SIGNED RESIDUALS FOR SETPOINT BAYESIAN OPTIMISATION: WHY AN ACTIVE-INFERENCE ACQUISITION COLLAPSES TO GREEDY

Anonymous authors
Paper under double-blind review

ABSTRACT

Setpoint problems, such as holding a physiological reading at a clinical target while the response drifts, are usually given to Bayesian optimisation by folding the residual into a distance to be maximised. We show that this encoding, rather than the acquisition function, decides whether a preference-based acquisition can work. Our case study is BOBA (Kelly et al., 2025), an active-inference acquisition function for dynamic BO. We prove that its pragmatic term orders candidates by the posterior mean whenever the preferred outcome lies above every prediction, which holds on 94.6% of recorded states, and that its epistemic term is too small to change the decision, so the method acts greedily. Folding guarantees this condition for every target, because it places the preferred outcome at the supremum of the objective. Keeping the residual signed restores the preference term's ordering, raising the share of states whose candidates bracket the target from 5.0% to 57.7% on a clinical digital twin and from 5.6% to 84.5% on fourteen synthetic setpoint problems. Once the encoding is fixed, the epistemic term no longer helps: in none of eight synthetic cells is it significantly beneficial, and a two-line proximity rule (GREEDY-H) matches or beats every active-inference variant we test. On a digital twin of closed-loop anaesthesia built from 50 patient-specific models fitted to surgical recordings, the proximity rule beats a fixed clinical protocol at every budget, although a forgetting-kernel baseline overtakes it at longer budgets. Where the target is unreachable, arms with an information term degrade $1.9\times$ in location error against $1.1\times$ for greedy and proximity rules. Redirecting BOBA's epistemic term at the optimum (BOBA-OPT) improves on BOBA but not on GP-UCB.

1 INTRODUCTION

Bayesian optimisation searches for the best setting of an expensive black-box function using a probabilistic surrogate, usually a Gaussian process, and an acquisition function that decides where to sample next (Jones et al., 1998; Mockus, 1994). The standard setting assumes the function does not change, while many real problems violate this. A control system tuned to a machine that heats up, an interface adapting to a user over time, and drug infusion affecting a patient whose physiology drifts during surgery are all problems in which the optimum setting changes over time. This is dynamic Bayesian optimisation, and it has been studied through surrogates that forget stale data (Bogunovic et al., 2016; Bardou et al., 2024; Bardou & Thiran, 2025) or that model time explicitly (Nyikosa et al., 2018; Chen & Li, 2021; Li et al., 2022), but far less through acquisition functions designed for a moving target (Kelly et al., 2025).

Many of these applications share a further property: the desired outcome is known in advance, even though the function is not. A clinician knows the depth of anaesthesia they want. A biofeedback system knows the physiological state required. Standard acquisition functions are built to find the maximum of an unknown function, so they can use a known target only indirectly, by rewriting the objective as the distance from it. Level-set estimation addresses a related question in the static setting (Bryan et al.; Mason et al., 2021).

Active inference offers a more direct route (Friston et al., 2023; Da Costa et al., 2023; Parr et al., 2020). An agent minimises Expected Free Energy, which combines a pragmatic term rewarding

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

movement toward preferred outcomes with an epistemic term rewarding information gain (Millidge et al., 2021). Kelly et al. (2025) instantiate this over a Gaussian process to obtain BOBA, an acquisition function for dynamic problems that in principle balances reaching the desired state against learning where it is.

This paper. We show that BOBA, as instantiated, reduces to greedy exploitation, and that the cause lies in how setpoint objectives are encoded rather than in the acquisition function.

BOBA's pragmatic term measures divergence from a preferred outcome that, on the problems it was evaluated on, sits at or above every predicted value. Divergence then shrinks monotonically as the predicted value rises, so ranking candidates by it is the same as ranking them by the posterior mean. Combined with an epistemic term several orders of magnitude too small at the published setting, BOBA acts greedily. Aiming the epistemic term at the latent an optimiser actually needs, the optimum's location and value, gives BOBA-OPT, which connects active inference to entropy search (Hvarfner et al., 2022).

The deeper cause is the encoding. A setpoint problem presented to a maximiser must be folded into a distance, $f = -|g - y^\star|$, and this places the preferred outcome at the supremum of f for every choice of $y^\star$, so the collapse is guaranteed by the algebra rather than observed by chance. Writing the residual signed instead, $h = g - y^\star$, puts the target strictly inside the objective's range and gives the preference term a non-monotone ranking, penalising overshoot exactly as it penalises undershoot. The same active-inference construction on the signed objective, with the epistemic term aimed at where the level is crossed, gives BOBA-LS. On a clinical twin and on fourteen synthetic problems (eight standard benchmarks and six one-dimensional restrictions of them), the signed encoding removes the collapse, and once it is made the epistemic term stops contributing.

Verifying a dynamic acquisition function also needs a benchmark with a moving optimum and the structure of a real application: a target fixed by domain knowledge, individuals who respond differently, and a non-adaptive protocol that an adaptive method must beat before it is worth deploying. We therefore build one from closed-loop anaesthesia, where an anaesthetist adjusts a drug concentration to hold a monitor reading at a clinically agreed value while surgery pushes that reading around. We fit patient-specific models to surgical recordings and use them as simulated patients. Our contributions are:

1. **An encoding analysis.** We show that folding a setpoint into a distance places the preferred outcome at the objective's supremum for every target. This guarantees that a Gaussian preference term collapses onto the posterior mean, which we prove and measure for BOBA. Keeping the residual signed restores the ordering on a clinical twin and on fourteen synthetic problems. Once it is restored, the epistemic term stops contributing and a proximity rule, GREEDY-H, which queries the candidate predicted closest to target, suffices.
2. **A clinical digital twin** for acquisition functions against moving optima. It has 50 patient-specific models fitted to open surgical recordings, an exact time-varying optimum, fitted noise, drift and dropout, and the deployable fixed protocol as the baseline. On it, information-seeking is harmful where the target is unreachable, and no method is safe for every patient.

Secondary. Two corrected acquisition functions: BOBA-OPT, with the epistemic term aimed at the optimum (§3.4), and BOBA-LS, aimed at the level crossing (§4). Neither beats the simplest competitor on its own suite: GP-UCB on synthetic maximisation, GREEDY-H on the twin.

2 BACKGROUND

2.1 DYNAMIC BAYESIAN OPTIMISATION

We optimise $f(x, t)$, a function of a controllable input x and time t. At each step the optimiser chooses one x, observes a noisy value, and time advances. It cannot go back and it cannot look ahead. The surrogate is a Gaussian process whose covariance function is a product over space and time,

$$k\big((x,t),(x',t')\big) = \sigma_f^2 \; k_{5/2}(x, x'; \ell_S) \; k_{3/2}(t, t'; \ell_T), \tag{1}$$

where σ_f^2 is the signal variance, $k_{5/2}$ is a Matérn kernel of smoothness $5/2$ over the input space with a single lengthscale ℓ_S shared by all input dimensions, and $k_{3/2}$ is a rougher Matérn kernel over time

with lengthscale ℓ_T. Small ℓ_T means a fast-moving objective and old observations rapidly stop being informative. Both lengthscales are constrained to $[0.005, 20]$, the signal variance is bounded below by 10^{-4}, and the observation noise is fixed from the data rather than fitted. Methods that search a discrete set score the same 128 scrambled Sobol points, fixed for the whole run, the *candidate set*, and are configured to use this surrogate with identical hyperparameters, so differences between them are differences between acquisition functions. The exceptions, and which arms have their surrogate verified from session records, are given in §6 and Appendix B.

At step t, the surrogate provides a predictive mean $\mu_t(x)$, its best guess at the objective, and a predictive standard deviation $\sigma_t(x)$, its uncertainty about that guess. We write σ_n for the standard deviation of the observation noise.

2.2 The acquisition function under study

Active inference (Friston et al., 2023) describes behaviour as the minimisation of Expected Free Energy G, which has two standard decompositions, equivalent under the usual approximations (Costa et al., 2020; Millidge et al., 2021). In the first, G is the *risk* $D_{\mathrm{KL}}[Q(o) \,\|\, P(\tilde{o})]$, the divergence between the agent's predicted observations $Q(o)$ and a distribution $P(\tilde{o})$ over preferred ones, plus an *ambiguity* term $\mathbb{E}_{Q(s)} H[P(o \mid s)]$ over hidden states s. In the second, G is an *extrinsic* term $-\mathbb{E}_{Q(o)}[\log P(\tilde{o})]$ minus the *epistemic* term $\mathbb{E}_{Q(o)} D_{\mathrm{KL}}[Q(s \mid o) \,\|\, Q(s)]$, the information an observation is expected to give about s. Minimising G means learning about the world while moving toward preferred outcomes.

Kelly et al. (2025) identify candidate inputs with actions and the surrogate's predictive distribution with $Q(o)$. The pragmatic term is the risk: the divergence between what the model predicts at x and a Gaussian of width σ_p centred on the preferred outcome $Y^\star$, small when the model confidently predicts a value close to $Y^\star$,

$$R(x) = D_{\mathrm{KL}}\big[\mathcal{N}(\mu(x), \sigma^2(x)) \,\big\|\, \mathcal{N}(Y^\star, \sigma_p^2)\big] = \log \frac{\sigma_p}{\sigma(x)} + \frac{\sigma^2(x) + (\mu(x) - Y^\star)^2}{2\sigma_p^2} - \frac{1}{2}, \quad (2)$$

and BOBA sets the preference width to the noise level, $\sigma_p = \sigma_n$. The epistemic term is the mutual information between the observation that would be made at x and the noise-free value of the objective at x, and is large where the model is uncertain relative to the noise:

$$I(x) = \frac{1}{2} \log\left(1 + \frac{\sigma^2(x)}{\sigma_n^2}\right). \quad (3)$$

The acquisition function is $a(x) = \beta\, I(x) - R(x)$, maximised over the candidate set, with $\beta > 0$ trading exploration against goal-seeking in the same role as the confidence multiplier in GP-UCB (Srinivas et al., 2012). Pairing the risk term from one decomposition with the information term from the other makes $a(x)$ an acquisition function inspired by Expected Free Energy rather than $-G$ itself; our analysis concerns $a(x)$ as implemented. Two weighting conventions appear in this literature, and we label each BOBA arm by the one it used: BOBA[pub], the published setting, with constant $\beta = 1$ and both terms unnormalised, and BOBA[wm] (weight-matched), which equalises the two terms' ranges before combining them and so raises the epistemic term's effective weight by roughly four orders of magnitude. The diagnostics of §3 pool the recorded states of all BOBA arms; Appendix A breaks each figure down by convention.

3 Why BOBA acts greedily

3.1 The pragmatic term does not rank candidates

Proposition 1 (Ordinal equivalence)**.** *Fix a time t and suppose $Y^\star \geq \mu_t(x)$ for every candidate x. Then over any set of candidates on which σ_t is constant, $-R$ is a strictly increasing function of μ_t, so* $\arg\max_x -R(x) = \arg\max_x \mu_t(x)$.

Proof. In Eq. 2 the only term involving μ is $(\mu - Y^\star)^2$. If $Y^\star \geq \mu$, this is strictly decreasing in μ, hence $-R$ is strictly increasing in μ. The remaining terms depend only on σ, which is constant by assumption. □

The proposition has a mirror image. If $Y^\star \le \mu_t(x)$ for every candidate, then $-R$ is strictly decreasing in μ_t and the ranking is again determined entirely by the posterior mean, with the sign reversed. Both regimes share one structural feature, which gives the general statement.

Proposition 2 (Bracketing is necessary)**.** *Over a set of candidates on which σ_t is constant, $-R$ fails to be a monotone function of μ_t only if the set brackets the preferred outcome, that is, only if there exist candidates x_1, x_2 with $\mu_t(x_1) < Y^\star < \mu_t(x_2)$.*

Proof. If no two candidates bracket $Y^\star$, then either $Y^\star \ge \mu_t(x)$ for all x or $Y^\star \le \mu_t(x)$ for all x, and Proposition 1 or its mirror image applies. □

We call a state *degenerate* when the preferred outcome lies outside the range of the candidates' predicted means: *upper-degenerate* if $Y^\star \ge \max_x \mu_t(x)$ (Proposition 1) and *mirror-degenerate* if $Y^\star \le \min_x \mu_t(x)$. Bracketing is necessary to escape degeneracy but not sufficient: when the target sits just inside one end of the predicted range, the ranking stays almost monotone (Appendix D).

Kelly et al. (2025) set the preferred outcome from the known optimum of each benchmark, so it sits at the attainable maximum by construction, and a Gaussian process with a constant mean rarely predicts a value above everything it has observed. We checked the condition at every recorded state from the synthetic pipeline (the eight benchmarks of §3.4 and thirteen further functions): it holds on 94.6% of $11{,}170$ states, and on 96.0% of the $6{,}770$ from the eight benchmarks. Section 4 shows that this rate is not a property of the benchmarks but of the objective's algebraic form.

Expanding Eq. 2, the acquisition function on such states is

$$a(x) \;=\; -\frac{\big(\mu(x) - Y^\star\big)^2}{2\sigma_p^2} \;+\; \Big[\log\sigma(x) - \frac{\sigma^2(x)}{2\sigma_p^2}\Big] \;+\; \beta\, I(x) \;+\; c, \tag{4}$$

where c collects constants. The first term is increasing in μ whenever $Y^\star$ lies above every prediction; the bracketed term depends only on σ and peaks at $\sigma = \sigma_p$. The method is therefore greedy in the posterior mean, with a weak preference for uncertainty near the noise level and against it above, plus an information bonus. Tuning β cannot rescue the pragmatic term, because scaling a quantity that induces no ordering of its own cannot make it induce one.

