Attributes: Selective Learning and Influence*

Arjada Bardhi †

April 15, 2020

Abstract

When different stages of the evaluation of a multi-attribute project rest with conflicting economic actors, which attributes are selectively explored and why? We provide a model of attribute sampling in which correlation across attributes is flexibly modeled through Gaussian processes. In the absence of conflict, the optimal sample of attributes maximizes informativeness by balancing out-of-sample extrapolation with correlation within the sample. It depends neither on the prior value of the project nor on the format of sampling. Agency conflict, in contrast, gives rise to distortions. Sampling serves a dual purpose of generating valuable information and influencing the co-player. When influence takes priority, optimal sampling either suppresses informativeness for both players or negatively correlates their interests. Casting site selection as an attribute problem, our framework provides a theoretical rationale for site selection bias in small-scale program evaluation.

Keywords: correlated attributes, Gaussian sample paths, sampling capacity, agency conflict, sample centrality, site selection bias

JEL classification: D83, D81, D72, D04

*The latter stages of this project were supported by NSF Award SES-1919329. I am indebted to my advisors, Asher Wolinsky, Bruno Strulovici, Yingni Guo, and Georgy Egorov, for their formative feedback and continued support. The project has benefited from discussions with Attila Ambrus, Nemanja Antic, Nina Bobkova, Steve Callander, Alex Frankel, Rachel Kranton, R. Vijay Krishna, Meg Meyer, Xiaosheng Mu, Alessandro Pavan, Andy Postlewaite, Debraj Ray, Todd Sarver, Takuo Sugaya, and Andy Skrypacz, as well as seminar audiences at UCSD, Stanford GSB, FSU, Cambridge, UCL, YorkU, USC, Duke, NYU, UPenn, Collegio Carlo Alberto, Bocconi, EUI, UChicago Booth, Oxford, Northwestern, NASMES, CETC, SEA Meetings, 29th Stony Brook GT Conference, Columbia, and NYU Stern. This version generalizes and thereby subsumes earlier versions.

†Department of Economics, Duke University. Email: arjada.bardhi@duke.edu.

1 Introduction

Important decisions rely on selective exploration of objects with multiple attributes: from potential buyers appraising complex products to employers evaluating the diverse skills of prospective employees and policymakers gauging the spatial impact of social programs. More often than not, such exploration is undermined by agency conflict: the party deciding which attributes to explore differs from the party that translates the findings into a decision. Understanding the nature of such selective and decentralized exploration – which attributes are optimally explored and why – has been of long-standing interest in economics (Lancaster (1966), MacCrimmon (1968), Keeney and Raiffa (1976)). An important feature that is shared by these examples but is absent from most of the literature on attribute learning is that attributes are correlated. Thus, evaluation hinges on non-trivial extrapolation from explored attributes to those left unexplored.

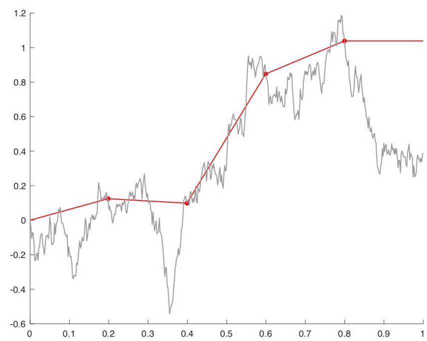
Our model features a principal and an agent who jointly evaluate an uncertain project consisting of a multitude of attributes. Players have separate authorities over evaluation: the agent decides which attributes to sample subject to a sampling capacity, whereas the principal decides whether to adopt the project. Attributes are observed perfectly if sampled, and players are symmetrically informed at all times. Conflict between players is therefore one of interest rather than information: players disagree on the weighing of different attributes and the value of the status quo.

The analysis in the paper is twofold. First, we characterize the subset of attributes that is optimally sampled in the benchmark of no agency conflict. The optimal sample maximizes a natural informativeness statistic that summarizes the extent to which sample attributes are informative about out-of-sample attributes and generate overlapping information within the sample. Second, we characterize sampling distortions that arise due to agency conflict. To address these aims, a novel and flexible framework for attribute sampling is introduced.

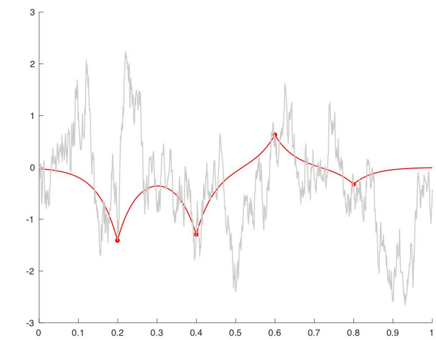
The project is characterized by the uncertain realizations of a mass of attributes $\mathcal{A}$.[1] Attribute realizations are assumed to follow an unknown mapping, drawn randomly from the space of sample paths of a Gaussian process. The process is pinned down by two commonly known parameters: an *attribute mean* function $\mu : \mathcal{A} \to \mathbb{R}$ and an *attribute covariance* function $\sigma : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$. Covariance σ introduces a natural similarity metric across attributes. The stronger the covariance between attributes a and a', the stronger is the inference drawn from one attribute to the other and therefore the more similar the

[1] For example, when evaluating the skill bundle of a job candidate, "writing skills" and "time-management" would be *attributes*, whereas the candidate's proficiency in each skill would be her *attribute realizations*.

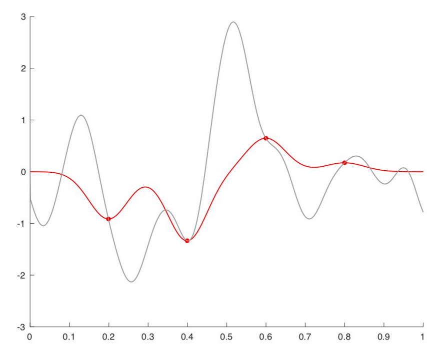
two attributes are. This setup is both general and tractable: it covers a vast array of attribute mappings and extrapolation patterns. Figure 1 illustrates extrapolation across a small sample of attributes for four such Gaussian processes.



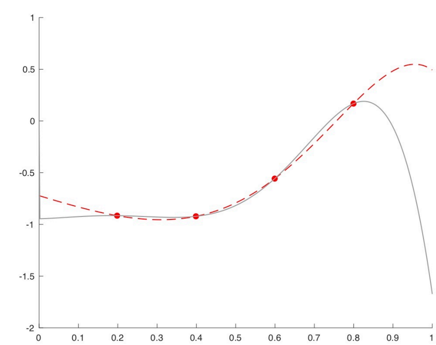
(a) Standard Brownian: $\sigma_{Br}(a, a') = \min(a, a')$



(b) Ornstein-Uhlenbeck: $\sigma_{OU}(a, a') = e^{-20|a-a'|}$



(c) Squared exponential: $\sigma_{SE}(a, a') = e^{-400(a-a')^2}$



(d) Polynomial: $\sigma_{pol}(a, a') = (1 + aa')^{10}$

Figure 1: The randomly drawn attribute mapping is in grey and the extrapolated mapping is in red. For all plots, $\mathcal{A} = [0, 1]$, $\mu(\cdot) = 0$ *and sample of attributes* $\mathbf{a} = (1/5, 2/5, 3/5, 4/5)$.

The single-player benchmark consists of a pure learning problem. Each sample is succinctly summarized by the *posterior variance* that the sample induces on the player's expected value from the project. The higher this posterior variance, the more informative the sample is for adoption. But due to correlation across attributes, the informativeness of a sample is potentially complex: it hinges on both generalizability to out-of-sample attributes and non-redundancy within the sample. Posterior variance elegantly encodes these considerations into a single tractable statistic (Theorem 3.1).

Moreover, the only building blocks of posterior variance are the covariance function and attribute weights. Hence, the optimal single-player sample depends neither on the ex-ante value of the project nor on attribute means or the player's outside option. What is more, the player samples in the same way under simultaneous sampling as under sequential sampling.

Both observations are a consequence of the Gaussian structure across attributes. Taken together, they teach us that a single player would evaluate any project – promising or not relative to the status quo, bound or not to simultaneous draws – informed by the same set of attributes. Once agency conflict makes an appearance, this is no longer the case.

In the principal-agent setting, sampling serves a dual purpose for the agent: it controls both the sheer probability of adoption and the extent to which adoption is aligned with the agent's posterior value. Correspondingly, Theorem 4.1 shows that the agent's payoff from a given sample depends on a pair of sufficient statistics: (i) the sample's informativeness for the principal, and (ii) its informativeness for the agent, adjusted to reflect his lack of adoption authority.[2] Importantly, this characterization reduces optimization over all feasible samples to optimization over a two-dimensional space.

