchoring-and-adjustment parameters and a shortage-augmented specification, following Sterman (1989), Sterman and Dogan (2015). Table EC.6 evaluates prediction by fitting twelve weeks and scoring the other twelve, repeating in the opposite direction, and averaging errors within person. The fitted model has higher held-out error than all three simple benchmarks. The median within-person ratio of held-out to training error is 2.28 under this protocol. This ratio describes the split-half comparison, rather than bounding the estimation error of a fit using all 24 weeks.

Table EC.6 Ordering-model predictive errors.

Predictor	In sample	Held out	Fitted rule better in sample
Fitted five-parameter rule	3.22	6.43	
Own mean order	4.37	5.01	87%
Pass-through (order = incoming order)	4.07	4.07	91%
Constant 4 (= customer demand)	4.58	4.58	97%

Median participant RMSE, 69 participants. Held-out errors use bidirectional split halves; the last column compares in-sample participant errors.

Across experimental arms, Holm-adjusted tests of the fitted parameters and scale-free fit statistics do not reject equality (adjusted p values 0.84–1.00). Median scale-free fit quality is 0.907, 0.816, and 0.848. Raw RMSE medians are 2.28, 3.94, and 3.63; the unadjusted Kruskal–Wallis p value is 0.020 and its family-adjusted value is 0.16. A bootstrap interval for the no-advice-minus-advice difference is $[-3.60, -0.31]$. Differences in state distributions and order scale can contribute to these patterns, but the tests do not establish conditional invariance of the ordering rule. Unadjusted tests of volatility, maximum order, and order standard deviation yield $p = 0.017$, 0.029, and 0.044; none survives the reported behavioral-family adjustment.

Role-aware sampling preserves the empirical marginal distribution within each position. Independent draws do not preserve team dependence, and the behavioral experiment observes one human planner with simulated partners in each chain. Common random numbers improve paired numerical comparisons without guaranteeing that behavioral misspecification cancels.

EC.2.6 Advisor specification and elicitation

The LLM-fitted advisor uses

$$a^A = \max\{0, \min\{U_i, A\, m_{i,r}(b)\}\}, \quad b = \frac{B_i}{\max\{\widehat{D}_i, 4\}}, \quad A = \begin{cases} \widehat{D}_i, & \widehat{D}_i > 0, \\ 4, & \widehat{D}_i = 0. \end{cases} \tag{EC.23}$$

Here U_i is the fitted upper recommendation cap, $m_{i,r}$ is a monotone piecewise-linear markup table for position i and regime r, and B_i is backlog. The table is flat beyond its last knot. The demand floor matters because received demand is zero in 5.6%, 23.5%, and 26.7% of weighted wholesaler, distributor, and factory states. Inventory and supply line are absent from this surrogate. A second-stage correction changes held-out error by less than 0.001 at three positions and increases factory weighted median absolute error from 4.87 to 5.52. Those comparisons support this fitted representation within its evaluated support, not the absence of value in inventory information for another advisor.

Six announced regimes represent cuts of 10, 25, 33, 50, 75, and 95%, with bucket boundaries 0.175, 0.2917, 0.4167, 0.625, and 0.85. Calm weeks use the unannounced regime. Fitted zero-backlog markups are approximately 1.000, 1.149, 1.086, and 1.237 by position; these values describe the fitted baseline and do not characterize every announced regime. The applicable regime is determined by the public disruption announcement.

Table EC.7 LLM elicitation corpus.

Pass	Variation	Calls	Use
Varied severity	varied-severity states, both announcement arms, full wording	7,316	severity fit
Constant world	constant-world states, both arms, full wording	4,466	supplies the base markup table
Terse wording	both arms, *terse* capacity wording	2,932	capacity-wording sensitivity
Replicate panel	659 states re-asked under an identical prompt	2,036	measures the noise law
Reasoning, terse	announced only, terse, reasoning variant	1,466	stated-rationale classification
Wording sensitivity	minimal and reasoning-framed variants	892	wording and rationale ablation
Initial query check	initial advisor queries	24	excluded from fitting
Total		19,132	

Prompt totals: role-framed 16,774; reasoning-framed 1,912; minimal 446. Calls include replication and ablations.