3.2 Measured on recorded optimisation states

We reloaded the surrogate saved with each recorded state, rebuilt its candidate set, and computed the Spearman correlation between the pragmatic term R and the negated predicted mean $-\mu$. In some states the predicted mean varies across the candidate set by less than one predictive standard deviation. We call these *flat* and the rest *live*; flat states make up 45.8% of the total, and on them the correlation mostly describes numerical noise. On the $6{,}049$ live states the median correlation is $+0.9970$, 64.4% exceed $+0.99$ and 78.5% exceed $+0.95$ (Appendix A, Fig. 1).

For BOBA[pub], the pragmatic term alone picks the highest-mean candidate on 78.7% of live states; the remainder reflects the σ-dependent terms of Eq. 4 and near-ties, which Proposition 1 holds fixed. The full acquisition function chose the same candidate as the pragmatic term alone on all 521 live states checked. The epistemic term changes so little because of scale: on recorded states the pragmatic term is of order 10^3 and the epistemic term of order 10^{-2}, so at $\beta = 1$ the information bonus cannot reorder candidates (Fig. 2). The collapse is also visible in behaviour: on the twin of §5, a pure exploitation arm and the pragmatic term alone are indistinguishable on all four paired tests (Holm $p = 1$; Appendix A).

3.3 Neither a perfect target nor an interior target helps

Proposition 1 predicts that the quality of the preferred outcome is irrelevant, because it does not enter the ranking. Direct tests at both extremes confirm this. Supplying the true global optimum across 16 benchmark–budget combinations improved performance in only 5 (median effect -0.09 SD, none significant after Holm correction). Placing the target strictly inside the observed range yielded no significant difference in 24 comparisons. Proposition 2 explains why: an interior target on a maximisation objective still leaves almost every candidate's predicted mean on one side of it, so the state is nominally bracketed but the target sits at the edge of the predicted range, where the ranking stays almost monotone.

3.4 CORRECTING THE LATENT

Active inference specifies the epistemic term over the hidden state of interest. Equation 3 measures information about the objective value at the chosen point x, whereas an optimiser needs information about the global optimum $\{x^\star, y^\star\}$. We therefore replace Eq. 3 with the joint entropy search quantity (Hvarfner et al., 2022),

$$I_{\text{opt}}(x) \;=\; H\big[p(y \mid x)\big] \;-\; \mathbb{E}_{\{x^\star,y^\star\}}\Big[H\big[p(y \mid x, x^\star, y^\star)\big]\Big], \tag{5}$$

where H is differential entropy and the expectation is over optima sampled from the surrogate posterior. Keeping the original pragmatic term gives BOBA-OPT: information about the optimum combined with a greedy pull toward the current best guess. Entropy search has the right latent but no preference; BOBA has a preference and the wrong latent; BOBA-OPT has both. BOBA-OPT uses the weight-matched convention, rescaling the information term at each step so the two terms' ranges match, so its like-for-like comparison is BOBA[wm].

Synthetic evaluation. We run dynamic versions of eight standard benchmarks (Ackley, Eggholder, Griewank, Hartmann-3, Hartmann-6, Powell, Schwefel, Shekel), with the last input coordinate taken as time as in Kelly et al. (2025), at budgets $T{=}30$ and $T{=}120$ with at least 50 repetitions. The field is GP-UCB (Srinivas et al., 2012), its sliding-window and resetting variants SW-UCB and R-UCB (Zhou & Shroff, 2021; Bogunovic et al., 2016), expected improvement (EI; Jones et al., 1998), alpha entropy search (AES; Fernández-Sánchez et al., 2025), both BOBA conventions and BOBA-OPT. We score mean instantaneous regret, $\bar{r} = \frac{1}{N-n_0}\sum_{i>n_0}[f^\star(t_i) - y_i]$, the average gap between the best value available at each step and the value obtained after the n_0 initial queries, which penalises a method that stops tracking (Appendix K).

Against its like-for-like control BOBA[wm], BOBA-OPT wins on seven of eight benchmarks at $T{=}120$ with no losses, and it beats BOBA[pub] on six (four at $T{=}30$), but at $T{=}120$ it has the lowest mean regret only on Hartmann-6, while GP-UCB has it on five of eight (Table 7). Raising BOBA's epistemic weight without correcting its latent makes it worse: BOBA[wm] loses to BOBA[pub] on three benchmarks at $T{=}120$. The surrogate matters as much as the acquisition function. An earlier configuration in which BOBA-OPT alone ran on an unconstrained surrogate, whose spatial lengthscale reached 403 against the bound of 20 applied elsewhere, made it appear the strongest method; matching the surrogate removed most of that advantage. In a factorial crossing the acquisition function with four surrogate settings, whether the noise variance is fixed or fitted explains more variance in regret than the choice of acquisition function does ($\Delta R^2 = 0.049$ against 0.030; Appendix B).

4 THE DEGENERACY IS IN THE ENCODING

Proposition 2 says the pragmatic term can only order candidates where the candidate set brackets the target. We now show that for setpoint problems the failure to bracket is forced by how the objective is written, and that changing the encoding removes it.

Folding places the target at the supremum. A setpoint problem has a known desired outcome $y^\star$ and asks for the input achieving it. Any maximiser needs a scalar to maximise, so the standard construction folds the residual into a distance,

$$f(x,t) = -\big|\, g(x,t) - y^\star \,\big|, \qquad \sup f = 0, \tag{6}$$

with the preferred outcome set to $Y^\star = 0$. The target then sits at the supremum of f for every choice of $y^\star$. The degeneracy of the folded encoding is therefore a consequence of its algebra, and the measured rates record only how often the surrogate's predicted means also stay below that ceiling.

The signed residual. Keeping the sign,

$$h(x,t) = g(x,t) - y^\star, \qquad Y^\star = 0, \tag{7}$$

places the target strictly inside the range of h. The simplest acquisition on each encoding is greedy: GREEDY-F queries $\arg\max_x \mu_f(x)$, and GREEDY-H, the *proximity rule*, queries $\arg\min_x |\mu_h(x)|$, the candidate predicted closest to target. The risk term's derivative becomes $\partial(-R)/\partial\mu_h = -\mu_h(x)/\sigma_p^2$,

Table 1: Degeneracy diagnostics under the two encodings. Bracketing is the fraction of acquisition states in which the candidates' predicted means straddle the preferred outcome; ρ is the median Spearman correlation between the pragmatic term and the negated predicted mean. Twin states pool the drifting and frozen surfaces (§5); synthetic states pool both budgets and all fourteen problems (Appendix D).

platform	encoding	states	$Y^\star \geq \mu_{\max}$	$Y^\star \leq \mu_{\min}$	bracketing	median ρ
twin	folded f	48,400	94.86%	0.10%	5.04%	$+0.9925$
twin	signed h	66,000	5.30%	37.00%	**57.70%**	-0.8225
synthetic	folded f	93,800	94.41%	0.00%	5.59%	$+0.9866$
synthetic	signed h	93,800	9.66%	5.88%	**84.46%**	$+0.1319$

positive below the target, negative above it and zero at it, so overshoot is penalised exactly as undershoot is. The preference width σ_p becomes a free parameter with a meaning of its own, the tolerance around the target, rather than being tied to the sensor noise (Appendix J sweeps it). Scoring is unaffected, since $|h| = -f$ exactly, so both encodings are evaluated against the same ground truth and the same regret. The uncertainty part of the risk term, $\log \sigma - \sigma^2/2\sigma_p^2$ (Eq. 4), has an interior maximum at $\sigma = \sigma_p$, so the risk term is not a straddle heuristic (Bryan et al.), which maximises predictive uncertainty near the level; it targets a particular uncertainty and is uncertainty-averse above it. Applying Eq. 5 to the signed objective with the latent taken to be the level-crossing location rather than the argmax gives BOBA-LS, which we estimate by sampling posterior paths and taking each path's crossing (Appendix D).

It removes the degeneracy. Table 1 and Figure 4 (Appendix D) report the diagnostics on the clinical twin of §5 and on the eight synthetic benchmarks given interior setpoints, together with six one-dimensional restrictions of them. Bracketing rises on all fourteen synthetic problems, every one by a large factor. The interior level was placed at the median of each objective's own distribution, the most favourable placement available to the folded encoding, and the folded encoding still collapses. In input space the candidate set brackets the true crossing whenever the level exists (Appendix D).

What the epistemic term is for. With the encoding fixed we can ask what information buys, by running the 2×2 design $\{$folded, signed$\} \times \{$greedy, greedy plus the information term$\}$. On the twin, information helps on the folded objective ($+0.545$ BIS, CI $[+0.131, +0.968]$) and does nothing measurable on the signed one (-0.226, CI $[-0.706, +0.237]$). In all eight synthetic cells (two suites $\times$ two budgets $\times$ two encodings) the information term is never significantly beneficial (Fig. 6), and on the full suite its main effect is negative at both budgets (-0.620 and -0.912 standard deviations, both intervals clear of zero). The reading that fits every cell is that the epistemic term substitutes for a preference term the encoding has disabled: on the signed objective the preference term orders candidates itself, and the bonus has nothing left to do. Appendix E gives all three suites, both versions of the design and every simple effect.

Where performance follows. The performance consequence is narrower. On the twin the signed encoding improves drifting location error ($+0.035$, CI $[+0.025, +0.046]$) with a positive interaction. On the full synthetic suite the interaction is null in all four cells, and the encoding's main effect is significantly negative in one (T=30, version A). Restricting to one dimension, where the level set is a point as on the twin, recovers the twin's sign in four of four cells (three significant, pre-registered). The separating variable appears to be whether the estimator has a true crossing to condition on (Fig. 7), so we claim the benefit for point level sets and treat $d > 1$ as untested.

5 A DIGITAL TWIN OF CLOSED-LOOP ANAESTHESIA

Under total intravenous anaesthesia, a pump delivers propofol and a monitor reports the bispectral index (BIS), a number between 0 and 100 derived from the patient's EEG that summarises depth of unconsciousness; high values mean close to awake, and clinical practice targets a value near 50. The controllable input is drug concentration and the observation is BIS. The target of 50 is interior in BIS, so too much drug is as wrong as too little, which makes this a setpoint problem in the sense of §4. Surgical stimulation, blood loss and drug accumulation move the relationship between concentration

and BIS within a single operation, so the concentration achieving 50 at one moment is not the one achieving it later. Patients differ enough that anaesthetists titrate rather than follow a fixed schedule.

Building simulated patients. We use VitalDB (Lee et al., 2022), an open dataset of $6{,}388$ surgical cases recorded at one-second resolution, including drug delivery, BIS, and a signal-quality index reporting when the monitor's reading is unreliable. It is observational: clinicians chose the concentrations, so we cannot evaluate an optimisation policy on the records directly. Instead, we fit a model to each patient and optimise against the model. The model has two standard components: the Eleveld pharmacokinetic model (Eleveld et al., 2018), which converts an infusion history into an effect-site concentration, and a Hill-sigmoid pharmacodynamic curve (V, 1910) with half-effect concentration Ce_{50} and slope γ, which maps that concentration to BIS. To these we add a slowly varying baseline capturing surgical stimulation, which is what makes each simulated patient's optimum move.

We selected adult cases lasting at least two hours with recorded concentration, BIS and signal quality, and drew 500 stratified by age, weight, sex and the extent of concentration variation. Fitting converged on all of them, and the fitted kinetics agree with the pumps' own concentration estimates at $r = 0.912$. Held-out patients fitted with their own parameters reach a median R^2 of 0.710, so the model form generalises. Held-out patients scored with population-average parameters reach $R^2 = -0.925$ and are worse than simply predicting each patient's mean BIS on 71 of 100 cases, with the half-effect concentration varying by a factor of 11.7 across individuals. So no single population-derived setting can serve every patient, which is the premise of personalisation and gives the fixed protocol a measured reason to fail. 115 of the 500 survive an eight-stage identifiability filter (Appendix F), and we draw 50 across strata. Each becomes an ordinary benchmark with its own input range, so every method runs on it unmodified.

Both encodings, and attainability. The twin exposes the objective in both forms of §4, $f = -|\mathrm{BIS} - 50|$ and $h = \mathrm{BIS} - 50$. Noise is applied to the BIS reading and the encoding map applied last, so $|h| = -f$ holds observation by observation and the two encodings differ only in what the acquisition function sees. Because the fitted model is analytic, the best achievable value at every instant is known in closed form, so instantaneous regret is exact rather than estimated, is measured in BIS units, and is directly interpretable: clinical practice treats readings within roughly 10 points of target as adequate. Each patient is simulated on two surfaces: *drifting*, where the stimulation baseline varies as fitted, and *frozen*, where it is held at its initial value. For some patients and some moments no concentration in the fitted range achieves BIS 50, and the ground-truth location is then clipped to a box edge. This is 45.2% of states on the frozen surface, where 23 of 50 patients are wholly unattainable, and 5.1 to 6.5% on the drifting surface. It is a real feature of the clinical problem, and §6 shows that methods differ sharply there.

Fitting three sources of randomness to the recordings gives cohort medians of 3.37 BIS observation noise, an Ornstein–Uhlenbeck residual drift with marginal standard deviation 4.76 BIS and a 1.86-minute timescale, and 10.2% sensor dropout (Appendix F).