This characterization enables us to explore two striking sampling distortions: suppression of informativeness for both players and controversial sampling. The first consists in the optimal choice of a sample that can be strictly improved in its informativeness for both players, whereas the second in a sample that correlates *negatively* the players' expected values for the project. We argue that both are manifestations of the same strategic concern by the agent. They arise when the principal's prior value is sufficiently close to her outside option relative to the agent's – so that there is more room for persuasion through sampling. Under such circumstances, the agent sees sampling primarily as a means to controlling the adoption probability rather than collecting useful information. Which of the two distortions arises depends on whether players are in prior agreement about the project. Section 4.4 establishes these distortions for general attribute weights and general covariance function.

Section 5 presents an application to optimal site selection in impact evaluation. The low generalizability of evaluation findings due to strategic site selection has recently been the focus of a growing empirical literature (Allcott (2015), Bold et al. (2018), Vivalt (Forthcoming)). We provide a theoretical rationale for this observation by formulating optimal site selection as an attribute-sampling problem between a utilitarian researcher and a partisan evaluator.[3] Ideally the researcher would place the pilot studies at sites with the greatest external validity.[4] But in the presence of the partisan evaluator, he suppresses information

[2]Statistic (i) corresponds to the principal's posterior variance, whereas (ii) corresponds to the portion of the agent's posterior variance that is explained by the principal's posterior variance.

[3]That is, we consider a social program evaluated for full-scale implementation across a set of target sites, the program outcomes of which are correlated. A utilitarian researcher has a budget of k pilot studies to place in select target sites, whereas a partisan evaluator interested in the outcome at a single site translates pilot findings into a decision about full-scale implementation.

[4]This is akin to the notion of *purposive site selection* in impact evaluation (Olsen et al. (2013)).

by selecting peripheral sites deemed to be of low external validity by both.

Overall, the analysis makes two methodological contributions. First, the Gaussian path approach that we introduce for modeling correlated attributes promises to be useful to problems beyond attribute sampling. A byproduct of our analysis is the generalization of the Brownian path approach introduced in experimentation and search literatures by Callander (2011) and Jovanovic and Rob (1990).[5] Second, section 6 presents a novel graph-theoretic approach to the attribute problem to complement our characterization of single-player sampling. We construct and interpret a natural centrality measure in the attribute graph, namely *sample centrality*, which is a generalization of betweenness centrality to non-singleton subsets of nodes (Freeman (1977), Everett and Borgatti (1999)). In doing so, sample centrality departs from most existing network centrality measures, typically defined only over single nodes.

The rest of this section discusses related work. Section 2 sets up the model and establishes preliminary results on extrapolation. Section 3 characterizes optimal sampling in the single-player benchmark, while section 3.2 illustrates the characterization. The principal-agent analysis may be found in section 4. Section 5 develops the application to optimal site selection, while section 6 presents a graph-theoretic approach to the attribute problem.

Related literature. First and foremost, the paper builds on models of costly attribute discovery. Klabjan, Olszewski and Wolinsky (2014) study a setting with finitely many independent attributes that can be learned perfectly at a cost. They establish that if attributes are ordered by second-order stochastic dominance and are equally costly, the optimal sample consists of dominated attributes. Our model differs from theirs in that we allow for a continuum of correlated attributes while restricting attention to jointly Gaussian distribution. Yet theorem 3.1 draws a common thread, as it shows that the distribution over posterior values induced by the single-player sample is SOSD-dominated.[6] Other related work combines attribute sampling with search across several multi-attribute objects (Sanjurjo (2017), Olszewski and Wolinsky (2016), Geng, Pejsachowicz and Richter (2017), Neeman (1995)). In contrast, our focus is on learning from a single multi-attribute object.

Liang, Mu and Syrgkanis (2020) study dynamic and noisy learning of a finite number of correlated attributes. Both our model and theirs leverage the tractability of Gaussian

[5] Outside of economics, Gaussian processes have been essential to spatial inference in geostatistics, also known as kriging (Matheron (1963), Chilés and Delfiner (2012)). They are also increasingly used in kernel methods in machine learning (Rasmussen and Williams (2006), Hofmann, Schölkopf and Smola (2008)).

[6] Due to correlation, attribute realizations are not necessarily SOSD-ordered. But the distributions over posterior values induced by attribute samples are.

correlation. There are, however, two key differences. First, our focus is on instantaneous learning from a continuum of attributes, whereas their model works with a finite number of attributes. As the attribute space becomes arbitrarily large, their main characterization holds only if attributes are approximately independent. Second, we study the implications of agency conflict for optimal sampling of attributes. Our focus on static learning from an infinite set of correlated information sources differentiates our model from dynamic learning models in Moscarini and Smith (2001), Fudenberg, Strack and Strzalecki (2018), Ke and Villas-Boas (2019), Mayskaya (2019) and Che and Mierendorff (2019).

Starting with Aghion et al. (1991), a large literature studies selective learning of an unknown payoff mapping. A productive technique in this literature has been to model this payoff mapping as a Brownian sample path. Jovanovic and Rob (1990) model gradual technological discovery through an infinite family of independent Brownian paths. Callander (2011) studies optimal strategic experimentation with correlated bandit arms modeled as the domain of a Brownian path. Garfagnini and Strulovici (2016) introduce costly learning and forward-looking agents in this experimentation setup, whereas Callander and Hummel (2014) explore forward-looking sequential experimentation with two players. Our model departs from this body of work in two respects. First, we work with a large class of Gaussian processes, where Brownian motion is but a special case.[7] Second, the payoff-relevant statistic here is the area under the sample path rather than a maximum of the path – this crucial feature differentiates attribute discovery from search. In this respect our model relates to Ilut and Vlachev (2017) and Callander and Clark (2017), in which players seek to learn the entire sample path. In Ilut and Vlachev (2017), the problem is not one of sampling: the agent merely selects amount of observational noise at exogenously drawn states.[8] In Callander and Hummel (2014) the Higher Court samples legal cases, but the nature of the adoption problem faced by the Lower Court differentiates this model from attribute learning.[9]

Our principal-agent sampling is related to models of persuasion through flexible experimental design (Brocas and Carrillo (2007), Rayo and Segal (2010), Kamenica and Gentzkow (2011)), Wald persuasion games (Henry and Ottaviani (2019)), and persuasion with multi-dimensional information (Glazer and Rubinstein (2004), Sher (2011)). Our setup contrasts

[7]In an earlier version (Bardhi, 2018), attribute realizations followed a Brownian path. The current framework strictly nests this special case, treated in depth in online appendix F.1.

[8]Ilut and Vlachev (2017) are a notable exception in that they model the payoff mapping through a squared exponential Gaussian path rather than the traditional Brownian path.

[9]In Callander and Hummel (2014) optimal case selection does not consist of the most uncertain legal case generically – similarly, the most uncertain attribute is not the most informative in section 3.

with such work in that (i) attribute sampling imposes endogenous constraints on the set of feasible information structures, and (ii) both players have attribute-dependent preferences.

Our application to site selection in section 5 shares with Di Tillio, Ottaviani and Sørensen (2017*a*) and Di Tillio, Ottaviani and Sørensen (2017*b*) the interest in strategic sample selection and its implications for the external validity of experiments. In contrast to our setting, the agent in Di Tillio, Ottaviani and Sørensen (2017*a*) and Di Tillio, Ottaviani and Sørensen (2017*b*) has state-independent preferences and discloses privately observed signals.[10] In contrast to Banerjee et al. (2017), our framework features a single principal and there is no value from randomization over samples. Our analysis also relates to Hirsch (2016) in the agent's dual goals of learning and motivating the principal, although disagreement here arises due to different attribute weighting rather than heterogenous beliefs.

2 Model

2.1 Setup

Players. A principal (P, she) and an agent (A, he) jointly evaluate a multi-attribute project of uncertain quality. The agent decides which attributes to sample. The principal observes the realizations of sample attributes and decides whether to adopt the project.