The query corpus contains only simulated numerical states and prompt variants. It contains 19,132 calls, including calibration, replication, and wording checks (Table EC.7); these are not 19,132 independent states used in the final fit. A temperature of zero did not produce deterministic responses. Changed replicate pairs were modeled with a multiplicative mixture: with fitted probability p_{change}, a Student-t relative perturbation is applied. The pooled degrees of freedom are 1.5497, estimated on 569 changed pairs; depth-specific scales are 0.2266, 0.2697, 0.3015, 0.2727, 0.3462, and 0.4061. Before truncation this distribution has finite mean and infinite variance. The clipping sequence, negative-draw treatment, and per-position caps therefore matter for simulated tails.

Table EC.8 Sensitivity to prompt wording.

Capacity wording in the prompt	Above demand	Median ratio	Per-position median (R/W/D/F)	Chain product
No capacity information	41.0%	1.000	1.000 / 1.000 / 1.000 / 1.130	1.130
Capacity numbers only (terse)	46.7%	1.000	1.000 / 1.000 / 1.002 / 1.134	1.137
Numbers + "Deliveries down the chain may be short for a while"	66.5%	1.125	1.000 / 1.175 / 1.179 / 1.251	1.734

These elicitation comparisons concern the specified prompts and state distribution.

The role-framed wording mentions possible delayed or partial downstream deliveries. Removing or changing that sentence materially changes the fitted responses (Table EC.8). These prompt conditions delimit the advisor being studied. Rationale classifications are descriptions of generated text, not evidence of the model's causal reasoning. Pair-weighted change rates, averages across states, and averages across regimes use different weighting schemes and need not coincide.

The surrogate differs from the queried recommendations in three respects. The surrogate overstates the frequency of above-demand recommendations in several deep-cut buckets (differences across buckets range from -0.5 to $+17.9$ percentage points), understates the weighted median order-to-demand ratio (1.013 versus 1.065), and cannot represent the decline in recommendations at very high backlog because its markup is monotone and eventually flat. The net effect on policy comparisons is not identified by these discrepancies.

Table EC.9 Information available to the learned gates.

Gate	Varying inputs	Operating information
Chain-wide	64	Inventories, backlogs, pipelines and related states at all four positions; current capacity and disruption history; chain cost and backlog-growth summaries.
Local	26 / 28 / 28 / 34	Recipient inventory, backlog, pipeline and incoming order; disagreement with advice; own response history; time and public disruption indicator.

Both gates use a 70-slot representation. The local mask fixes other-position states and chain aggregates at constants. Its number of varying inputs depends on the recipient's position. Masked capacity entries are unobserved quantities, not observations of zero available capacity.

The chain-wide observation includes current capacity and summaries of disruption history, including recent incidence and run length. The local observation retains the public disruption indicator and recipient-specific operating information. Human participants observed the stated capacity reduction during each disrupted week, but received neither a computed disruption-history summary nor an announced recovery date. The learned-policy comparison therefore varies more than access to other positions' inventories. Even without a direct capacity signal, a gate can respond to capacity losses through the orders, backlogs and shipments they induce.

EC.2.7 Training and policy controls

PPO uses eight parallel environments, 2,048 rollout steps, 500,000 training timesteps, $\gamma = 1$, and $\lambda_{\mathrm{GAE}} = 0.99$. The quantity $1/(1-\gamma\lambda) = 100$ is the geometric decay scale of the advantage weights. Each learned family uses five participant folds and three candidate seeds per fold.

The period reward is the reduction in chain cost relative to a matched silence episode with the same planners, disruption path, and exogenous draws, divided by that episode's cost over the advice window. Reward is assigned once per week after all four positions have decided, with no intervention penalty or additional scaling. The smallest reference advice-window cost is 428.51, so no denominator approaches zero. Evaluation measures realized chain cost directly. The reward normalization affects training by placing greater weight on savings in worlds with lower silent cost; it does not change the cost measure used for evaluation.