6 Results on the digital twin

We run the field on 50 simulated patients with ten random seeds each, at budgets $T \in \{30, 60, 120\}$ and two initial-design sizes. The baseline that matters is not another acquisition function but the deployable alternative: a fixed concentration of $1.713\,\mu$g/mL, the value that achieves BIS 50 under population-median parameters, held for the whole operation. Two comparators sit outside the shared surrogate. Decay-BO (Jansen & Colley, 2026) forgets old data through a temporal decay kernel whose rate is fitted by marginal likelihood, in the forgetting-factor tradition of Bogunovic et al. (2016); its control, Decay-BO (no forgetting), fixes that rate at zero. Both use their own Gaussian process and search continuously rather than over the candidate set, so they form their own matched pair.

No method dominates. At $T{=}30$, BOBA-LS, GREEDY-H, BOBA-OPT and Decay-BO are statistically tied on location error. Decay-BO pulls ahead of GREEDY-H on regret by $T{=}60$ and on both metrics by $T{=}120$. GREEDY-H beats BOBA-OPT on regret from $T{=}60$ and BOBA-LS at $T{=}120$, and where the target is unreachable it degrades far less than the information-carrying arms.

Table 2: The field on the digital twin under D′ (dropped samples discarded), drifting surface, T=30 with eight initial queries, the cell in which every arm and both surfaces were run. Regret and distance are medians of per-patient means; distance is normalised location error. "On target" is the fraction of queries within 10 BIS of target; "worst patient" is the regret on each method's hardest patient; "beats protocol" is the share of the 50 patients on which the method has lower regret than the fixed protocol. ‡ uses its own surrogate and continuous search.

method	regret (BIS)	distance	on target	worst patient	beats protocol
BOBA-LS (signed)	**8.198**	0.1513	**0.672**	15.15	**68%**
Decay-BO (no forgetting)‡	8.363	0.1804	0.617	27.78	66%
Decay-BO‡	8.419	0.1619	0.638	15.60	66%
BOBA-OPT (folded)	8.835	0.1558	0.621	16.58	60%
GREEDY-H (signed)	8.937	**0.1492**	0.655	**14.73**	64%
GP-UCB	9.507	0.1814	0.607	17.19	60%
AES	9.775	0.2008	0.564	21.72	60%
EI	10.241	0.1975	0.579	17.10	58%
Fixed protocol	11.264	0.2327	0.447	31.50	n/a
Random	12.719	0.2663	0.438	28.64	42%

Realism ladder. To isolate the cost of each source of randomness, we evaluate (A) the noise-free fitted surface, (B) plus observation noise, (C) plus drift, and (D, D′) plus dropout. Under naive handling (D), dropped samples are passed to the surrogate as valid readings; under our primary condition (D′), they are discarded. Across fifteen arms, observation noise increases regret by 0.2–2.2 BIS and drift adds a further 0.2–1.5 BIS. Feeding the dropped readings to the surrogate as if they were real adds 2.1–4.2 BIS, while discarding them costs at most 0.5. Effort is better spent rejecting bad samples than recovering them (Appendix G).

6.1 Adaptation beats the protocol, and a two-line rule suffices

Table 2 reports the primary cell. Every adaptive method beats the fixed protocol for most patients. Because patients differ enormously, the comparison that matters is paired: for each patient we take the difference between two methods on that patient, then summarise those differences, with Holm correction over all 120 pairs among the sixteen arms run in this cell (Appendix H).

GREEDY-H beats the deployable clinical protocol on regret in all four cells tested, by 2.33 and 3.13 BIS at the short budget with eight and four initial queries, 3.13 at T=60 and 4.78 at T=120. It beats GP-UCB, EI and AES on regret in all four cells, and GP-UCB, EI, AES, Decay-BO (no forgetting) and random search on location error in every cell except AES at T=30 with four initial queries.

It also matches the active-inference arms. GREEDY-H is never distinguishable from BOBA-LS on location error and beats it on regret at T=120 (0.666 BIS); it overtakes BOBA-OPT on regret once the budget grows, by 0.596 BIS at T=60 (Holm $p = 0.0099$) and 0.832 at T=120. Once the objective is encoded so that the preference term can order candidates, the Expected Free Energy machinery adds nothing measurable over two lines of code, and at longer budgets it costs.

At the top of the field nothing is separable. Correcting over all 120 arm pairs rather than over a smaller family, BOBA-LS, GREEDY-H, BOBA-OPT and Decay-BO are mutually indistinguishable on drifting location error, so we claim no ordering among them. The three other than Decay-BO beat Decay-BO (no forgetting); Decay-BO's own margin over its control does not reach significance at this scope.

6.2 The comparison against the dynamic-BO family is budget and metric dependent

GREEDY-H matches Decay-BO at T=30 and falls behind by T=60 on regret (-0.761 BIS, Holm $p = 3.6 \times 10^{-4}$) and at T=120 on both metrics (-1.142 BIS and -0.0287). Any claim about the two must carry both a budget and a metric. Appendix H gives all four cells.

6.3 INFORMATION-SEEKING IS HARMFUL WHERE THE TARGET IS UNREACHABLE

Where the level is unattainable, methods separate sharply by whether they carry an information term. Taking the ratio of location error on unattainable states to attainable ones, arms carrying an information term degrade $1.93\times$, greedy and proximity rules $1.09\times$, and UCB-style and dynamic methods $1.38\times$ (Appendix I). Information-carrying arms degrade most in all five combinations of cell and surface that contain both kinds of state, spanning two budgets, two initial designs and both surfaces.

The mechanism is that there is nothing to explore for: the ground-truth location sits on a box edge for the whole slice, sitting on it is optimal, and an arm that keeps looking elsewhere pays for nothing. Clinically this regime matters most: the patient who cannot reach the target at any feasible dose is the one for whom a wandering controller is least acceptable.

6.4 OUTCOMES

No method is safe for every patient: the best worst-case performance in Table 2 is 14.73 BIS, well outside the 10 BIS adequate band. No method is misled by its own model, however. On a flat control surface where drug concentration has no effect, every arm finishes within 0.0021 BIS of random action, so the surrogates do not actively concentrate queries in harmful regions.

7 LIMITATIONS

The twin is fitted to observational data restricted to clinical concentrations, and the identifiability filter keeps records in which the anaesthetist varied the concentration, so absolute performance is optimistic; relative comparisons are unaffected. A slowly varying baseline accounts for much of the model's explanatory power, lowering the median fitted R^2 from 0.748 to 0.437 when removed.

Proposition 1 concerns BOBA's Gaussian preference over a scalar outcome, not active inference in general. The signed encoding's performance benefit is shown for point level sets; in $d > 1$ we attribute the negative cells to the estimator conditioning on a single sampled crossing rather than on the level manifold, and a manifold-aware estimator remains future work.

The factorial cannot separate the surrogate family from two other settings, Decay-BO uses its own surrogate, and frozen-surface runs exist only at $T{=}30$. The twin is a simulation: this paper reports no clinical deployment, and the worst-case results show why none should follow without further safeguards.

8 CONCLUSION

The encoding, not the acquisition function, decides whether a preference-based acquisition can work on a setpoint problem. Folding a setpoint into a distance places the target at the supremum, so BOBA's pragmatic term collapses onto the posterior mean; keeping the residual signed restores its ordering. Once the objective is well posed, a two-line proximity rule matches every active-inference variant we test, beats the clinical protocol at every budget, and avoids the near-doubling of location error that information-seeking arms incur where the target is unreachable, although a forgetting-kernel baseline is stronger at longer budgets. Redirecting BOBA's epistemic term at the optimum improves on BOBA but not on GP-UCB once surrogates are matched. Our results concern a Gaussian preference over a scalar outcome; they do not show that active-inference acquisition is greedy in general. We release the twin as a benchmark for setpoint acquisition.

REPRODUCIBILITY STATEMENT

The benchmark code, the patient-fitting pipeline and the analysis scripts are released. VitalDB is distributed under CC BY-NC-SA 4.0; we release the case-identifier list, every selection threshold and the fitting pipeline rather than per-patient fitted parameters, so the cohort can be reconstructed exactly from the source, and the per-case simulation outputs we release carry the same licence. Ground truth on the twin is analytic. Every arm is labelled with the weighting convention and the surrogate class recorded in its own session files, and Appendix B states which of those are verified and which are

unrecorded. Anonymised repository URL: https://anonymous.4open.science/r/setpoint-bo-release-4723/README.md

ETHICS STATEMENT

This work uses de-identified, publicly released clinical records from a dataset with its own institutional approval, and involves no new data collection and no intervention on any patient. All results are simulated. Automated control of anaesthetic depth carries evident risk, and we report explicitly that no method evaluated here keeps its worst-case patient within the clinically adequate range.

AI USE STATEMENT

In this work, we used generative AI tools to implement parts of the code for the digital twin and for the experiment, analysis and figure pipelines; to regenerate the paper's figures from our results files; to check the numbers reported in the manuscript against those files and identify errors and inconsistencies; and to edit the manuscript for clarity and length. We did not use generative AI tools for research ideation, and the propositions, experimental design and conclusions are the authors' own. We have reviewed all AI-assisted work: all generated code was checked for correctness by one of the authors, the corrected derivation and the proof were verified by hand, every cited reference was checked against its source, and every section of the manuscript was proofread by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Anthony Bardou and Patrick Thiran. Optimizing Through Change: Bounds and Recommendations for Time-Varying Bayesian Optimization Algorithms, January 2025. URL http://arxiv.org/abs/2501.18963. arXiv:2501.18963 [stat].

Anthony Bardou, Patrick Thiran, and Giovanni Ranieri. This Too Shall Pass: Removing Stale Observations in Dynamic Bayesian Optimization, October 2024. URL http://arxiv.org/abs/2405.14540. arXiv:2405.14540.

Ilija Bogunovic, Jonathan Scarlett, and Volkan Cevher. Time-Varying Gaussian Process Bandit Optimization, January 2016. URL http://arxiv.org/abs/1601.06650. arXiv:1601.06650 [stat].

Brent Bryan, Jeff Schneider, Robert C Nichol, Christopher J Miller, Christopher R Genovese, and Larry Wasserman. Active Learning For Identifying Function Threshold Boundaries.

Renzhi Chen and Ke Li. Transfer Bayesian Optimization for Expensive Black-Box Optimization in Dynamic Environment. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1374–1379, October 2021. doi: 10.1109/SMC52423.2021.9659200. URL https://ieeexplore.ieee.org/document/9659200/.

Lancelot Da Costa, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. Active inference on discrete state-spaces: a synthesis. *Journal of Mathematical Psychology*, 99:102447, December 2020. ISSN 00222496. doi: 10.1016/j.jmp.2020.102447. URL http://arxiv.org/abs/2001.07203. arXiv:2001.07203 [q-bio].

Lancelot Da Costa, Noor Sajid, Thomas Parr, Karl Friston, and Ryan Smith. Reward Maximization Through Discrete Active Inference. *Neural Computation*, 35(5):807–852, April 2023. ISSN 0899-7667. doi: 10.1162/neco_a_01574. URL https://doi.org/10.1162/neco_a_01574.

D. J. Eleveld, P. Colin, A. R. Absalom, and M. M. R. F. Struys. Pharmacokinetic-pharmacodynamic model for propofol for broad application in anaesthesia and sedation. *British Journal of Anaesthesia*, 120(5):942–959, May 2018. ISSN 1471-6771. doi: 10.1016/j.bja.2018.01.018.

Daniel Fernández-Sánchez, Eduardo C. Garrido-Merchán, and Daniel Hernández-Lobato. Alpha entropy search for new information-based Bayesian optimization. *Knowledge-Based Systems*, 322:113612, July 2025. ISSN 0950-7051. doi: 10.1016/j.knosys.2025.113612. URL https://www.sciencedirect.com/science/article/pii/S0950705125006586.

Karl Friston, Lancelot Da Costa, Noor Sajid, Conor Heins, Kai Ueltzhöffer, Grigorios A. Pavliotis, and Thomas Parr. The free energy principle made simpler but not too simple. *Physics Reports*, 1024:1–29, June 2023. ISSN 0370-1573. doi: 10.1016/j.physrep.2023.07.001. URL https://www.sciencedirect.com/science/article/pii/S037015732300203X.

Carl Hvarfner, Frank Hutter, and Luigi Nardi. Joint Entropy Search For Maximally-Informed Bayesian Optimization. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11494–11506. Curran Associates, Inc., 2022. doi: 10.52202/068431-0835. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4b03821747e89ce803b2dac590f6a39b-Abstract-Conference.html.

Pascal Jansen and Mark Colley. Pascal-Jansen/Bayesian-Optimization-for-Unity: v1.4.0, April 2026. URL https://zenodo.org/doi/10.5281/zenodo.19786494.

Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, December 1998. ISSN 1573-2916. doi: 10.1023/A:1008306431147. URL https://doi.org/10.1023/A:1008306431147.

Merlin Kelly, Rishan Patel, Alexander Thomas, Ziyue Zhu, Zikun Quan, Tom Carlson, and Youngjun Cho. BOBA: Dynamic Bayesian Optimization through Bayesian Active Inference. October 2025. URL https://openreview.net/forum?id=Yd6XR8jX5t.

Hyung-Chul Lee, Yoonsang Park, Soo Bin Yoon, Seong Mi Yang, Dongnyeok Park, and Chul-Woo Jung. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data*, 9(1):279, June 2022. ISSN 2052-4463. doi: 10.1038/s41597-022-01411-5. URL https://www.nature.com/articles/s41597-022-01411-5.

Ke Li, Renzhi Chen, and Xin Yao. A Data-Driven Evolutionary Transfer Optimization for Expensive Problems in Dynamic Environments, November 2022. URL http://arxiv.org/abs/2211.02879. arXiv:2211.02879 [cs].