Attributes. The project consists of a mass of attributes $\mathcal{A} := [0,1]^d$ where $d \geq 1$. The realization of attribute $a \in \mathcal{A}$ is denoted by $f(a) \in \mathbb{R}$. Attribute realizations follow an unknown mapping $f: \mathcal{A} \to \mathbb{R}$, drawn from the space of sample paths of a Gaussian process (GP) with *prior mean function* $\mu : \mathcal{A} \to \mathbb{R}$ and symmetric positive semidefinite *covariance function* $\sigma : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$.[11] Per standard notation, the distribution over mappings is

$$f \sim \mathcal{GP}(\mu, \sigma). \tag{1}$$

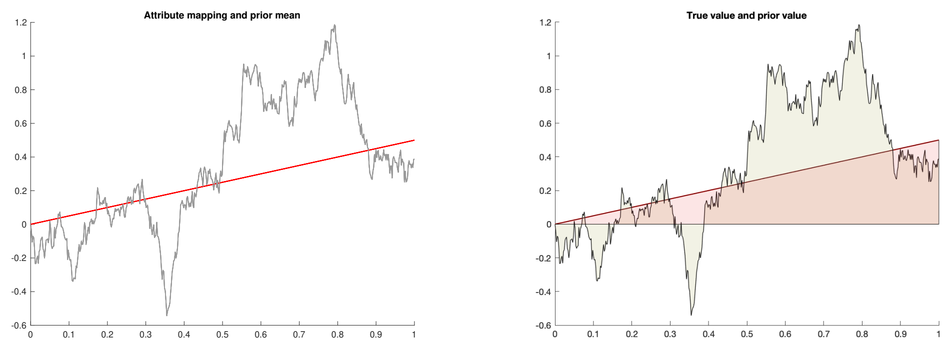
Parameters (μ, σ) are commonly known and fully pin down the distribution of f. The prior mean $\mu(a)$ specifies the expected realization of attribute $a \in \mathcal{A}$, whereas $\sigma(a, a')$ specifies the covariance between $f(a)$ and $f(a')$ for any attribute pair $a, a' \in \mathcal{A}$. Correspondingly, $\sigma(a, a)$ is the variance of $f(a)$. Figure 2a depicts the attribute mapping and the prior mean for a familiar one-dimensional Gaussian process – namely, the Brownian motion. The following assumption imposes a regularity condition on the prior distribution of f.

[10] Of the two, Di Tillio, Ottaviani and Sørensen (2017*b*) is more closely related as there the state is multi-dimensional and the agent designs an experiment on his dimension of choice. Yet again, selective sampling in their setup arises out of researcher's access to private information.

[11] Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a stochastic process $f := \{f(a,\omega)\}_{a\in\mathcal{A},\omega\in\Omega}$ is a GP if and only if $(f(a_1), \ldots, f(a_n))$ is jointly Gaussian for any $a_1, \ldots, a_n \in \mathcal{A}$ and $n \geq 1$. Appendix A.1 provides technical preliminaries on GPs. For a comprehensive introduction, see Rasmussen and Williams (2006).

Assumption 1 (Sample-path continuity)**.** *Almost surely any realization of f is continuous.*

Assumption 1 implies that, for any realization of f, attributes that are close in the attribute space have similar realizations. Importantly, this implies that μ and σ are both continuous.[12] That is, for any two attributes arbitrarily close in $\mathcal{A}$, their respective realizations are almost perfectly correlated and arbitrarily close both ex-post and in expectation.



(a) Realized attribute mapping f in grey and prior mean $\mu(a) = 2a$ in red.

(b) Yellow (resp., red) area depicts a player's true value v_i (resp., prior value ν_0^i).

Figure 2: For the depicted Brownian motion, $\sigma(a, a') = \min(a, a')$ for all $a, a' \in \mathcal{A} = [0, 1]$. The right panel 2b assumes that $\omega_i(a) = 1$ for all $a \in [0, 1]$.

Sampling. Because f is drawn from the space of Gaussian sample paths, the realizations of a sample of k attributes $\mathbf{a} = (a_1, \ldots, a_k)$ are jointly Gaussian, that is

$$f(\mathbf{a}) := \begin{pmatrix} f(a_1) \\ \vdots \\ f(a_k) \end{pmatrix} \sim \mathcal{N}\left(\underbrace{\begin{pmatrix} \mu(a_1) \\ \vdots \\ \mu(a_k) \end{pmatrix}}_{\mu(\mathbf{a})}, \underbrace{\begin{pmatrix} \sigma(a_1, a_1) & \ldots & \sigma(a_1, a_k) \\ \vdots & \ddots & \vdots \\ \sigma(a_k, a_1) & \ldots & \sigma(a_k, a_k) \end{pmatrix}}_{\Sigma(\mathbf{a})} \right).$$

Let $\mu(\mathbf{a})$ and $\Sigma(\mathbf{a})$ denote the sample mean and covariance matrix respectively. With some notational abuse, let $a \in \mathbf{a}$ denote a sample attribute. The acquired sample $(\mathbf{a}, f(\mathbf{a}))$ is observed publicly and players are symmetrically informed throughout the game. Fixing k, it is without loss to restrict attention to the set of *non-redundant samples* of size at most k, as any redundant attributes can be dropped from the sample:

$$\mathcal{A}_k := \left\{(a_1, \ldots, a_n) \in \mathcal{A}^n,\ \forall n \leq k, n \in \mathbb{N} \mid \Sigma((a_1, \ldots, a_n)) \text{ is non-singular}\right\}.^{13}$$

[12] Proposition A.1 provides sufficient conditions on (μ, σ) to guarantee sample-path continuity. On the other hand, Proposition A.2 shows that sample-path continuity implies continuity of μ and σ.

[13] As is standard, covariance matrix $\Sigma(\mathbf{a})$ is non-singular if and only if there exists no vector $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $\mathrm{Var}\left(\sum_i x_i f(a_i)\right) = 0$. In particular, there exists no $a \in \mathbf{a}$ such that $\sigma(a, a) = 0$.

The agent samples attributes subject to sampling capacity $k \in \mathbb{N}$. The cost of drawing n attributes is zero if $n \leq k$ and $+\infty$ otherwise. We contrast two sampling formats – *simultaneous* and *sequential* – even though our main focus is on simultaneous sampling. Under simultaneous sampling, agent draws the entire sample $\mathbf{a}$ before $f(\mathbf{a})$ is observed. Under sequential sampling, attributes are drawn sequentially in batches, with attribute realizations in batch j observed before batch $(j+1)$ is drawn. There is no cost of delay and the principal decides after the agent concludes all sampling.

Payoffs. The value from adoption for player i is a weighted sum of attribute realizations

$$v_i = \int_{\mathcal{A}} f(a)\omega_i(a)\,\mathrm{d}a, \tag{2}$$

where $\omega_i : \mathcal{A} \to \mathbb{R}$ is a Lebesgue-integrable attribute weight function for player i. Generically, for each player i the sum of attribute weights can be normalized to one.[14] Weight functions (ω_A, ω_P) are commonly known. The value from rejection for player i is a known outside option $r_i \in \mathbb{R}$. Hence there are two sources of inter-player conflict: (i) conflict about the *relative* importance of attributes, and (ii) conflict about *outside options*.

Due to assumption 1 and the Lebesgue-integrability of ω_i, the true adoption value v_i and its distribution are well-defined. At the start of the game, the true adoption value for player i is distributed according to

$$v_i \sim \mathcal{N}\left(\underbrace{\int_{\mathcal{A}} \mu(a)\omega_i(a)\,\mathrm{d}a}_{:=\nu_0^i},\ \int_{\mathcal{A}}\int_{\mathcal{A}} \sigma(a,a')\omega_i(a)\omega_i(a')\,\mathrm{d}a\,\mathrm{d}a'\right). \tag{3}$$

Let $\nu_0^i = \mathbb{E}[v_i]$ denote i's prior (expected) value from adoption. Figure 2b illustrates ν_0^i as the area under prior mean μ. Similarly, let $\nu^i(\mathbf{a}, f(\mathbf{a})) = \mathbb{E}\left[v_i \mid \mathbf{a}, f(\mathbf{a})\right]$ denote the posterior expected value for player i given sample $(\mathbf{a}, f(\mathbf{a}))$.[15]

Assumption 2 (Underdetermination by data)**.** *For any* $\mathbf{a} \in \mathcal{A}_k$*, any* $f(\mathbf{a})$*, and at least some player* $i \in \{A, P\}$*,* $Var(v_i \mid \mathbf{a}, f(\mathbf{a})) > 0$.