For each fold and learned-policy family, seed selection minimizes the paired median cost difference from blanket advice using training participants and no binding advice budget. Held-out participants do not enter this selection. Selection among a finite set of seeds does not establish that the chosen policy is optimal within its information class.

Training concatenates two independent 24-week draws rather than extending one shock schedule to 48 weeks. Disrupted-week densities are 33.7% in training and 33.3% in evaluation, with disjoint schedule-index ranges. Similar disruption densities do not imply identical operational-state distributions.

The training samples overlap across folds, although their held-out participants differ. Fold comparisons therefore reflect variation across related fits, not five independent repetitions of the full estimation and training procedure.

The common architecture holds the nominal input dimension fixed, while masking changes the effective representation and optimization problem.

The signed-gap threshold 4.74 was extracted from a depth-two tree approximating a policy trained with an adaptive base-stock advisor in a fixed-disruption environment and transformed out of symlog units. The two-condition rule reproduced 74.3% of that policy's choices on 20,000 held-out decisions, compared with 61.2% for the majority class; the one-condition rule reproduced 65.6%. A depth-three tree achieved accuracy 0.769 and AUC 0.821; a logistic surrogate achieved 0.819 and 0.899. These fidelity measures concern the earlier policy from which the threshold was extracted; they do not guarantee the operating performance of the evaluated gap filter or hand rule.

The evaluations impose neither a binding display cap nor an explicit attention penalty. The two matched-dose controls randomize timing at the display count induced by their respective learned gates. This retrospective construction conditions on a policy-dependent quantity, so it does not estimate the value of an ex ante fixed attention budget.

EC.2.8 Paired worlds and disruption schedules

All eight policies are evaluated on the same 1,500 worlds. Silence and blanket advice provide common reference outcomes for the chain-wide and local comparisons. The pooled median cost under silence is 1,210.78. Common world identifiers pair planner draws and exogenous shocks; the resulting trajectories remain policy-dependent.

Disruption counts within the advice window define the duration strata. Fixed-bin counts are 41, 329, 742, 346, and 42 for zero, 1–4, 5–8, 9–12, and at least 13 disrupted advice-window weeks. Quartile groups contain 486, 413, 310, and 291 worlds for 0–5, 6–7, 8–9, and 10–18 weeks. Ties account for unequal quartile sizes. The 41 worlds in the zero-count category have shocks wholly within the four warm-up weeks. Counts therefore describe exposure while advice is available, rather than shock exposure over the entire episode. These strata combine different outage paths and are not randomized duration treatments.

EC.3 Policy Comparisons

Each comparison pairs costs within the same world. Negative differences favor the first policy. The mean is an expected-cost estimate in the fitted environment; the pooled median describes the central within-world contrast. Fold medians and their average are separate summaries. We report both because averaging medians and pooling observations need not agree.

EC.3.1 Variation across participant folds

Table EC.10 gives the five fold medians for the principal contrasts. The final column is the pooled paired median, not their average. Table EC.11 reports the average of the fold mean differences and t_4 reference intervals based on their variation. As described in Section 4.7, overlapping training samples leave the repeated-sampling coverage of these intervals undetermined. No multiplicity-adjusted testing claim is attached to them.

Table EC.10 Within-fold medians of paired cost differences.