Blake Mason, Romain Camilleri, Subhojyoti Mukherjee, Kevin Jamieson, Robert Nowak, and Lalit Jain. Nearly Optimal Algorithms for Level Set Estimation, November 2021. URL http://arxiv.org/abs/2111.01768. arXiv:2111.01768 [stat.ML].

Beren Millidge, Alexander Tschantz, and Christopher L. Buckley. Whence the Expected Free Energy? *Neural Computation*, 33(2):447–482, February 2021. ISSN 0899-7667. doi: 10.1162/neco_a_01354. URL https://ieeexplore.ieee.org/abstract/document/9346140.

Jonas Mockus. Application of Bayesian approach to numerical methods of global and stochastic optimization. *Journal of Global Optimization*, 4(4):347–365, June 1994. ISSN 1573-2916. doi: 10.1007/BF01099263. URL https://doi.org/10.1007/BF01099263.

Favour M. Nyikosa, Michael A. Osborne, and Stephen J. Roberts. Bayesian Optimization for Dynamic Problems, March 2018. URL http://arxiv.org/abs/1803.03432. arXiv:1803.03432 [stat].

Thomas Parr, Rajeev Vijay Rikhye, Michael M Halassa, and Karl J Friston. Prefrontal Computation as Active Inference. *Cerebral Cortex*, 30(2):682–695, March 2020. ISSN 1047-3211. doi: 10.1093/cercor/bhz118. URL https://doi.org/10.1093/cercor/bhz118.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, May 2012. ISSN 1557-9654. doi: 10.1109/TIT.2011.2182033. URL https://ieeexplore.ieee.org/document/6138914.

HILL A. V. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *J. Physiol.*, 40:iv–vii, 1910. URL https://cir.nii.ac.jp/crid/1571980074887256960.

Xingyu Zhou and Ness Shroff. No-Regret Algorithms for Time-Varying Bayesian Optimization, April 2021. URL http://arxiv.org/abs/2102.06296. arXiv:2102.06296 [cs.LG].

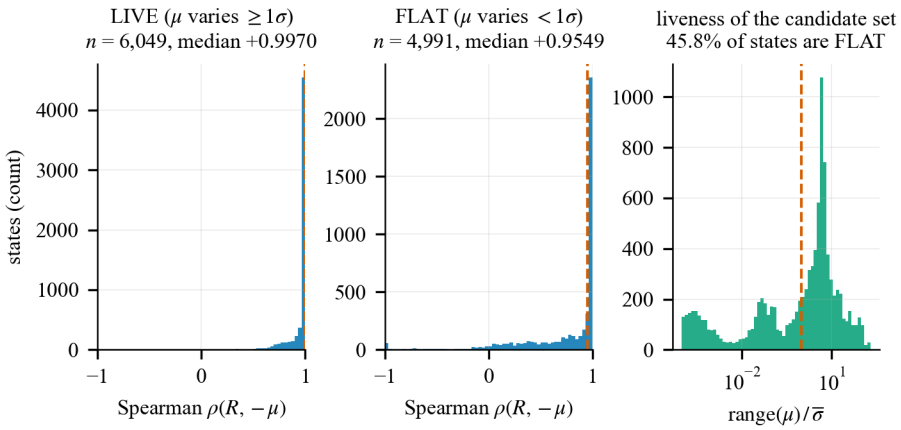

Figure 1: The pragmatic term ranks candidates the same way the predicted mean does. Left, on live states, where the predicted mean varies enough across candidates for a ranking to exist, the rank correlation between the pragmatic term and the negated predicted mean is concentrated at $+1$ (median $+0.9970$, $n = 6{,}049$). Centre, on flat states the coefficient describes numerical noise rather than the model, which is why the two must be separated before any correlation is quoted (median $+0.9549$, $n = 4{,}991$ with the correlation defined). Right, the distribution of how much the predicted mean varies across the candidate set, in units of the predictive standard deviation, over all $n = 11{,}170$ states; the dashed line is the threshold separating the two groups.

A THE DIAGNOSIS IN MORE DETAIL

How often the condition fails. The condition of Proposition 1 is recorded directly at each acquisition state as whether the preferred outcome exceeds every candidate's predicted mean (Table 3).

Table 3: States on which the condition of Proposition 1 fails, pooled over all BOBA arms on the synthetic suite.

states	total	condition fails	rate
all recorded states	11,170	603	5.40%
live	6,049	600	9.92%
flat	5,121	3	0.06%

By weighting convention, the condition holds on 97.96% of the 3,240 states recorded under BOBA[pub], 95.33% of 1,540 under BOBA[pub] with output normalisation (the objective standardised before scoring, as in the BOBA code release), 92.11% of 4,550 under BOBA[wm], and 94.2% of 1,840 under an adaptive-weight variant of BOBA (1,770 unnormalised, 70 normalised). The spread tracks which benchmarks each convention was run on rather than the convention itself: neither the weighting mode nor the normalisation appears in the target projection or in μ, so the condition is structurally independent of both.

Where flatness comes from. Flat states are caused mostly by the lower bound on the signal variance, not by the acquisition function. In a separate variance-floor ablation, with the bound at 10^{-4} 48.5% of states are live and 37.2% of candidates have their predicted mean pinned; with it at 10^{-10}, 62.7% are live and none are pinned. Median standardised regret is the same in both (2.545 against 2.541). Flatness and regret have different causes.

Rank correlation by set. Median Spearman correlation between the pragmatic term and the negated predicted mean, with denominators attached: live states pooled over conventions $+0.9970$ ($n = 6{,}049$); flat states $+0.9549$ ($n = 4{,}991$); all defined states $+0.9932$ ($n = 11{,}040$); live states under BOBA[pub] $+0.9951$ ($n = 1{,}262$); live states under BOBA[wm] $+0.9975$ ($n = 3{,}411$). On the twin, the folded encoding gives $+0.9925$ over 48,368 states with the correlation defined and the signed encoding -0.8225 over 65,957.

The decision, by convention. Under BOBA[pub], the rate at which the pragmatic term alone selects the highest-mean candidate is 0.7868 over 1,262 live states, and the rate at which the full acquisition function selects the same candidate as the pragmatic term alone is 1.0000 over the 521 live states checked. Under BOBA[wm] the latter falls to 0.0495 over 3,411 states, so raising the epistemic term's effective weight by four orders of magnitude does change which candidate is chosen in roughly 95% of live states, without improving performance (§3.4).

Behavioural confirmation. A pure exploitation arm and the preference term scored alone on the folded objective, paired over 50 patients with a 10,000-sample bootstrap and Holm correction over the four tests (Table 4).

Table 4: Pure exploitation against the pragmatic term alone on the folded twin objective, T=30, eight initial queries, D′. Difference is pragmatic term minus exploitation; $n = 50$ patients.

surface	metric	difference	95% CI	Holm p
drifting	mean instantaneous regret	−0.0774	[−0.467, +0.325]	1
drifting	distance to optimum	−0.000078	[−0.0166, +0.0169]	1
frozen	mean instantaneous regret	+0.103	[−0.220, +0.451]	1
frozen	distance to optimum	+0.0092	[−0.0072, +0.0256]	1

Four of four indistinguishable, every interval straddling zero. The per-state rate at which the two rules select the same candidate is 56.8% on the folded objective, not close to 100%; a rank correlation of +0.99 with 57% selection agreement is the signature of near-ties, and the outcome table shows those disagreements are not systematic.

Scale. Figure 2 recomputes the acquisition function on one recorded state (Eggholder) across a logarithmic grid of the exploration weight w and plots the candidate it selects. The selection is a step function with 26 treads, the last taking over at $w = 1461.27$. The weight of BOBA[pub], $w = \beta = 1$, sits inside the first tread, which is why the epistemic term never changes the decision there. The median recorded weight under BOBA[wm] is $w = 1.258 \times 10^4$, past the last tread.

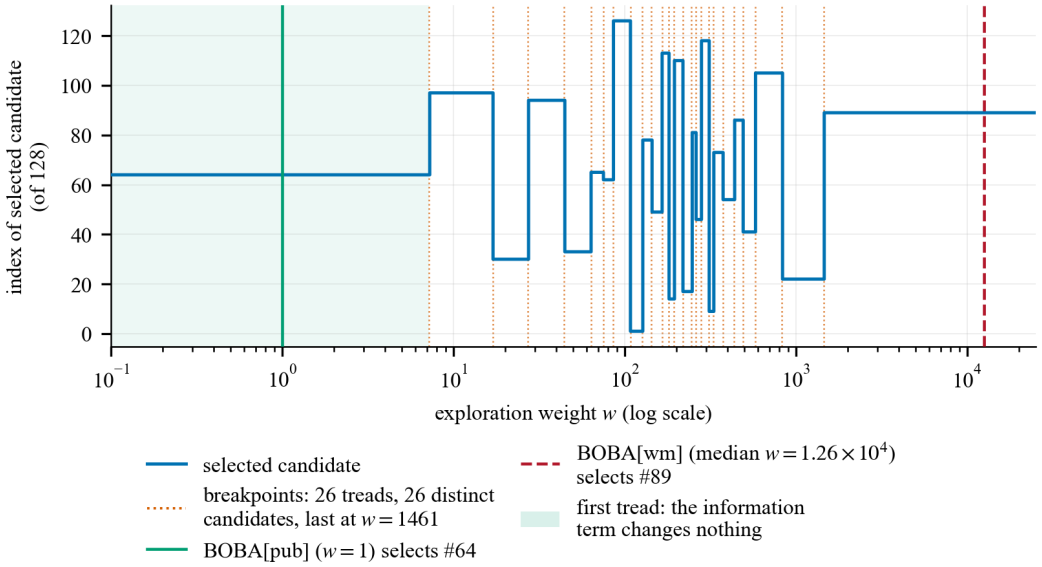

Figure 2: Candidate selected as a function of the exploration weight w on one recorded live acquisition state, 128 candidates, 26 treads, last breakpoint at $w = 1461.27$. Drawn under BOBA[pub], where w equals β. The information term only begins to change the decision when the weight is several orders of magnitude above 1.

B SURROGATE CONFIGURATION, PROVENANCE, AND THE FACTORIAL EXPERIMENT

Table 5: The shared surrogate of Eq. 1 against the default configuration of entropy-search code.

property	shared surrogate (Eq. 1)	entropy-search convention
spatial lengthscales	one, shared, in $[0.005, 20]$	one per dimension, unbounded
observation noise	fixed from the data	fitted freely
signal variance	bounded below by 10^{-4}	no bound
output standardisation	once	twice

In the matched setup, methods that need batch or fantasy semantics (BOBA-OPT, AES, BOBA-LS) run on a BoTorch single-task GP whose hyperparameters are copied from the shared surrogate rather than refitted; predictive means agree with the shared surrogate to within 2.1×10^{-10} on every session checked.

Which arms carry a verified surrogate class. Every arm in the twin tables is verified on the shared space-time surrogate of Eq. 1 from its own session record, except two pre-fix arms that are verified to have run on an unconstrained single-task GP and are reported only as archive evidence, never in a results table. On the synthetic suite only BOBA[pub], BOBA[pub] with normalisation, BOBA-OPT and AES carry a recorded surrogate class; the baseline columns predate the provenance field, so their surrogate is unrecorded rather than verified. BOBA[wm] at the short budget is mixed, with 190 sessions recorded on the shared surrogate, 160 on an unconstrained single-task GP and 700 unrecorded. We state this rather than assert uniformity we cannot demonstrate.

The factorial, and what is aliased. The experiment crosses the acquisition function (BOBA-OPT against a control using BOBA's acquisition function) with four surrogate settings, 80 runs per cell: the surrogate family (the shared space-time GP of Eq. 1, or the native single-task GP used by entropy-search code); *noise anchoring*, fixing the noise variance from the data rather than fitting it; *ARD*, one lengthscale per input dimension rather than one shared; and the *floor*, a lower bound on the signal variance. The native family was run only without a floor and only with noise fitted freely, so its effect cannot be separated from those two settings and is not estimable. We therefore analyse only the space-time runs ($n = 1{,}280$), in which the other four factors are fully crossed (Table 6). Regret is standardised within benchmark, since raw regret spans four orders of magnitude across benchmarks. In the table the acquisition effect is coded so that positive means BOBA-OPT has lower regret.

Table 6: Factorial decomposition of standardised regret on the space-time runs, $n = 1{,}280$, model $R^2 = 0.0914$. Terms ranked by ΔR^2, the drop in R^2 when the term is removed and the model refitted. Effect is half the difference between the factor's two levels.

rank	term	effect (SD)	ΔR^2
1	noise anchoring	-0.2312	0.0491
2	acquisition function	$+0.1815$	0.0303
3	anchoring $\times$ ARD	$+0.0680$	0.0043
4	acquisition $\times$ floor	$+0.0659$	0.0040
5	ARD	-0.0422	0.0016
6	floor	$+0.0341$	0.0011
7	acquisition $\times$ ARD	-0.0310	0.0009
8	anchoring $\times$ floor	-0.0140	0.0002

The held-fixed contrast, BOBA-OPT minus the control within each configuration of anchoring, floor and ARD, is expressed as a difference in standardised regret, so negative favours BOBA-OPT: mean -0.3840 SD, median -0.4386, range -0.7537 to -0.0526. Ten of ten configurations favour BOBA-OPT on the point estimate and seven of ten have an interval entirely below zero, so quoting ten of ten alone overstates the resolution. Halved and sign-flipped to the table's coding this is $+0.1920$ SD against the main effect's $+0.1815$, agreement to 0.011 SD.