Assumption 2 ensures that the sampling problem is non-trivial: capacity k is insufficient to fully learn the adoption values for both players. For a simple violation, consider $\mathcal{A} = [0,1]$, $\sigma(a,a') = (1 + aa')$, and $\omega_i(a) = 1$ for all $a \in [0,1]$. In this case any attribute mapping is linear; therefore player i can fully learn his value from the project with $k = 1$.[16]

[14] See lemma E.1. In the single-player benchmark, it is without loss to have $\omega_i(a) \geq 0$ as well. The covariance function is redefined to reflect the normalization to positive attribute weights. Hence ω_i can be reinterpreted as a probability density over attributes, which proves useful in section 6.

[15] For brevity, we often refer to ν_0^i and $\nu^i(\mathbf{a}, f(\mathbf{a}))$ as *prior* value and *posterior* value, respectively.

[16] See online appendix E.4 for calculations.

Finally, players are said to be in *prior disagreement* if $\left(v_0^P - r_P\right)$ and $\left(v_0^A - r_A\right)$ have opposite signs. They are in *prior agreement* otherwise.

2.2 Inference and extrapolation

Player i's inferential problem proceeds in two steps: he first extrapolates from sample realizations to the rest of the attribute space, and then derives his posterior value from the extrapolated mapping. Lemma 2.1 establishes that players' best guess for an out-of-sample attribute is a linear combination of sample realizations. Because sample realizations are observed without noise, the extrapolated mapping traverses these realizations.

Lemma 2.1 (Extrapolation). *Fix sample $\mathbf{a} = (a_1, \dots, a_n) \in \mathcal{A}_k$, realizations $f(\mathbf{a}) \in \mathbb{R}^k$ and attribute $\hat{a} \in \mathcal{A}$. The expected realization $f(\hat{a})$ is given by*

$$\mathbb{E}[f(\hat{a}) \mid \mathbf{a}, f(\mathbf{a})] = \mu(\hat{a}) + \sum_{j=1}^{k} \tau_j(\hat{a}; \mathbf{a}) \left(f(a_j) - \mu(a_j)\right), \tag{4}$$

where $\tau_j(\hat{a}; \mathbf{a})$ is the $(1, j)^{th}$ entry of the vector $\begin{pmatrix}\sigma(a_1, \hat{a}) & \dots & \sigma(a_k, \hat{a})\end{pmatrix} [\Sigma(\mathbf{a})]^{-1}$. For any $j = 1, \dots, k$ and any $m \neq j$, $\tau_j(a_j; \mathbf{a}) = 1$ and $\tau_m(a_j; \mathbf{a}) = 0$.

Simple as it may seem, equation (4) is quite general. In particular, this equation generalizes the Brownian-bridge approach used in Callander (2011). The exact shape of extrapolation – and relatedly, the expressions for sample weights $(\tau_1, \dots, \tau_k)$ – are dictated by the parametric form of the covariance function. Figure 1 in the introduction illustrated four such extrapolation patterns. For instance, certain covariances give rise to *local* extrapolation: the expectation for each attribute is based only on its nearest sample neighbors. The Brownian path in Figure 1a and the Ornstein-Uhlenbeck path in Figure 1b are two such examples.[17] In contrast, Figure 1c and Figure 1d correspond to non-local extrapolation: sample attributes both near and far are informative.

Lemma 2.2 shows that the posterior value is also a linear combination of sample realizations. The sensitivity of the posterior value to each sample realization aggregates the sensitivity of the entire extrapolated mapping to it. Players agree on how to extrapolate but hold generically different posterior values due to different attribute weights.

Lemma 2.2 (Posterior expected value). *Fix sample $\mathbf{a} = (a_1, \dots, a_k) \in \mathcal{A}_k$ and $f(\mathbf{a}) \in \mathbb{R}^k$. Player i's posterior expected value is a linear combination of sample realizations, i.e.*

[17] Local extrapolation arises more generally for any Markov GP with $d = 1$. This is because for any such process and any sample of size $k \geq 3$, $\Sigma^{-1}(\mathbf{a})$ is tridiagonal (cf. Theorem 1 in Ding and Zhang (2018)).

$$\nu^i(\mathbf{a}, f(\mathbf{a})) = \nu_0^i + \sum_{j=1}^{k} \tau_j^i(\mathbf{a}) \left(f(a_j) - \mu(a_j)\right) \tag{5}$$

where realization $f(a_j)$ is weighted by

$$\tau_j^i(\mathbf{a}) := \int_{\mathcal{A}} \tau_j(a; \mathbf{a}) \omega_i(a) \, \mathrm{d}a \tag{6}$$

and $\tau_j(a; \mathbf{a})$ is as in lemma 2.1.

2.3 Remarks on the model

Covariance as similarity of attributes. Covariance σ can be interpreted as a *similarity metric* over attributes: the more similar two attributes are, the more accurately the realization of one attribute can be predicted from that of the other. The choice of covariance, therefore, determines the nature of the multi-attribute project under evaluation: the salience of certain attributes, the varying uncertainty across attributes, as well as the complexity of the project.

The model imposes two assumptions on attribute correlation: (i) jointly Gaussian distribution, and (ii) continuity of f. The latter guarantees that the project's value is well-defined and that the attribute space is rich, insofar as for any attribute there is another arbitrarily similar to it. The Gaussian structure, on the other hand, is to be understood as reflecting players' maximal ignorance regarding the project. That is, for any subset of attributes with a corresponding mean and covariance matrix, the multivariate Gaussian distribution is the distribution that maximizes entropy.

Cost structure. Capacity cost implies that the single-player problem is one of exploration rather than exploitation: how to best learn f with k select attributes. It is in the principal-agent interaction that the exploitation motive becomes important. For example, the agent samples attributes of low uncertainty or decides to forgo part of the capacity altogether in order to better influence the principal.

Noisy observations. Because our focus is on which attributes should be learned from a large attribute space rather than how well each attribute is to be learned, we assume zero observational noise. Yet the model can be extended to include Gaussian noise. That is, players observe $y(a) = f(a) + \epsilon(a)$ where $\epsilon(a) \sim \mathcal{N}\left(\mu^0(a), \eta^2(a)\right)$. Although optimal sampling is affected by the presence of noise, the single-player equivalence between sequential and simultaneous sampling still holds (Corollary E.2).[18]

[18] Also, example E.1 illustrates that with Brownian covariance greater observational noise steers the player towards attributes that are more uncertain ex ante.

3 Single-player sampling

This section establishes a benchmark for optimal sampling in the absence of conflict: sampling and adoption are decided by the same player. Section 3.1 characterizes single-player sampling for any prior mean and covariance and establishes an equivalence between simultaneous and sequential sampling. Section 3.2 demonstrates the tractability of this characterization by applying it to a class of distance-based covariance functions.

3.1 General characterization

The sample choice determines the distribution of posterior expected value $\nu(\cdot, f(\cdot))$.[19] From lemma 2.2, this posterior value follows a Gaussian distribution for any $\mathbf{a} \in \mathcal{A}_k$, that is,

$$\nu(\mathbf{a}, f(\mathbf{a})) \sim \mathcal{N}\left(\nu_0, \psi^2(\mathbf{a})\right),$$

where ψ^2 is the *posterior variance* induced by the sample. As expected from Bayesian reasoning, for any sample $\mathbf{a}$ the mean of the posterior value is ν_0. Therefore, ψ^2 provides a sufficient informativeness statistic based on which the agent ranks feasible samples. A higher posterior variance implies a more variable adoption decision, which in turn corresponds to a better informed adoption decision.

The player adopts the project if and only if the realized posterior value is above his outside option. His payoff from sample $\mathbf{a}$ is

$$V(\mathbf{a}) := \mathbb{E}\left[\max\left\{r, \nu(\mathbf{a}, f(\mathbf{a}))\right\}\right] = r + (\nu_0 - r)\Phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right) + \psi(\mathbf{a})\phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right), \tag{7}$$

where ϕ and Φ denote the pdf and cdf of the standard Gaussian distribution. The player's payoff depends on the sample only through the posterior variance that the sample induces: $V(\mathbf{a})$ is strictly increasing and convex in $\psi(\mathbf{a})$. The higher the posterior variance of a sample, the more informative the sample is and the lower is the residual uncertainty about the project's value and the frequency of type I and type II errors.[20] Because a more informative sample better tailors adoption to the true value of the project, it also leads to a higher expected value of an adopted project $\mathbb{E}\left[\nu(\mathbf{a}, f(\mathbf{a})) \mid \nu(\mathbf{a}, f(\mathbf{a})) \geq r\right]$. On the other hand, the frequency of adoption decreases in posterior variance if the prior value favors adoption (i.e. $\nu_0 \geq r$) and increases in it otherwise. That is, a more informative sample dampens the player's prior attitude on adoption.