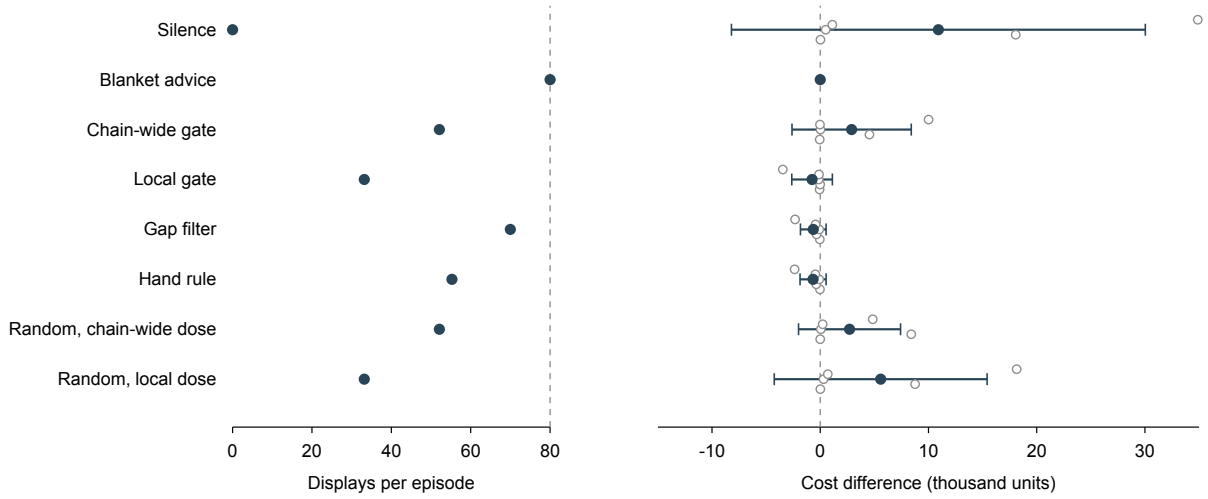
Comparison	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Pooled
Local − chain-wide	+5.13	-55.10	-75.24	-132.92	-265.62	-45.00
Local − blanket	-0.53	+22.66	-89.88	-68.98	-19.85	-27.51
Chain-wide − blanket	-3.73	+68.44	+4.44	+17.89	+82.74	+14.60
Gap filter − blanket	-2.13	-9.47	-24.21	-59.08	-28.64	-13.22
Hand rule − blanket	+0.09	-12.46	-21.38	-134.52	-37.38	-19.52
Local − matched random	-15.91	-216.92	-192.13	-540.75	-1,991.28	-199.27
Chain-wide − matched random	-11.58	-47.33	-11.45	-33.27	+14.94	-15.05
Local − chain-wide, Q4	+14.36	-192.17	-90.18	-57.68	-123.53	-39.99

All-world contrasts use 300 worlds per fold. Q4 contains 291 worlds with 10–18 disrupted advice-window weeks. Matched random denotes the control at the corresponding gate's realized dose. Folds contain distinct held-out participants but share training participants.

Table EC.11 Paired mean differences and variation across fitted populations.

Comparison	Mean fold difference	t_4 reference interval	t_{mean}
Local − chain-wide	−3,654.74	[−10,874.88, +3,565.39]	−1.41
Local − blanket	−754.67	[−2,630.35, +1,121.01]	−1.12
Chain-wide − blanket	+2,900.07	[−2,610.28, +8,410.43]	+1.46
Gap filter − blanket	−645.05	[−1,829.18, +539.08]	−1.51
Hand rule − blanket	−662.97	[−1,866.40, +540.47]	−1.53
Local − matched random	−6,340.12	[−17,848.78, +5,168.54]	−1.53
Chain-wide − matched random	+192.21	[−3,804.45, +4,188.88]	+0.13
Local − chain-wide, Q4	−2,961.22	[−8,803.69, +2,881.24]	−1.41

Point estimates equally weight the five fold means; intervals add and subtract $2.776\,s/\sqrt{5}$. Equal fold sizes make the all-world estimates identical to pooled paired means. Q4 has unequal fold sizes; Table EC.12 reports its pooled mean. Reference intervals are not calibrated confidence intervals for re-estimation and retraining.



(a) Mean displays per episode (b) Paired mean cost relative to blanket advice

Figure EC.1 Display frequency and variation in cost across fits. Panel (a) gives mean displays out of 80 opportunities. Panel (b) shows policy-minus-blanket paired means; open points are the five fold means, filled points their average, and lines are t_4 reference intervals. Costs are in thousands of units. Each fold contains 300 worlds. Training samples overlap, so the intervals describe variation across fitted populations rather than calibrated coverage under repeated estimation and training.