C SYNTHETIC RESULTS IN FULL

Table 7: Long budget (T=120), mean instantaneous regret, mean ± standard error over n=50 sessions per cell. Lower is better; the lowest mean in each row is in bold. Per-slice optima, which regret is measured against, were computed by global search (differential evolution followed by L-BFGS-B).

benchmark	GP-UCB	SW-UCB	R-UCB	EI	AES	BOBA[pub]	BOBA[wm]	BOBA-OPT
Ackley	4.29 ± 0.19	$\mathbf{3.64} \pm 0.15$	3.80 ± 0.14	4.43 ± 0.14	5.40 ± 0.07	5.62 ± 0.15	5.61 ± 0.08	4.98 ± 0.11
Eggholder	$\mathbf{448.8} \pm 9.3$	479.8 ± 6.8	475.1 ± 6.0	475.6 ± 6.6	562.8 ± 6.7	494.6 ± 8.1	501.9 ± 10.0	487.7 ± 10.9
Griewank	$\mathbf{31.9} \pm 4.5$	59.2 ± 6.4	95.3 ± 7.7	84.2 ± 1.6	50.3 ± 4.1	82.4 ± 6.7	147.7 ± 0.4	49.3 ± 5.5
Hartmann-3	0.736 ± 0.060	0.952 ± 0.063	0.791 ± 0.044	0.425 ± 0.034	$\mathbf{0.407} \pm 0.051$	1.311 ± 0.056	0.704 ± 0.035	0.501 ± 0.039
Hartmann-6	1.531 ± 0.053	1.862 ± 0.045	2.026 ± 0.036	1.826 ± 0.025	1.753 ± 0.022	1.957 ± 0.049	1.913 ± 0.028	$\mathbf{1.516} \pm 0.033$
Powell	$\mathbf{1024} \pm 56$	1881 ± 138	3834 ± 356	1788 ± 54	1875 ± 159	3722 ± 369	6077 ± 95	1668 ± 128
Schwefel	$\mathbf{729.1} \pm 25.8$	1003.3 ± 19.0	1001.3 ± 19.1	1018.6 ± 12.6	1215.4 ± 4.6	801.3 ± 25.6	1226.3 ± 6.2	1149.8 ± 10.2
Shekel	$\mathbf{2.436} \pm 0.016$	2.560 ± 0.018	2.607 ± 0.020	2.576 ± 0.016	2.585 ± 0.010	2.592 ± 0.016	2.610 ± 0.018	2.481 ± 0.017

Table 8: Short budget (T=30), mean instantaneous regret, mean ± standard error. Lower is better; the lowest mean in each row is in bold. This cell is unbalanced: BOBA[wm] ranges from 60 to 214 sessions per benchmark, GP-UCB from 50 to 155 and BOBA-OPT from 100 to 116, AES has 100 throughout, and SW-UCB, R-UCB, EI and BOBA[pub] have exactly 50. Per-benchmark n is given in italics beneath each row. The BOBA[wm] column includes 160 sessions run on an unconstrained single-task GP (Appendix B).

benchmark	GP-UCB	SW-UCB	R-UCB	EI	AES	BOBA[pub]	BOBA[wm]	BOBA-OPT
Ackley	5.53 ± 0.17	5.72 ± 0.14	$\mathbf{5.32} \pm 0.19$	5.48 ± 0.15	6.06 ± 0.06	6.17 ± 0.14	5.72 ± 0.08	5.78 ± 0.06
n	*55*	*50*	*50*	*50*	*100*	*50*	*60*	*100*
Eggholder	561.3 ± 4.7	579.2 ± 7.3	572.4 ± 5.1	559.0 ± 7.4	579.3 ± 5.2	559.5 ± 6.3	597.1 ± 4.9	$\mathbf{554.9} \pm 5.9$
n	*100*	*50*	*50*	*50*	*100*	*50*	*100*	*100*
Griewank	$\mathbf{63.4} \pm 4.3$	104.4 ± 3.6	106.8 ± 7.7	102.4 ± 4.2	124.0 ± 2.8	84.8 ± 6.1	122.0 ± 2.8	91.1 ± 2.8
n	*55*	*50*	*50*	*50*	*100*	*50*	*124*	*116*
Hartmann-3	0.889 ± 0.038	0.961 ± 0.055	0.963 ± 0.054	0.920 ± 0.049	0.864 ± 0.034	1.346 ± 0.064	1.030 ± 0.023	$\mathbf{0.760} \pm 0.024$
n	*155*	*50*	*50*	*50*	*100*	*50*	*174*	*116*
Hartmann-6	$\mathbf{1.819} \pm 0.025$	1.958 ± 0.040	1.895 ± 0.038	1.824 ± 0.038	2.072 ± 0.016	1.987 ± 0.039	1.922 ± 0.015	1.859 ± 0.024
n	*150*	*50*	*50*	*50*	*100*	*50*	*164*	*116*
Powell	4185 ± 243	$\mathbf{2570} \pm 283$	4157 ± 345	4778 ± 325	5061 ± 168	2858 ± 196	7467 ± 326	4790 ± 176
n	*150*	*50*	*50*	*50*	*100*	*50*	*100*	*100*
Schwefel	1021.5 ± 25.6	1053.2 ± 37.3	1055.3 ± 30.4	1079.5 ± 32.9	1254.7 ± 5.9	$\mathbf{889.8} \pm 38.9$	1210.2 ± 8.8	1218.6 ± 8.1
n	*50*	*50*	*50*	*50*	*100*	*50*	*114*	*116*
Shekel	2.506 ± 0.011	2.590 ± 0.015	2.579 ± 0.019	2.529 ± 0.014	2.594 ± 0.006	2.573 ± 0.017	$\mathbf{2.497} \pm 0.007$	2.523 ± 0.008
n	*100*	*50*	*50*	*50*	*100*	*50*	*214*	*116*

Win and loss counts. Per-benchmark Mann-Whitney tests on session values with Holm correction over the eight benchmarks. At the long budget, BOBA-OPT against BOBA[pub] is six wins, one loss and one tie on regret, and five wins with no losses on location error; against BOBA[wm] it is seven wins and no losses on regret. BOBA[wm] against BOBA[pub] is one win and three losses on regret at the long budget. At the short budget BOBA-OPT against BOBA[pub] is four wins and two losses on regret, the losses being on Powell and Schwefel.

Timing on the synthetic suite. Over the first third of each long-budget run, unpaired effect sizes of BOBA-OPT against expected improvement have a median of -0.35 across the eight benchmarks, and over the remainder $+0.17$ (Fig. 3).

D THE ENCODING RESULT IN FULL

Per-benchmark bracketing rates under the two encodings are in Table 9.

By budget and surface. On the twin the folded encoding brackets on 3.98% of drifting states and 6.11% of frozen ones, the signed encoding on 67.02% and 48.38%. On the synthetic suite the folded encoding brackets on 6.15% at T=30 and 5.48% at T=120, the signed encoding on 72.75% and 86.76%.

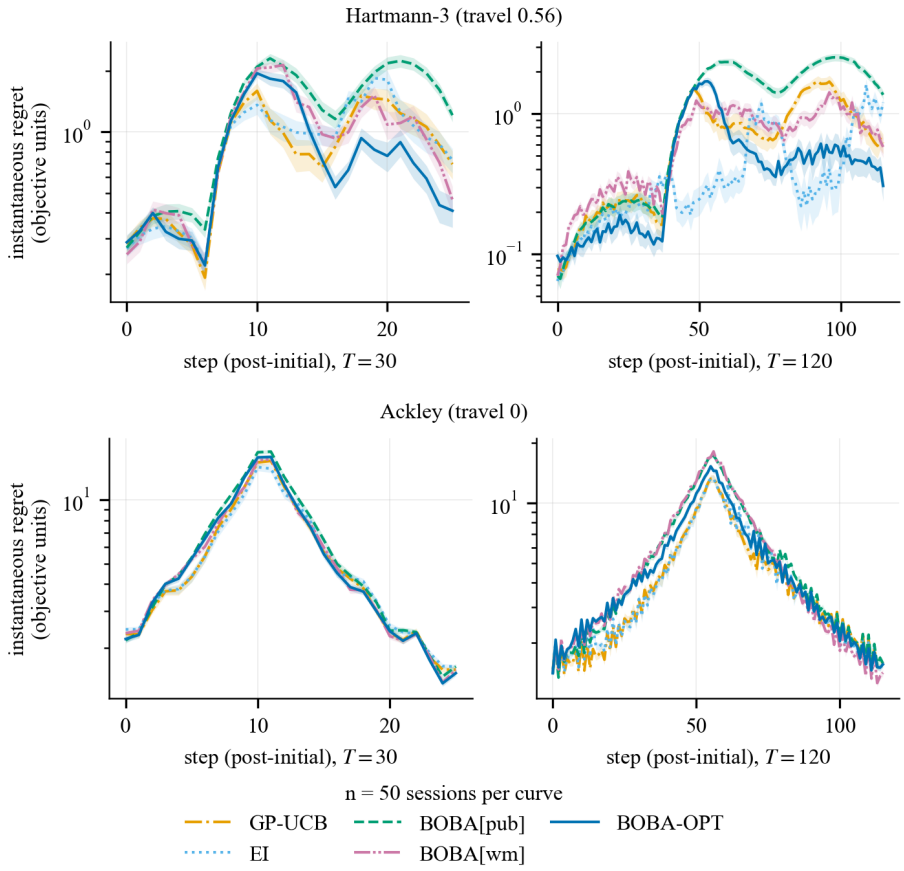

Figure 3: Instantaneous regret over the course of a run on a benchmark whose optimum moves (Hartmann-3; the optimum travels 0.56 of the box diagonal) and one whose optimum does not (Ackley, travel 0), at both budgets, matched surrogate, mean and standard error across repetitions.

Table 9: Share of acquisition states whose candidate set brackets the preferred outcome, both budgets pooled. The eight benchmarks are listed first and their one-dimensional restrictions second.

benchmark	folded f	signed h
Ackley	1.1%	91.2%
Eggholder	0.5%	74.4%
Griewank	0.9%	96.1%
Hartmann-3	2.3%	86.2%
Hartmann-6	11.4%	95.6%
Powell	19.9%	99.1%
Schwefel	0.6%	88.2%
Shekel	4.8%	97.4%
Ackley (1-D)	2.4%	60.3%
Griewank (1-D)	10.1%	82.9%
Hartmann-6 (1-D)	7.1%	97.1%
Powell (1-D)	12.5%	98.1%
Schwefel (1-D)	1.0%	33.4%
Shekel (1-D)	3.7%	82.5%

Bracketing is necessary but not sufficient. Within the two degenerate regimes the absolute rank correlation is 0.98 to 0.99, as the monotonicity argument requires. But the folded encoding's bracketing states still show a median correlation of $+0.980$ on the twin, because bracketing there is marginal: the median position of the target within the predicted range is 0.041 and only 8.9%

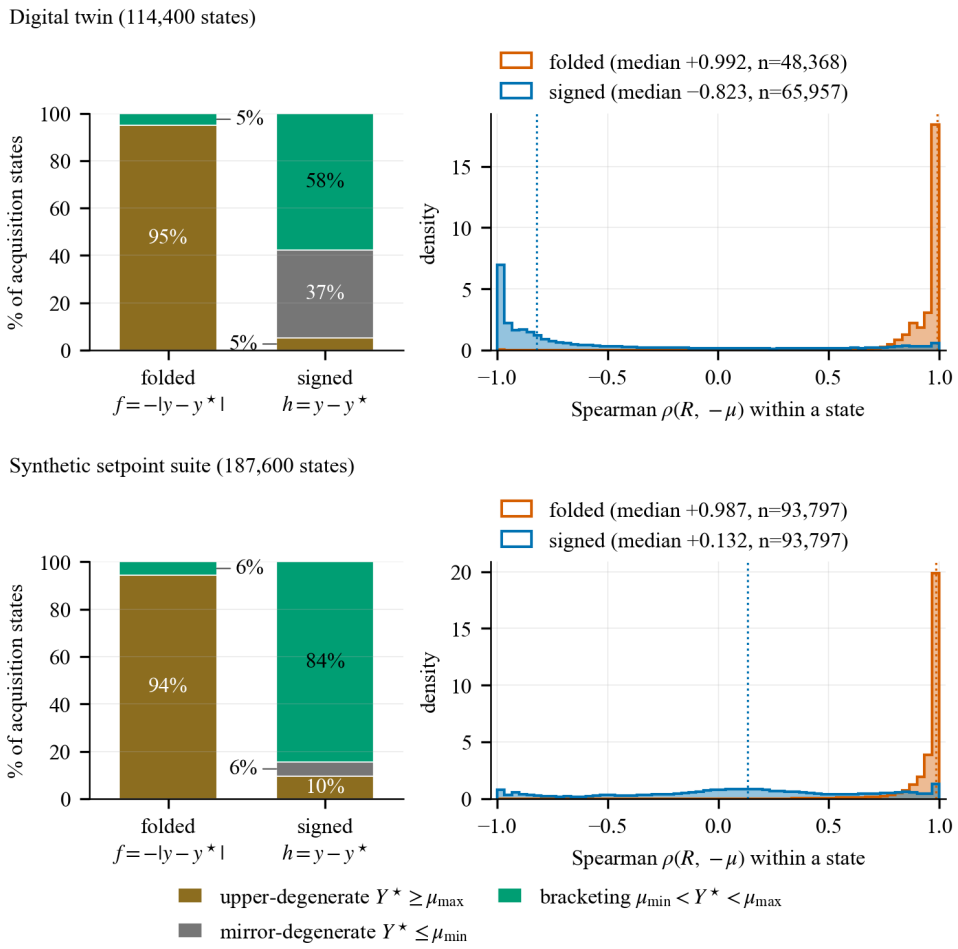

Figure 4: Left of each row, the three regimes as a share of acquisition states under the folded and signed encodings. Right, the full distribution of the within-state rank correlation between the pragmatic term and the negated predicted mean, medians marked. Top row, the clinical twin, $n = 48{,}400$ folded and $66{,}000$ signed states. Bottom row, fourteen synthetic setpoint problems, $n = 93{,}800$ states per encoding.

of those states place it centrally (Fig. 5). The signed encoding's bracketing states sit at 0.651 with 63.3% central, and the correlation falls to -0.616 with 29% of states below 0.5 in absolute value. On the synthetic suite the same quantities are 0.010 with 0.7% central for the folded encoding and 0.556 with 56.8% central for the signed one.