Theorem 3.1 characterizes optimal sampling and unpacks posterior variance ψ^2. We refer

[19] For notational ease, we drop player's index in this single-player benchmark.

[20] The residual uncertainty about the project is $\mathrm{Var}[v] - \mathrm{Var}[\nu(\mathbf{a}, f(\mathbf{a}))]$.

to an optimal sample in this benchmark as a *single-player sample*. Posterior variance has two building blocks: (i) extrapolative weights $\tau_1(\mathbf{a}), \ldots, \tau_k(\mathbf{a})$ assigned to sample realizations, and (ii) intra-sample correlation, captured by $\sigma(a_j, a_m)$ for any pair of sample attributes $a_j, a_m \in \mathbf{a}$. As equation (8) shows, posterior variance adds up product terms of the form $\tau_j(\mathbf{a})\tau_m(\mathbf{a})\sigma(a_j, a_m)$ across all pairs of sample attributes. A pair of attributes adds to posterior variance if and only if their realizations are expected to reinforce each other towards the same adoption decision, i.e. $\tau_j(\mathbf{a})\tau_m(\mathbf{a})\sigma(a_j, a_m) > 0$ for any such $a_j, a_m \in \mathbf{a}$.

Theorem 3.1 (Single-player sampling)**.** *Fix $k \in \mathbb{N}$. Any single-player sample $\mathbf{a}^*$*

(i) consists of k distinct attributes, i.e. $\mathbf{a}^* \in \mathcal{A}_k \setminus \mathcal{A}_{k-1}$;

(ii) maximizes posterior variance $\psi^2(\cdot)$, given by

$$\mathbf{a}^* \in \arg\max_{\mathbf{a} \in \mathcal{A}_k} \quad \sum_{j=1}^{k} \tau_j(\mathbf{a}) \left(\sum_{m=1}^{k} \tau_m(\mathbf{a}) \sigma(a_j, a_m) \right) := \psi^2(\mathbf{a}); \tag{8}$$

(iii) does not depend on μ, ν_0, and r.

The theorem also shows that single-player sampling is independent of the player's expectations about the value of the project or the realizations of particular attributes. The desirability of a sample is determined only by its informativeness for adoption, which in turn depends neither on ν_0 nor on μ. For a fixed covariance, any project – promising or not – is evaluated based on the same sample of attributes. Moreover, the player uses the entire sampling capacity to minimize the residual uncertainty about the project.

An immediate consequence of part (iii) of theorem 3.1 is that whether the player is allowed to sample attributes sequentially is immaterial: the player would sample the same attributes. The key observation is that the posterior expected value follows a martingale. While sample realizations do inform the player's expected value from adoption, they do not affect the covariance across attributes that remain unsampled and the optimal continuation sample. The latter depends only on attributes sampled so far.[21] Therefore, costless delay of sampling is of no value to the player. Proposition 3.2 establishes the result.

Proposition 3.2 (Equivalence of sampling formats)**.** *In the single-player benchmark, a sample is optimal under sequential sampling of attributes if and only if it is optimal under simultaneous sampling.*

[21]Suppose that up to round t a sample $(\mathbf{a}_t, f(\mathbf{a}_t))$ of size t has been drawn. The updated covariance function for $a, a' \in \mathcal{A}$ is:

$$\sigma_t(a, a') = \sigma(a, a') - \Sigma(a, \mathbf{a}_t)[\Sigma(\mathbf{a}_t)]^{-1}\Sigma^\top(\mathbf{a}_t, a'),$$

where $\Sigma(a, \mathbf{a}_t)$ is the $1 \times t$ vector that lists the covariance of attribute a and that of each attribute in $\mathbf{a}_t$, and $\Sigma(\mathbf{a}_t)$ is the covariance matrix of the sample.

3.2 Example: Ornstein-Uhlenbeck covariance

For tractable covariance functions, theorem 3.1 can be operationalized to derive properties of the single-player sample. We illustrate this using a particular distance-based covariance function – the Ornstein-Uhlenbeck (OU) covariance – which arises as a natural choice in the context of optimal site selection in section 5. This discussion characterizes the posterior variance and the single-player sample when the player weighs all attributes equally.

We let $\mathcal{A} = [0, 1]$. For ease of exposition, we assume $\mu(\cdot) = 0$ and $\omega(\cdot) = 1$: attributes have identical means and are of equal importance to the player. For any pair of attributes, the covariance of their realizations depends only on the distance between the attributes. That is, for any $a, a' \in [0, 1]$,

$$\sigma_{OU}(a, a') := e^{-|a-a'|/\ell}. \tag{9}$$

This parametric form implies that $\sigma_{OU}(a, a) = 1$ for all $a \in [0, 1]$. The OU covariance is suitable for describing a project the attributes of which are ex ante equally uncertain. Therefore, $f(a) \sim \mathcal{N}(0, 1)$ for any $a \in [0, 1]$. Attributes differ only insofar as they sit at different positions in the attribute space.

The length-scale parameter ℓ measures how challenging it is to extrapolate away from an attribute. The higher ℓ is, the more predictable the attribute mapping is locally. Correlation decreases with distance, and ℓ captures the strength of that correlation across a fixed distance. Note that attributes become *perfectly correlated* as $\ell \to +\infty$ and *perfectly independent* as $\ell \to 0$. This observation allows a comparison of the single-player sample to each of these limit cases.

The first step is to derive the weights assigned to each sample realization. Lemma B.1 establishes that the expected realization of an out-of-sample attribute is inferred from sample attributes to its immediate left and right only. Extrapolation is therefore local. Due to the zero prior mean, the extrapolated mapping bends towards zero as illustrated in Figure 1b in the introduction. Because of such local extrapolation, the weight assigned to each sample attribute depends on its distance from its left and right sample neighbor. In fact, lemma 3.3 shows that the weight is strictly increasing and concave in such distance. The further away a sample attribute is from other sample attributes, the more its realization is weighted by in the posterior value.

Lemma 3.3. *Fix sample* $\mathbf{a} = (a_1, \ldots, a_k)$ *where* $k \geq 2$ *and* $0 \leq a_1 < \ldots < a_k \leq 1$. *Realization* $f(a_j)$ *is weighted by*

$$\tau_j(\mathbf{a}) = \begin{cases} \ell\left(1 - e^{-a_1/\ell} + \tanh\left(\frac{a_2 - a_1}{2\ell}\right)\right) & \textit{if } j = 1 \\ \ell\left(\tanh\left(\frac{a_{j+1} - a_j}{2\ell}\right) + \tanh\left(\frac{a_j - a_{j-1}}{2\ell}\right)\right) & \textit{if } j \in (1, k) \\ \ell\left(1 - e^{-(1-a_k)/\ell} + \tanh\left(\frac{a_k - a_{k-1}}{2\ell}\right)\right) & \textit{if } j = k. \end{cases}$$

For a singleton sample $\mathbf{a} = a_1$, $\tau_1(\mathbf{a}) = \ell\left(2 - e^{-a_1/\ell} - e^{-(1-a_1)/\ell}\right)$.

Lemma 3.3 provides the critical step in obtaining an explicit expression for posterior variance, the maximization of which then gives us the single-player sample. One might intuitively guess that for $k = 1$ the player optimally samples $a^* = 1/2$: attributes are equally uncertain, so the most informative attribute is that on average closest to all other attributes. Proposition 3.4 generalizes this intuition for any $k \geq 1$.

The unique single-player sample has an elegant structure: it is symmetric around the median attribute and adjacent sample attributes are equidistant. Sample attributes are spaced exactly so that they are weighted equally in the player's posterior value. As attributes become increasingly more correlated, the sample strictly expands to more extreme attributes.[22] Figure 3 illustrates the single-player sample for various capacities and values of ℓ. As capacity gets larger, the leftmost and rightmost attributes in the sample get closer to $a = 0$ and $a = 1$ respectively.