EC.3.2 Recorded disruption duration

Table EC.12 conditions on the number of disrupted weeks during which advice is available. Local-minus-chain-wide mean differences are negative in every displayed stratum. Pooled medians are also negative, but they are small in the zero-week and longest fixed-bin categories. The grouped comparisons do not form a monotone duration gradient. Each stratum changes the mix of outage timing, depth and planner draws, whereas Proposition 3 varies persistence holding the benefit distribution fixed.

Table EC.12 Local minus chain-wide gate by disrupted advice-window weeks.

Disrupted weeks	Worlds	Paired mean	Paired median	t_{med}
0	41	−3,846.72	−6.81	+0.02
1–4	329	−4,633.76	−43.78	−2.65
5–8	742	−3,475.38	−45.73	−2.10
9–12	346	−3,236.70	−68.72	−2.55
13–18	42	−2,410.98	−4.92	−1.39
10–18 (Q4)	291	−2,952.51	−39.99	−2.62

The first five rows partition the evaluation worlds. The final row is the upper quartile under tied duration counts and overlaps the 9–12 and 13–18 bins. The zero-week group experiences disruption only during warm-up.

EC.3.3 The distribution of total operating cost

Table EC.13 gives marginal cost quantiles and the average cost in each policy's upper five percent. These are different from quantiles of paired policy differences. The upper-tail sets are defined separately for each policy, so changes in their averages combine changes in membership and in cost. The five-percent cost shares range from 41.7% to 45.9%, indicating that a small fraction of worlds contributes substantially to every policy's mean cost.

Table EC.13 Total-cost tails under each delivery policy.

Policy	95th percentile	Upper-tail mean	Upper-tail cost share
Silence	242,099.42	300,427.60	41.7%
Blanket advice	175,515.38	226,562.90	45.1%
Chain-wide gate	184,939.07	243,443.00	43.4%
Local gate	169,885.64	223,029.18	45.7%
Gap filter	173,727.04	224,928.63	45.9%
Hand rule	173,724.76	224,893.88	45.9%
Random, chain-wide dose	191,398.87	239,681.44	43.0%
Random, local dose	208,235.26	264,841.33	43.1%

Each marginal distribution contains 1,500 episode costs. The upper-tail mean averages the 75 highest-cost worlds, and its cost share divides their total by aggregate cost under that policy. The 95th percentile uses linear interpolation between adjacent ordered costs.

Blanket advice is more costly than silence in 328 of the 1,500 worlds (21.9%). This is an outcome-defined subset, not an ex ante state a delivery policy observes. Within it, the local-minus-chain-wide paired median is −24.92, with fold-median statistic −1.71. The local-minus-blanket paired median is −159.79, with statistic −2.30. These selected comparisons describe complete policy trajectories and do not identify the causal effect of any particular display.

A retrospective choice between blanket advice and silence has cost $\min\{C_{\text{blanket}}, C_{\text{silence}}\}$ in each world. Relative to blanket advice, its mean difference is −48.48 and its paired median is zero; the policies coincide in the 78.1% of worlds where blanket advice is no more costly. This diagnostic is restricted to two constant policies. It is not an upper bound on the value attainable by within-episode gating.

EC.3.4 Agreement with experimental no-advice behavior

The no-advice condition supplies the behavioral reference for simulated silence. Table EC.14 compares an endpoint order-dispersion ratio, retailer backlog incidence and the factory's median order. The simulated factory order level exceeds the experimental interval. The other two intervals are wide, reflecting the five to seven no-advice participants per position; their inclusion of the simulated value does not establish accurate model fit.

Table EC.14 Observed and simulated silence moments.

Moment	Observed	95% interval	Simulated	Outside interval
Endpoint order-dispersion ratio	3.851	[0.938, 51.022]	7.791	No
Retailer backlog incidence	0/5	[0.000, 0.434]	0.238	No
Factory median order	4.00	[0.00, 5.00]	12.922	Yes

Intervals are bootstrap for continuous moments and Wilson for incidence. Five to seven no-advice participants contribute per position.