Level-crossing estimation. BOBA-LS estimates the epistemic term by sampling posterior paths and taking each path's crossing of the target level. Measured over $4{,}400$ states, 26% of sampled paths have no crossing and 52% have more than one, with a mean of 6.0 per path. Paths with no crossing are kept and their closest approach used, since discarding them would condition only on draws that agree the level exists; paths with several crossings contribute one drawn uniformly, so the sample spans the posterior over crossing locations rather than collapsing to the lowest dose.

Input-space bracketing. Over all 50 reported patients and $1{,}000$ time slices each, on the same 128-point candidate set the experiments search, the candidate set brackets the true level-crossing location on 91.06% of drifting slices and 54.00% of frozen ones. The level itself exists on 91.18% and 54.00% of those slices; these are shares of all slices, whereas Appendix I counts post-initial query states, hence the slightly different rates. Where the level exists the candidate set brackets it on 99.84% of drifting and 100.00% of frozen slices; where it does not, it never can, because the clipped

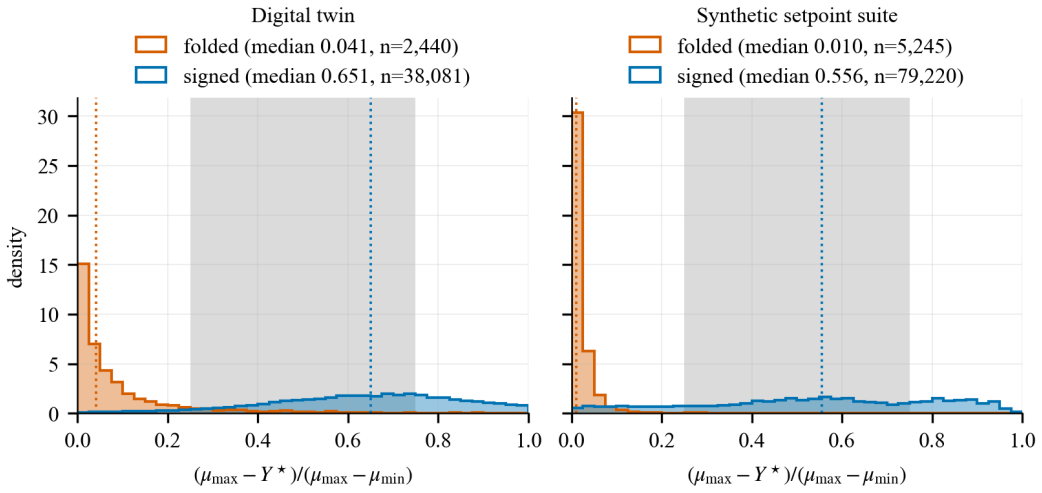

Figure 5: Position of the preferred outcome within the predicted range, over bracketing states only. Twin $n = 2{,}440$ folded and $38{,}081$ signed, medians 0.041 and 0.651; synthetic $n = 5{,}245$ folded and $79{,}220$ signed, medians 0.010 and 0.556. The shaded band is the centrally placed interval $[0.25, 0.75]$.

location lies on a box edge and no candidate lies beyond it. Outcome bracketing follows from input bracketing, which follows from the target being reachable.

E THE 2×2 ACROSS ALL THREE SUITES

The three suites are the twin, the full-dimensional synthetic suite (eight benchmarks) and its one-dimensional restrictions (six). The acquisition factor is {greedy, greedy plus the information term} and the objective factor is {folded, signed}, giving four arms. The greedy arms are GREEDY-F and GREEDY-H in both versions. In version A the two information arms are the shipped ones: BOBA-OPT on the folded objective and BOBA-LS on the signed one. Version B replaces the signed information arm with GREEDY-H plus BOBA-LS's level-crossing information term, so on the signed objective the two acquisition levels differ only by that term. The two versions agree on every sign and every verdict. Positive is better throughout, since both metrics are lower-is-better and every effect is a reduction.

Table 10: The 2×2 on the twin, version B, T=30, raw units (BIS for regret, normalised location error for distance), $n = 50$ patients, 10,000-sample bootstrap intervals.

surface	metric	objective ($f \to h$)	acquisition	interaction
Eight initial queries				
drifting	regret	$+0.804$ $[+0.477, +1.166]$	$+0.160$ $[-0.196, +0.497]$	$+0.771$ $[+0.225, +1.340]$
drifting	distance	$+0.0352$ $[+0.0253, +0.0463]$	$+0.0275$ $[+0.0159, +0.0393]$	$+0.0419$ $[+0.0222, +0.0618]$
frozen	regret	$+0.111$ $[-0.188, +0.416]$	-1.402 $[-1.817, -1.007]$	$+0.799$ $[+0.333, +1.266]$
frozen	distance	$+0.0056$ $[-0.0084, +0.0202]$	-0.0477 $[-0.0676, -0.0275]$	$+0.0230$ $[+0.0016, +0.0458]$
Four initial queries				
drifting	regret	$+0.133$ $[-0.213, +0.535]$	-0.197 $[-0.613, +0.199]$	$+0.644$ $[+0.195, +1.074]$
drifting	distance	$+0.0121$ $[+0.0025, +0.0222]$	$+0.0267$ $[+0.0108, +0.0432]$	$+0.0302$ $[+0.0109, +0.0493]$
frozen	regret	$+0.101$ $[-0.288, +0.506]$	-2.017 $[-2.633, -1.442]$	$+1.124$ $[+0.514, +1.780]$
frozen	distance	$+0.0072$ $[-0.0051, +0.0197]$	-0.0529 $[-0.0779, -0.0278]$	$+0.0369$ $[+0.0123, +0.0626]$

Simple effects. The acquisition main effect averages over the objective and must not be read as uniform. Table 12 and Figure 6 give the information term's simple effect within each encoding, version B.

Two readings follow. The twin's positive interaction is positive because one arm of it is a gain: information helps on the folded encoding and does nothing on the signed one. The one-dimensional

Table 11: The 2×2 on the synthetic suites, both versions, regret standardised within benchmark; n is the number of benchmarks.

suite	ver	T	objective	acquisition	interaction	n
full-d	A	30	-0.201 $[-0.313, -0.079]$	-0.736 $[-1.077, -0.344]$	-0.129 $[-0.492, +0.262]$	8
full-d	B	30	-0.085 $[-0.191, +0.011]$	-0.620 $[-0.847, -0.324]$	-0.361 $[-0.684, +0.043]$	8
full-d	A	120	$+0.177$ $[-0.219, +0.649]$	-1.010 $[-1.334, -0.688]$	$+0.077$ $[-0.552, +0.709]$	8
full-d	B	120	$+0.275$ $[-0.188, +0.757]$	-0.912 $[-1.156, -0.645]$	-0.119 $[-0.540, +0.255]$	8
1-D	A	30	$+0.595$ $[+0.128, +1.062]$	-0.329 $[-0.587, -0.020]$	$+0.361$ $[+0.068, +0.595]$	6
1-D	B	30	$+0.637$ $[+0.093, +1.181]$	-0.287 $[-0.568, +0.003]$	$+0.277$ $[-0.035, +0.628]$	6
1-D	A	120	$+0.394$ $[-0.032, +0.820]$	-0.327 $[-0.765, +0.111]$	$+0.725$ $[+0.264, +1.182]$	6
1-D	B	120	$+0.489$ $[-0.074, +1.044]$	-0.232 $[-0.748, +0.249]$	$+0.535$ $[+0.308, +0.836]$	6

Table 12: Simple effect of adding the information term, within each encoding, version B. Twin rows at T=30 with eight initial queries in raw units; synthetic rows in within-benchmark standard deviations.

suite	cell	on folded	on signed
twin	drifting, regret	$+0.545$ $[+0.131, +0.968]$	-0.226 $[-0.706, +0.237]$
twin	drifting, distance	$+0.0485$ $[+0.0328, +0.0645]$	$+0.0066$ $[-0.0078, +0.0216]$
twin	frozen, regret	-1.003 $[-1.518, -0.481]$	-1.801 $[-2.221, -1.376]$
twin	frozen, distance	-0.0362 $[-0.0597, -0.0103]$	-0.0591 $[-0.0805, -0.0389]$
full-d	T=30	-0.801 $[-1.153, -0.315]$	-0.439 $[-0.596, -0.294]$
full-d	T=120	-0.971 $[-1.375, -0.559]$	-0.852 $[-1.066, -0.663]$
1-D	T=30	-0.149 $[-0.450, +0.216]$	-0.426 $[-0.779, -0.140]$
1-D	T=120	$+0.035$ $[-0.594, +0.647]$	-0.500 $[-0.930, -0.105]$

synthetic interaction is positive for a different reason, since information is neutral on the folded encoding there and harmful on the signed one. The sign is shared and the mechanism is not. What unifies them is the negative statement: across all eight synthetic rows the information term never significantly helps.

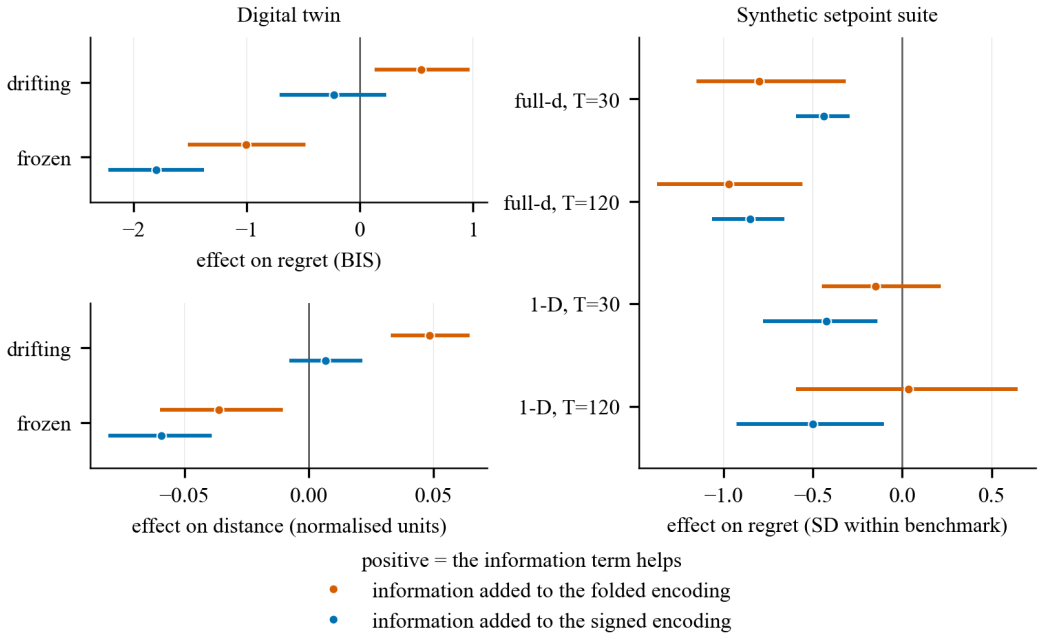

Figure 6: The information term's simple effect within each encoding, version B, with 10,000-sample bootstrap intervals. Twin $n = 50$ patients per cell; synthetic $n = 8$ benchmarks for the full-dimensional suite and $n = 6$ for the one-dimensional one, regret standardised within benchmark.

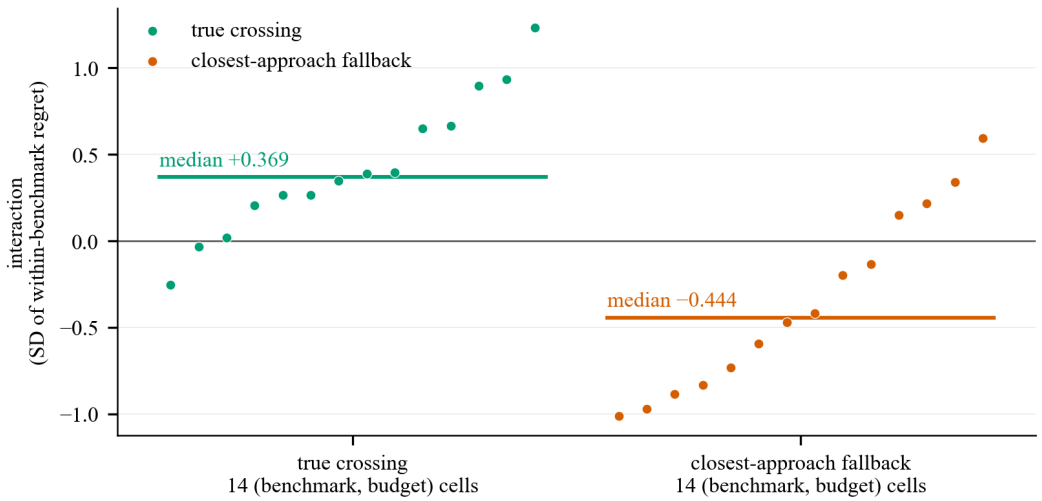

Figure 7: Per-benchmark and per-budget interaction in units of within-benchmark regret standard deviation, split by whether the epistemic estimator conditioned on a true level crossing (seven benchmarks, fourteen cells, median $+0.369$, twelve positive) or on the closest-approach fallback (seven benchmarks, fourteen cells, median -0.444, four positive). The estimator mode is recorded per session; it coincides with the problem's spatial dimension, since a true crossing exists on the seven problems with one spatial dimension (the six restrictions and Eggholder).