Proposition 3.4 (Single-player sampling)**.** *Let* $\mathbf{a}^* = (a_1^*, \ldots, a_k^*)$*, where* $a_1^* < \ldots < a_k^*$*, denote a single-player sample.*

(i) Realizations are weighted equally in $\mathbf{a}^*$*, i.e.* $\tau_j(\mathbf{a}^*) = \tau_m(\mathbf{a}^*)$ *for all* $m, j \in \{1, \ldots, k\}$*.*

(ii) The single-player sample is unique and characterized by the equations:

$$1 - e^{-a_1^*/\ell} = \tanh\left(\frac{1 - 2a_1^*}{2\ell(k-1)}\right), \quad a_j^* = a_1^* + (j-1)\frac{1 - 2a_1^*}{k-1}. \tag{10}$$

The single-player sample is symmetric with respect to $a = 1/2$*, i.e.* $a_j^* = 1 - a_{k-j+1}^*$ *for any* $j = 1, \ldots, k$*.*

The proof of proposition 3.4 builds on the fact that the player can increase posterior variance by shifting a non-symmetric sample slightly to the left or right − keeping the distance between any two adjacent sample attributes fixed. Extreme attributes $a = 0$ and $a = 1$ are never sampled optimally, but as either k or ℓ increase the single-player sample gets closer to these extreme attributes.[23]

[22] Simulations with the squared exponential covariance (Figure 1c) suggest that Proposition 3.4 qualitatively holds for a large class of distance-based covariances, parametrized as $\sigma(a, a') =: g(|a - a'|/\ell)$ for all $a, a' \in \mathcal{A}$, where $\ell > 0$ is a length-scale parameter.

[23] See Proposition G.4 for monotonicity in ℓ.

To further illustrate the use of theorem 3.1, online appendices F.1 and F.2 perform the same exercise with two more covariance functions. The Brownian covariance is suitable to study the presence of *benchmark attributes*, i.e., attributes that are known ex ante and which determine the expected outcome and uncertainty of all other attributes. The polynomial covariance, on the other hand, illustrates non-local extrapolation across a sample.[24]

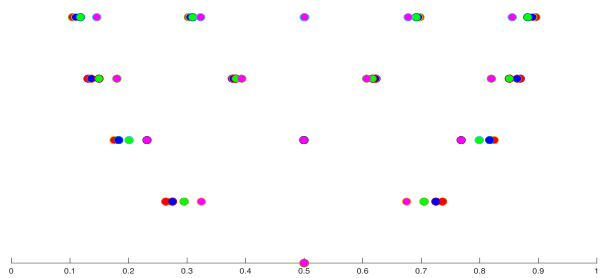


Figure 3: The single-player sample corresponding to covariance σ_{OU} illustrated for $k \in \{1, \dots, 5\}$ (bottom up) and $\ell = 1$ (red), $\ell = 1/2$ (blue), $\ell = 1/5$ (green), $\ell = 1/20$ (magenta).

4 Influence and distortion in principal-agent sampling

4.1 Sufficient statistics for a sample

We first simplify the agent's problem. In the principal-agent interaction the adoption decision is based on the principal's posterior value. But different samples induce different correlation between the players' posterior values. So the agent's choice of sample balances two considerations: he seeks to identify a sample that in expectation both leads to a well-informed adoption decision and aligns the players' posterior values. That is, he solves[25]

$$\max_{\mathbf{a} \in \mathcal{A}_k} \quad r_A + \underbrace{\Pr\left(\nu^P(\mathbf{a}) \geq r_P\right)}_{\text{probability of adoption}} \cdot \left(\underbrace{\mathbb{E}\left[\nu^A(\mathbf{a}) \mid \nu^P(\mathbf{a}) \geq r_P\right]}_{\substack{\text{expected value of an adopted} \\ \text{project for the agent}}} - r_A \right). \tag{11}$$

In striking the balance between informativeness and alignment of posterior values, the agent might distort the choice of attributes and forgo part of the sampling capacity.

Let $\rho(\mathbf{a})$ denote the correlation between posterior values $\nu^A(\mathbf{a})$ and $\nu^P(\mathbf{a})$ induced by sample $\mathbf{a}$. As expected, this pair of posterior values is jointly Gaussian, i.e.,

[24]This is due to the fact that unlike the Brownian motion and the Ornstein-Uhlenbeck process, the Gaussian process resulting from a polynomial covariance is non-Markov.

[25]Hereafter we adopt shorthand notation $\nu^i(\mathbf{a}) := \nu^i(\mathbf{a}, f(\mathbf{a}))$.

$$\begin{pmatrix} \nu^P(\mathbf{a}) \\ \nu^A(\mathbf{a}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \nu_0^P \\ \nu_0^A \end{pmatrix}, \begin{pmatrix} \psi_P^2(\mathbf{a}) & \rho(\mathbf{a})\psi_A(\mathbf{a})\psi_P(\mathbf{a}) \\ \rho(\mathbf{a})\psi_A(\mathbf{a})\psi_P(\mathbf{a}) & \psi_A^2(\mathbf{a}) \end{pmatrix} \right), \tag{12}$$

where ψ_A^2 and ψ_P^2 denote the posterior variance for the agent and principal respectively. There is an intuitive interpretation of the correlation induced by a sample: squared correlation $\rho^2(\mathbf{a})$ measures the proportion of the variance in $\nu^A(\mathbf{a})$ that is explained by the variance in $\nu^P(\mathbf{a})$. That is, if the agent were to indirectly predict his posterior value from observing only the principal's posterior value – i.e. no observation of individual attribute realizations – ρ^2 would measure the accuracy of that prediction.

Theorem 4.1 establishes that the agent's expected payoff from a sample depends on only two sufficient statistics. As explained below, these statistics summarize the extent to which the sample is informative for each player. This characterization reduces the agent's optimization over all feasible samples to optimization over a two-dimensional space.

Theorem 4.1 (Sufficient statistics for a sample)**.** *The agent's expected payoff from an arbitrary sample* $\mathbf{a} \in \mathcal{A}_k$ *is given by*

$$V_A(\mathbf{a}) = r_A + \left(\nu_0^A - r_A\right) \Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a})}\right) + \alpha_2(\mathbf{a})\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a})}\right), \tag{13}$$

where $\alpha_1(\mathbf{a}) := \psi_P(\mathbf{a})$ *and* $\alpha_2(\mathbf{a}) := \rho(\mathbf{a})\psi_A(\mathbf{a})$*. Moreover, for any* $\mathbf{a}, \mathbf{a}' \in \mathcal{A}_k$ *such that* $\alpha_1(\mathbf{a}) = \alpha_1(\mathbf{a}')$ *and* $\alpha_2(\mathbf{a}) > \alpha_2(\mathbf{a}')$*, agent strictly prefers* $\mathbf{a}$ *to* $\mathbf{a}'$*.*

First, $\alpha_1^2(\mathbf{a})$ corresponds to the principal's posterior variance, which is equal to the variance induced on the adoption decision. In his single-player benchmark, the principal would maximize this statistic. Hence $\alpha_1(\mathbf{a})$ is the sample's *informativeness for the principal.*

On the other hand, statistic $\alpha_2(\mathbf{a})$ corresponds to the amount of variation in agent's posterior value that is explained by the principal's posterior value. Intuitively, it captures the share of the posterior variance for the agent that gets reflected in the adoption decision. Naturally, $\alpha_2^2(\mathbf{a})$ is weakly smaller than $\psi_A^2(\mathbf{a})$. We refer to $\alpha_2(\mathbf{a})$ as the sample's *adjusted informativeness for the agent*, adjusted for the fact that the agent lacks adoption authority. Its sign reflects the sign of $\rho(\mathbf{a})$. The higher a sample scores in this statistic, the more accurately the adoption decision reflects the agent's interest.

Equation (13) makes plain the two channels through which the agent's sampling strategy influences adoption. First, the sample choice determines the probability of adoption. The extent to which the agent prefers more frequent adoption based on her prior value is reflected in the payoff term:

$$\left(\nu_0^A - r_A\right)\Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a})}\right). \qquad \text{(Adoption frequency)}$$

A more informative sample for the principal (i.e. higher α_1) moves the principal closer to agent's ex ante preferred decision if players are in prior disagreement. It moves her further away from the jointly preferred decision if the players are in prior agreement. So, more information seeks to *persuade* under prior disagreement and *caution* under prior agreement.[26]

The sample choice determines also whether the prospect of adoption is good news for the agent. This depends on the correlation induced between the players' posterior values: a better informed principal is beneficial for the agent if and only if the sample correlates their posterior values positively. This is reflected in the second term of the agent's payoff:

$$\alpha_2(\mathbf{a})\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a})}\right). \qquad \text{(Adoption accuracy)}$$

Using this payoff characterization, we address two questions. First, will the agent ever suppress informativeness for both himself and the principal, in the sense that the optimal sample is dominated in both sufficient statistics? Second, will the agent ever find it optimal to draw a sample that induces negative correlation between players' values? Sections 4.2 and 4.3 analyze two stark special cases, whereas section 4.4 generalizes the discussion.