F Digital twin: construction

Table 13: The eight-stage identifiability filter applied to the 500 fitted cases.

stage	criterion	remaining	removed
0	drawn, stratified	500	
1	Hill slope γ not on a fitting bound	356	144
2	Ce_{50} not on a fitting bound	320	36
3	baseline identifiable (minimum concentration below 0.5 Ce_{50})	170	150
4	maximal effect not on a fitting bound	127	43
5	effect span at least 10 BIS	120	7
6	own-parameter held-out $R^2 \geq 0.50$	119	1
7	BIS 50 attainable at some concentration	116	3
8	concentration achieving BIS 50 inside the case's input range	115	1

Thresholds come from the data: $R^2 \geq 0.50$ because the fifth percentile on the training set is 0.503, and an effect span of 10 BIS because that is about 1.75 times the median residual standard deviation of 5.74 BIS. The 50 evaluation patients are drawn one per stratum from 50 of the 70 strata the eligible pool occupies. Baseline identifiability is the largest single exclusion.

Fitted randomness, and what the primary condition realised. Cohort medians with fifth and ninety-fifth percentiles: slow surgical-stimulation standard deviation 6.10 BIS (3.10 to 10.88); residual drift standard deviation 4.76 BIS (3.22 to 7.39) with an autoregressive coefficient at one minute of 0.583 and a pull-back timescale of 1.86 minutes (1.01 to 3.81); observation noise standard deviation 3.37 BIS (2.24 to 4.31); total residual standard deviation 5.74 BIS; measured signal-quality dropout 10.2% of samples, at 3.75 events per hour with a median event length of 0.5 minutes. As a check that the simulator reproduces what it is given, one patient (case 1667) was simulated under the primary condition and its realised values compared with that patient's own fitted values: observation noise 1.8927 against 1.8986 BIS, one-minute drift innovation standard deviation 3.2003 against 3.2094, autoregressive coefficient 0.6764 against 0.6787, and dropout fraction 0.0425 against 0.0420, all within 1%. The fixed protocol concentration is 1.7126 μg/mL, which clips for four of the fifty patients, all downward.

Fit quality and the personalisation premise. Kinetics agree with the pumps' own estimates at Pearson $r = 0.912$, with bias $-0.299\,\mu$g/mL and root mean squared error $0.473\,\mu$g/mL. Held-out cases fitted with their own parameters reach median $R^2 = 0.710$ against a training median of 0.748. Held-out cases scored with population-median pharmacodynamic parameters reach median $R^2 = -0.925$ with root mean squared error 15.505 BIS, and 71 of 100 are worse than predicting the case's mean BIS. The half-effect concentration spans a factor of 11.7 across the cohort, the Hill slope a factor of 11.3 and the maximal effect a factor of 3.4. Removing the slowly varying baseline lowers the training median R^2 from 0.748 to 0.437.

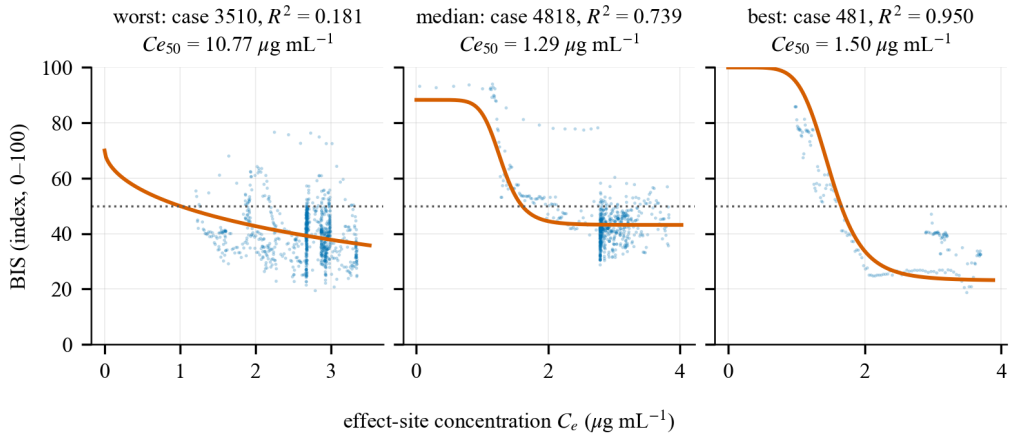

Figure 8: Three fitted patients from the 500 fitted cases, worst, median and best by goodness of fit (cases 3510, 4818 and 481; $R^2 = 0.181$, 0.739, 0.950). The worst case fails the filter and is not among the 50 evaluated. Fit quality tracks how much the anaesthetist varied the concentration rather than how well the model form works: when the concentration is held nearly steady, the data cannot identify how the patient would respond to a change.

G Digital twin: the realism ladder

Across the fifteen arms the marginal cost of observation noise ranges from $+0.218$ to $+2.161$ BIS, of drift from $+0.181$ to $+1.483$, of dropout treated as real data from $+2.117$ to $+4.196$, and of dropout with low-quality samples discarded from $+0.015$ to $+0.510$. The strongest single result is the last pair: dropout costs two to four BIS if the corrupted samples are kept and essentially nothing if they are discarded, which is why the discarding condition is the one we quote throughout.

H The field across budgets

The rule's rank on location error is fourth of the ten arms of Table 2 at T=30 with four initial queries, first with eight, and second at T=60 and T=120; restricted to attainable slices at its best cell it is second rather than first. We therefore do not quote a rank.

I Unattainability

A state is a post-initial query whose time slice has no attainable level, so the ground-truth location is clipped to a box edge. On the frozen surface this is a property of the patient, since the baseline is constant; on the drifting surface it varies within a run. The predicate and the metric are identical in both cases.

The preference term scored alone is classified by what it scores rather than by the name it shares, since it is behaviourally a greedy rule (Appendix A).

What the restriction changes. Restricting analysis to the 27 patients with a real crossing affects the two metrics differently: because regret is referenced to the attainable optimum, its gap is largely

Table 14: Mean instantaneous regret in BIS ± standard error at each realism condition, $n = 50$ patients per cell, drifting surface, T=30, four initial queries, folded objective. The final four columns give the marginal cost of each layer. Arms are ordered by regret at A; the two non-adaptive references are listed below the rule. Eleven of the fifteen folded-objective arms are shown; the other four (LogEI, Thompson sampling, knowledge gradient and a parity-weighted BOBA) fall inside the ranges quoted in the text. The signed-encoding arms were run under D′ only, so they have no ladder.

	Realism condition					Marginal cost of each layer			
arm	A: noise-free	B: +noise	C: +drift	D: +dropout kept	D′: +dropout discarded	B−A	C−B	D−C	D′−C
Decay-BO	5.892 ± 0.342	7.743 ± 0.313	9.038 ± 0.313	13.234 ± 0.460	9.288 ± 0.343	$+1.852$	$+1.295$	$+4.196$	$+0.250$
BOBA[pub]	7.229 ± 0.449	9.389 ± 0.415	10.872 ± 0.417	14.078 ± 0.529	11.029 ± 0.477	$+2.161$	$+1.483$	$+3.206$	$+0.157$
BOBA-OPT	7.399 ± 0.428	9.259 ± 0.405	10.311 ± 0.447	13.648 ± 0.488	10.820 ± 0.484	$+1.860$	$+1.051$	$+3.338$	$+0.510$
GP-UCB	7.659 ± 0.438	9.258 ± 0.352	10.371 ± 0.376	13.650 ± 0.460	10.757 ± 0.481	$+1.599$	$+1.113$	$+3.279$	$+0.386$
Decay-BO (no forgetting)	8.226 ± 0.461	9.094 ± 0.443	9.880 ± 0.470	13.149 ± 0.441	9.947 ± 0.460	$+0.868$	$+0.786$	$+3.269$	$+0.067$
EI	8.401 ± 0.485	9.965 ± 0.410	10.956 ± 0.409	13.986 ± 0.509	11.161 ± 0.459	$+1.565$	$+0.990$	$+3.030$	$+0.206$
AES	8.521 ± 0.490	10.492 ± 0.526	11.937 ± 0.539	14.973 ± 0.648	11.975 ± 0.587	$+1.971$	$+1.445$	$+3.036$	$+0.037$
R-UCB	8.523 ± 0.425	10.112 ± 0.454	10.883 ± 0.454	14.142 ± 0.503	11.042 ± 0.482	$+1.588$	$+0.771$	$+3.259$	$+0.160$
SW-UCB	8.619 ± 0.458	9.835 ± 0.434	11.074 ± 0.469	14.001 ± 0.480	11.089 ± 0.418	$+1.216$	$+1.239$	$+2.927$	$+0.015$
Fixed protocol	13.084 ± 1.094	13.302 ± 1.059	13.596 ± 1.038	15.774 ± 0.944	13.706 ± 1.031	$+0.218$	$+0.294$	$+2.178$	$+0.110$
Random	13.485 ± 0.760	13.945 ± 0.756	14.126 ± 0.711	16.243 ± 0.749	14.181 ± 0.715	$+0.460$	$+0.181$	$+2.117$	$+0.055$

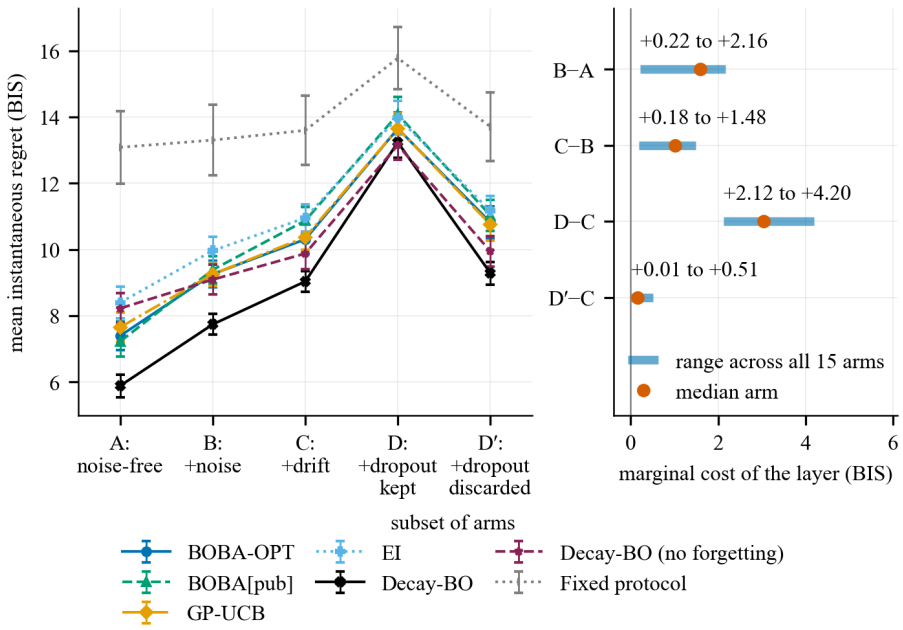

Figure 9: Left, regret as each source of randomness is added, for the arms of Table 14 except AES, R-UCB, SW-UCB and Random (omitted for legibility), with the fixed protocol in grey. Right, the cost of each layer across methods. Fitting to fabricated readings is the dominant cost; discarding them costs at most 0.5 BIS.

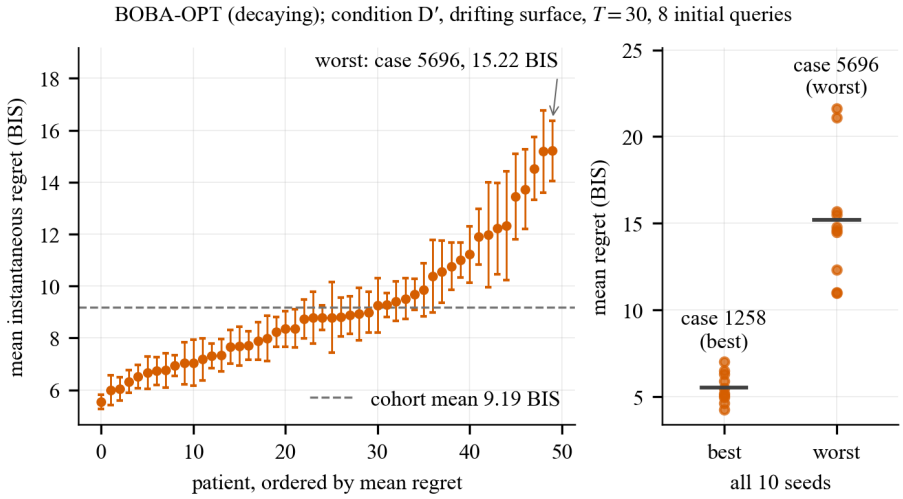

Figure 10: BOBA-OPT with the decaying schedule of Appendix J. Left, every simulated patient ordered by regret (cohort mean 9.19, median 8.79 BIS), $n = 50$ patients over ten seeds each, condition D′, T=30, eight initial queries; the worst patient reaches 15.22 BIS (case 5696). No method keeps every patient inside the adequate band. Right, the ten seeds behind the best and worst patients, showing that the worst-patient result is a property of the patient rather than of an unlucky run.

preserved, whereas location error relies on physical locations where clipping alters the target itself. Consequently, on the frozen surface with eight initial queries, all eleven of the proximity rule's regret comparisons survive restriction with identical signs and significance. On location error, however, its advantage over the three information-carrying arms is lost, though its losses to the Decay-BO pair persist unchanged on both metrics. On the drifting surface, every significance verdict survives

Table 15: Paired differences of GREEDY-H against each reference on the drifting surface, D′, positive favouring GREEDY-H, Holm corrected over the nine references in each cell. Entries without a marker are significant after correction.

reference	T=30, init 4	T=30, init 8	T=60	T=120
Fixed protocol, regret	+3.128	+2.331	+3.131	+4.784
Fixed protocol, distance	+0.0408 (n.s.)	+0.0738	+0.0820	+0.0886
Random, regret	+2.751	+3.767	+5.296	+5.454
Random, distance	+0.0701	+0.1237	+0.1338	+0.1364
GP-UCB, regret	+0.485	+0.854	+0.945	+1.007
GP-UCB, distance	+0.0236	+0.0229	+0.0194	+0.0113
EI, regret	+0.595	+1.451	+1.424	1.733
EI, distance	+0.0217	+0.0430	+0.0277	+0.0278
AES, regret	+1.313	+1.424	2.383	2.239
AES, distance	+0.0251 (n.s.)	+0.0512	+0.0389	+0.0444
Decay-BO (no forgetting), regret	−0.054 (n.s.)	−0.321 (n.s.)	+0.039 (n.s.)	+0.343
Decay-BO (no forgetting), distance	+0.0265	+0.0346	+0.0542	+0.0513
BOBA-LS, regret	+0.494 (n.s.)	−0.151 (n.s.)	+0.325 (n.s.)	+0.666
BOBA-LS, distance	−0.0119 (n.s.)	−0.0033 (n.s.)	+0.0014 (n.s.)	+0.0061 (n.s.)
BOBA-OPT, regret	+0.157 (n.s.)	+0.463 (n.s.)	+0.596	+0.832
BOBA-OPT, distance	−0.0136 (n.s.)	+0.0127 (n.s.)	+0.0086 (n.s.)	+0.0046 (n.s.)
Decay-BO, regret	−0.611 (n.s.)	−0.073 (n.s.)	−0.761	−1.142
Decay-BO, distance	−0.00003 (n.s.)	+0.0118 (n.s.)	−0.0068 (n.s.)	−0.0287

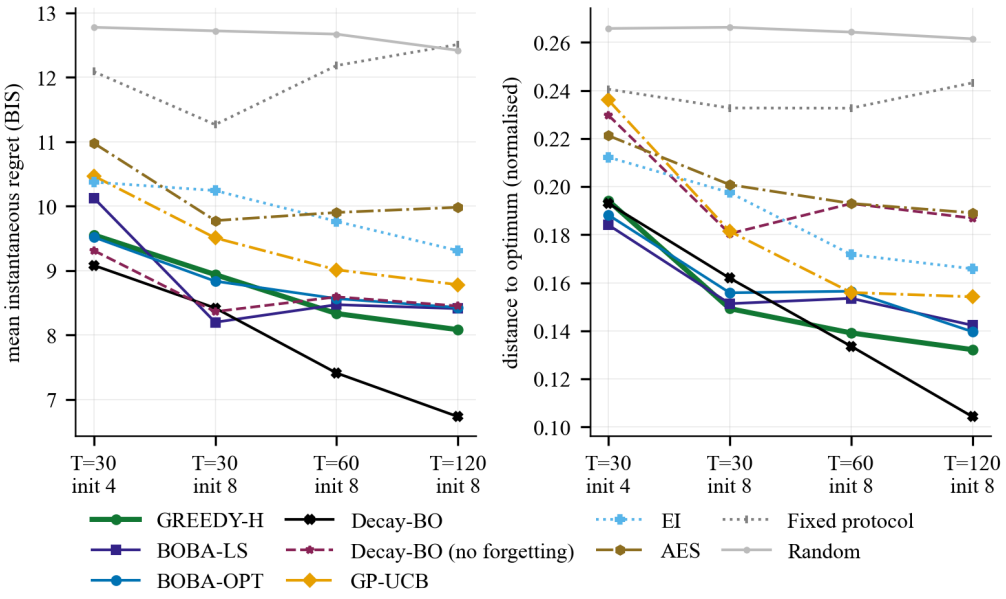

Figure 11: Median per-patient regret and median per-patient location error for the ten arms of Table 2 across four cells, drifting surface, D′, $n = 50$ patients over ten seeds per arm per cell.

restriction to attainable slices with identical signs, but top-rank ordering shifts: the proximity rule drops behind BOBA-LS, with the remaining three methods mutually indistinguishable.

J SCHEDULES AND SWEEPS

Schedules for the exploration weight. Two variants of BOBA-OPT change only how the weight evolves over a run. The decaying variant multiplies it by $1 - 0.9\,\tau$ with τ the fraction of the run completed, floored at 0.1; the floor and the linear shape were set in advance and not tuned. The warm-start variant runs expected improvement for the first third of the time axis and the decaying schedule thereafter. At the long budget the decaying variant beats plain BOBA-OPT (+0.456, Holm

Table 16: How often the target is unattainable, by cell and surface.

cell	surface	sessions	states	unattainable	patients wholly unattainable
init 8, T=30	drifting	8,000	150,471	5.11%	0
init 8, T=30	frozen	8,000	150,462	45.24%	23 of 50
init 4, T=30	drifting	6,000	136,796	5.67%	0
init 4, T=30	frozen	2,500	56,972	45.44%	23 of 50
init 8, T=120	drifting	5,000	496,186	6.50%	0

Table 17: Median degradation ratio, unattainable over attainable location error, where 1.0 means unaffected. Medians over 22 combinations of arm, cell and surface for the information-carrying class, 13 for the greedy class and 24 for the third. The frozen, four-initial-query cell ran only five arms, none in the third class, hence the dash.

arm class	$T$30/i8 drift	$T$30/i8 froz	$T$30/i4 drift	$T$30/i4 froz	$T$120/i8 drift	overall
information-carrying	2.25×	1.98×	1.86×	1.85×	1.85×	1.93×
greedy or proximity	1.28×	1.05×	1.20×	0.69×	1.37×	1.09×
UCB, dynamic or other	1.53×	1.12×	1.17×	—	1.34×	1.38×

$p = 3.7\times10^{-5}$), the warm start ($+0.222$, $p = 0.0038$), GP-UCB ($+0.590$, $p = 0.0012$) and expected improvement ($+0.910$, $p = 8.1 \times 10^{-8}$), and is level with Decay-BO (no forgetting). The warm start never pays: against the decaying schedule it is inert at the short budget and worse at the long one. This is the same contrast as above; corrected within the warm start's own family of five comparisons its Holm p is 0.0076. Both variants exist on the twin at eight initial queries only and have not been run on the synthetic suite.

Preference width. Sweeping the preference width σ_p over $\{0.25, 0.5, 1, 2, 5, 10\}$ BIS on the full fifty-patient design, the predictive standard deviation on the signed objective is about 11.4 BIS, so the whole sweep sits below it. Location error improves from 0.1606 at a width of 10 to 0.1385 at a width of 2 and then saturates, varying by 0.0018 over the remaining eightfold range; regret shows no trend and is best at a width of 5. All twelve contrasts against the proximity rule are indistinguishable after Holm correction, and the difference is flat in the width rather than growing as it falls. The obvious explanation for the null is ruled out: the predictive standard deviation varies across the candidate set by about its own mean, with a median relative spread of 0.961 and only 16.2% of states below 0.10, so the term the two rules differ by is present and still does not reach the outcome.

The proximity family. Scoring candidates by $\kappa\sigma(x) - |\mu_h(x) - Y^\star|$, which is the proximity rule at $\kappa = 0$ and the straddle heuristic at $\kappa = 2$, and sweeping $\kappa \in \{0, 0.5, 1, 2, 4\}$ on the reduced design, regret rises overall from 8.135 at zero to 12.766 at four, and location error from 0.1508 to 0.2396. The best point is the boundary at zero, which is the proximity rule itself, and BOBA-LS sits there rather than anywhere on the interior of the family. This is consistent with the algebra of §4: the risk term's uncertainty part has an interior maximum at the preference width rather than rising without bound, so it is uncertainty-averse in the regime these runs occupy.

K CHOICE OF METRIC

Mean instantaneous regret and mean running-best value select a different best method in 40.1% of 162 comparisons over 63,419 sessions, and in 42.6% when the same session set is scored both ways, so the two measure different things rather than the same thing noisily. Holding the session set fixed is the version in which the metric is the only thing that differs. In a dynamic problem a running-best value can credit a method for a value it can no longer reach: 31.0% of running-best value at the short budget, and 17.9% at the long budget, was attained at points the optimum had since moved away from. The two metrics also choose a different best exploration weight in 13.0% of cells.

Rank correlations between cumulative regret, simple regret and best-so-far across the five experimental cells range from $+0.000$ to $+0.895$, and the largest rank movements are substantial: Decay-BO (no forgetting) moves from third on the primary metric to twelfth on simple regret in one cell, and

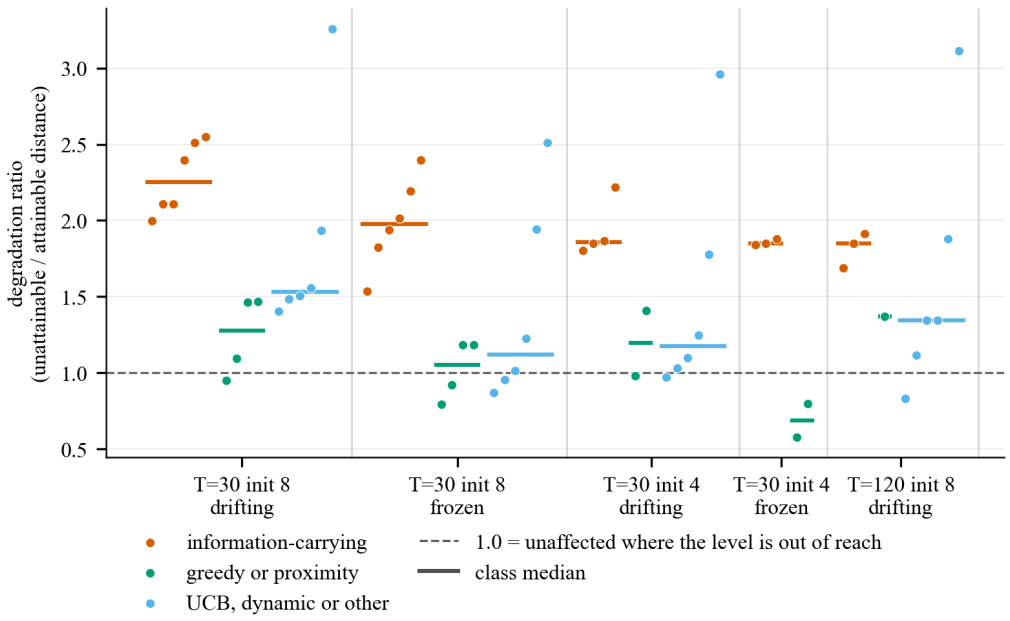

Figure 12: Per-arm degradation ratio grouped by arm class, over the five combinations of cell and surface containing both kinds of state. Bars are class medians. A ratio needs a patient with both attainable and unattainable states, so n ranges from 18 to 26 patients per arm depending on the cell.

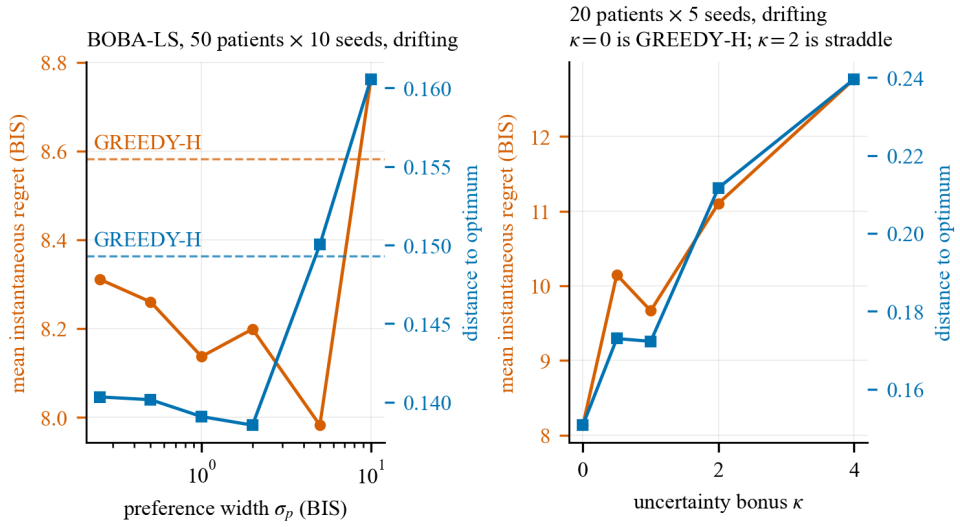

Figure 13: Left, the two metrics against the preference width σ_p, $n = 50$ patients over ten seeds, with the within-cell proximity rule marked. Right, the proximity family against the uncertainty bonus κ, $n = 20$ patients over five seeds. The two panels share a visual grammar but not an axis, since the two parameters are different quantities.

GP-UCB from second to eleventh in another. Reporting a single metric without saying which would therefore change the paper's ordering, which is why every comparison here names its metric.