Remark 4.1 (Delay in sampling)**.** *The equivalence between simultaneous and sequential sampling in the single-player benchmark no longer holds here. The agent benefits from the flexibility of gradually tailoring the sample to the realized path of posterior values. Our analysis below shows that the solution to the agent's simultaneous-sampling problem depends on* (ν_0^P, ν_0^A)*. By a similar logic, under sequential sampling the posterior expected values that the players hold at sampling round* t *will inform the agent's optimal choice in round* $(t+1)$*.*

4.2 No conflict over attribute weights

We first illustrate suppression of information in a simple setting in which players have equal attribute weights but different outside options. Because they weigh attributes in the same way, players hold the same prior and posterior values: $\nu_0 := \nu_0^P = \nu_0^A$ and $\nu_A(\mathbf{a}) = \nu_P(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}_k$. Also, the two sufficient statistics of Theorem 4.1 collapse to a single one: $\psi_P(\mathbf{a}) = \psi_A(\mathbf{a}) =: \psi(\mathbf{a})$. So players agree entirely on which samples are most informative.

Despite holding the same posterior value, players might disagree on the adoption decision to be made based on it. Hence the agent uses sampling to influence the probability of an

[26] If the agent simply sought to maximize the probability of an adoption $\Phi\left((\nu_0^P - r_P)/\psi_P(\mathbf{a})\right)$, the optimal sample would consist of the single-principal sample if $\nu_0^P \leq r_P$ and no informative sampling otherwise.

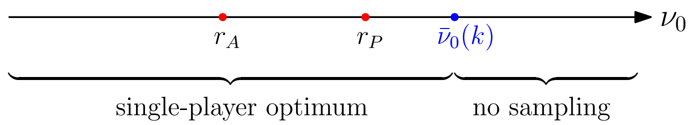


Figure 4: Proposition 4.2(i)

adoption. His payoff from a sample can be reformulated as equal to the principal's payoff plus an adoption wedge proportional to the difference in their outside options, i.e.

$$V_A(\mathbf{a}) = V_P(\mathbf{a}) + \underbrace{(r_A - r_P)\Phi\left(\frac{r_P - \nu_0}{\psi(\mathbf{a})}\right)}_{\text{adoption wedge}}.$$

The principal's payoff strictly increases in posterior variance – he always prefers a more informative sample – but the adoption wedge need not. If the agent is the less conservative player $r_A < r_P$, he seeks to induce more adoption. The way to do that is to provide more information if $\nu_0 \leq r_P$ and restrict information otherwise. It is straightforward to show that for ψ sufficiently small, the sensitivity of the wedge term to ψ dominates. Hence, the agent's payoff is single-troughed in ψ: he prefers either zero posterior variance (i.e. no informative sampling) or the highest attainable posterior variance (i.e. a single-player sample).

Proposition 4.2 establishes that if the prior value is sufficiently extreme the agent forgoes the entire sampling capacity. On the other hand, whenever sampling takes place a single-player sample is optimal. Therefore, the only distortion possible here is the full suppression of informativeness for both players. This implies that some conflict over attribute weights is necessary for partially informative samples to be optimal.

Proposition 4.2 (All-or-nothing sampling)**.** *Let $\omega_A = \omega_P$ and $r_A \neq r_P$.*

(i) If $r_A < r_P$ (resp., $r_A > r_P$), there exists a threshold $\bar{\nu}_0(k) > r_P$ (resp., $\underline{\nu}_0(k) < r_P$) such that the agent samples a single-player sample for $\nu_0 \leq \bar{\nu}_0(k)$ (resp., $\nu_0 \geq \underline{\nu}_0(k)$) and forgoes all sampling otherwise.

(ii) Threshold $\bar{\nu}_0(k)$ weakly increases in k and $\bar{\nu}_0(k) \to \bar{\nu}_0 < +\infty$ as $k \to \infty$. Threshold $\underline{\nu}_0(k)$ weakly decreases in k and $\underline{\nu}_0(k) \to \underline{\nu}_0 > -\infty$ as $k \to \infty$.

Three remarks are in order about proposition 4.2. First when in prior disagreement the agent always draws the most informative sample. Second, the agent forgoes sampling whenever he is sufficiently more confident than the principal in his ex ante decision. In Figure 4 sampling is forgone for sufficiently promising projects: players both favor adoption ex ante, but the agent is the less conservative player. From the agent's perspective, informative sampling is too risky in such circumstances. Third, agent's willingness to sample expands as capacity increases: if ν_0 is such that he samples for some k, he also samples for all $k' > k$.

4.3 Extreme conflict over attribute weights

Consider now two players who are diametrically opposed about attributes: $\omega_A(a) = -\omega_P(a)$ for all $a \in \mathcal{A}$. The pair of sufficient statistics boils down to $(\psi(\mathbf{a}), -\psi(\mathbf{a}))$, where $\psi_A^2(\mathbf{a}) = \psi_P^2(\mathbf{a}) := \psi^2(\mathbf{a})$. So even though all feasible samples induce perfect negative correlation between the players' posterior values, players rank samples in the same way according to their informativeness.

For expositional ease, we focus on $r_A = r_P := r$.[27] First, it is immediate that if $r = 0$ no informative sampling is uniquely optimal. For any possible realization of posterior values, the principal uses the accumulated information to do the opposite of what the agent would do.[28] With $r \neq 0$, the possibility of more nuanced distortions arises. Rather than choosing between a single-player sample and no sample at all, the agent gradually suppresses posterior variance by choosing a partially informative sample, the posterior variance of which is strictly between zero and the maximal one. Even though all available samples induce negative correlation, the agent resorts to sampling when the principal's prior attitude towards the project is particularly unfavorable. Through sampling, the agent seeks to overturn the principal's prior bias at the cost of greater misalignment.

Proposition 4.3. *Let $r_A = r_P = r$ and $\omega_A(a) = -\omega_P(a)$ for any $a \in \mathcal{A}$.*

1. *If $r = 0$, the agent forgoes all informative sampling for any k.*
2. *Let $r > 0$ (resp., $r < 0$). There exists a unique $\underline{\nu}_0(k) < -r$ (resp., $\bar{\nu}_0(k) > -r$) such that the agent optimally:*
   - *(i) forgoes all informative sampling for $\nu_0^A \geq -r$ (resp., $\nu_0^A \leq -r$) ,*
   - *(ii) samples a partially informative sample for $\nu_0^A \in [\underline{\nu}_0(k), -r]$ (resp., $\nu_0^A \in [-r, \bar{\nu}_0(k)]$),*
   - *(iii) samples a single-player sample for $\nu_0^A \leq \underline{\nu}_0(k)$ (resp., $\nu_0^A \geq \bar{\nu}_0(k)$).*
3. *Threshold $\underline{\nu}_0(k)$ weakly decreases in k and threshold $\bar{\nu}_0(k)$ weakly increases in k.*

---

[27] The results of this section remain qualitatively the same for any $r_A \neq r_P$.

[28] Formally, for any given ν_0^A and any sample $\mathbf{a} \in \mathcal{A}_k$, the agent's payoff from sampling it is

$$\nu_0^A \Phi\left(-\frac{\nu_0^A}{\psi(\mathbf{a})}\right) - \psi(\mathbf{a})\phi\left(-\frac{\nu_0^A}{\psi(\mathbf{a})}\right)$$

which is strictly greater than the payoff from not sampling given by $\min\{\nu_0^A, 0\}$.

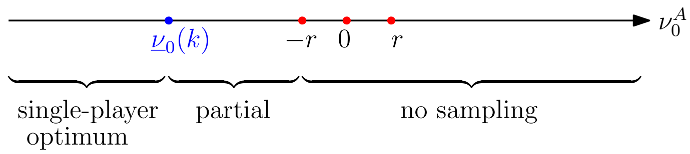


Figure 5: Proposition 4.3(ii). Given $r > 0$, principal adopts ex ante iff $\nu_0^A \leq -r$ and agent adopts iff $\nu_0^A \geq r$.

Key to the proof of proposition 4.3 is the observation that the agent's payoff is now single-peaked in ψ. Intuitively, negatively correlation posterior values is an instrument of last resort for the agent. He turns to it only if players are in prior disagreement and he is relatively more confident in his preferred decision than the principal prior to sampling. This explains why in Figure 5 projects with $\nu_0^A < -r$ are the ones for which the agent samples: intervals $(-\infty, -r]$ and $[r, \infty)$ are both disagreement regions, but only in the former is the agent the ex ante more confident player.

4.4 General attribute weights and outside options

4.4.1 Joint suppression of informativeness

No informative sampling in proposition 4.2 offers the most extreme manifestation of suppression of informativeness for both players. With general attribute weights, the agent might find it beneficial to settle for a sample with a low α_1 and α_2 without lowering both statistics all the way to zero. This discussion identifies general sufficient conditions for such suppression of informativeness to arise. The pair (α_1, α_2) determines a partial ranking of feasible samples according to their informativeness. We say that a sample is *dominated* if there exists another sample that induces higher informativeness for both players.

Definition 1 (Dominated sample)**.** For any $\mathbf{a}, \mathbf{a}' \in \mathcal{A}_k$, $\mathbf{a}'$ *dominates* $\mathbf{a}$ if $\alpha_1(\mathbf{a}') \geq \alpha_1(\mathbf{a})$ and $\alpha_2(\mathbf{a}') \geq \alpha_2(\mathbf{a})$ with at least one strict inequality. Sample $\mathbf{a}$ is *dominated* if there exists a feasible $\mathbf{a}' \neq \mathbf{a}$ that dominates $\mathbf{a}$.

All else equal, the agent's payoff given by (13) is (i) strictly increasing in α_1 if and only if players are in prior disagreement, and (ii) strictly increasing in α_2 for any (ν_0^A, ν_0^P). That is, higher variance in the principal's posterior value benefits the agent whenever he seeks to challenge the principal's prior view of the project. Therefore, if the agent ever prefers a sample that is dominated, it must be that the players are in prior agreement.

To get to sufficient conditions for information suppression, we exploit the observation made in section 4.2 that the agent needs to be relatively more confident than the principal in his prior value. If the principal is exactly indifferent ex ante, the optimal sample attains the highest α_2 that is feasible: we refer to it as α_2-maximal.[29] This is because the ex-ante probability that the principal adopts is 1/2 for any sample choice. Given that the agent cannot influence principal's probability of adoption, he opts for maximizing the accuracy of adoption instead.

Consider now an almost-indifferent principal who is in prior agreement with the agent: say, they both favor adoption. If the agent is sufficiently more confident than the principal about adoption, steering the principal towards more adoption is a first-order concern for him. But the only way he can encourage more adoption is by reducing posterior variance for the principal, given that the principal is inclined towards adoption ex ante. Therefore the agent sacrifices some adjusted informativeness for himself in order to reduce informativeness for the principal. If his prior value is sufficiently extreme, he forgoes all sampling. But if his prior value is not too extreme, the optimal sample will be both somewhat informative and dominated. Proposition 4.4 formalizes this insight.

Proposition 4.4 (Joint suppression of informativeness)**.**

(1) An optimal sample is dominated only if players are in prior agreement.

(2) If all the following hold:

(i) the players are in prior agreement,

(ii) there exists at least one sample $\mathbf{a} \in \mathcal{A}_k$ such that $\rho(\mathbf{a}) > 0$,

(iii) at any α_2-maximal feasible sample, there exists a sample arbitrarily close to it that is dominated by it

then there exist $\bar{x}^P$ and $\underline{x}^A < \bar{x}^A$ such that for $|\nu_0^P - r_P| \leq \bar{x}^P$ and $\underline{x}^A \leq |\nu_0^A - r_A| \leq \bar{x}^P$ any optimal sample $\mathbf{a}^*$ is dominated and has $\alpha_1(\mathbf{a}^*) > 0$.

4.4.2 Controversial sampling

Proposition 4.3 suggests that a sample that induces negative correlation among players' posterior values can be optimal. In such a case, sampling induces players to grow farther apart in their preference for adoption. But in section 4.3 all feasible samples were controversial. More generally, we argue that a controversial sample can arise as optimal even when other non-controversial samples are available. But first a definition.

Definition 2 (Controversial sampling)**.** A sample $\mathbf{a} \in \mathcal{A}_k$ is *controversial* if $\rho(\mathbf{a}) < 0$.

[29] See lemma C.1. If the indifferent principal breaks the tie in favor of adoption, the agent samples an α_2-maximal sample if ν_0^A is not too large and forgoes sampling otherwise. The opposite holds if the principal breaks the tie in favor of rejection.

The prospect of an adoption from a controversial sample is unfavorable for the agent: the agent's expected value from an adopted project is even lower than his prior value. He is willing to face such a prospect only if players are in prior disagreement, in which case sampling might induce the principal to switch to agent's ex ante preferred decision with some probability. That is, prior disagreement is necessary for controversial sampling.

But prior disagreement is not enough. For a controversial sample to be optimal, it is also necessary that (i) all feasible controversial samples induce sufficiently weak negative correlation, and (ii) all available non-controversial samples generate too little information for the principal and hence too little uncertainty in the adoption decision. That is, a controversial sample has to be sufficiently appealing in terms of the adoption frequency that it generates relative to the mismatch of players' interests. Proposition 4.5 formalizes the intuition.[30]

Proposition 4.5 (Influence via controversial sampling)**.**

(i) A controversial sample is optimal only if players are in prior disagreement.

(ii) When in prior disagreement, agent forgoes informative sampling if and only if all feasible samples are controversial and $\rho(\mathbf{a})$ *is sufficiently negative for all* $\mathbf{a} \in \mathcal{A}_k$*.*

(iii) If the optimal sample $\mathbf{a}^*$ *is controversial, then for any feasible non-controversial sample* $\mathbf{a}$*,* $\psi_P(\mathbf{a}^*) > \psi_P(\mathbf{a})$*.*

5 Application: Site selection and generalizability

This application recasts the attribute sampling problem as a problem of site selection in program evaluation. We reinterpret $\mathcal{A} = [0, 1]$ as the set of *target sites* of a novel social program, the outcomes of which are uncertain and site-specific. That is, sites are ordered from left to right according to an observable outcome-relevant characteristic, e.g., median income.[31] In this context, a sample is a collection of small-scale pilot studies located at particular sites. Ultimately the program is either adopted across all target sites or in none of them.

We adopt the Ornstein-Uhlenbeck covariance introduced in section 3.2. Site outcomes are positively correlated and sites in close proximity are likely to witness similar outcomes

[30] For brevity, proposition 4.5 focuses on necessary conditions for controversial sampling. If the condition of proposition 4.5(iii) is modified so that *any* controversial sample is more informative for the principal than any other non-controversial sample, the provided conditions are also sufficient.

[31] This could easily be generalized to a vector of $d > 1$ outcome-relevant observables for a site by adopting a multi-dimensional OU covariance function (with an appropriate metric $||a - a'||$ for $a, a' \in [0, 1]^d$).

from the program. Parameter $\ell \in [0, \infty)$ captures the extent to which pilot findings can be extrapolated across a fixed distance.

A researcher interested in the average outcome across all target sites has the authority to choose pilot sites. He interacts with an evaluator, who decides on large-scale implementation. The evaluator is exclusively interested in the outcome of a single site $a_P \in [1/2, 1)$. That is, the agent holds attribute weight $\omega_A(a) = 1$ for all $a \in [0, 1]$, whereas the principal holds $\omega_P(a_P) = +\infty$ and $\omega_P(a) = 0$ for all $a \neq a_P$.[32] We let $r_A = r_P = 0$. The prior value of the evaluator is given by $\mu(a_P)$ whereas that of the researcher is the average attribute realization $\bar{\mu} := \int_0^1 \mu(a)da$.

The goal of this application is to illustrate how the researcher distorts site selection in order to exert influence over program adoption. In empirical work there is a growing awareness that pilot programs, although grounded in scientific practices, are often utilized strategically among political actors with conflicting interests (Allcott (2015), Rogers-Dillon (2004), Weiss (1993)). Even when pilot findings are disclosed truthfully, the selection of pilot sites is in itself an important source of influence.

Researcher's benchmark. In the absence of the evaluator, the researcher would select sites as described in proposition 3.4: he would pick sites that are most informative about the average outcome across all sites. His solution relates naturally to the concept of *purposive site selection* in impact evaluations – a procedure that aims to identify sites that "yield the most information and have the greatest impact in the development of knowledge" (Patton (2014), Olsen et al. (2013)).[33]

Reinterpreted, theorem 3.1 has three implications for purposive site selection. First, the selection of pilot sites should be free of site selection bias, where bias is defined as non-zero correlation between the probability that a program is evaluated at particular sites and the program's outcome at those sites (Allcott, 2015). The single-player sample does not depend on expected site outcomes. Second, the choice of pilot sites maximizes the generalizability of pilot findings, where generalizability of a sample $\mathbf{a}$ is to be understood

[32] The evaluator's ω_P is the Dirac-delta function. For any integrable f, the following are well-defined:

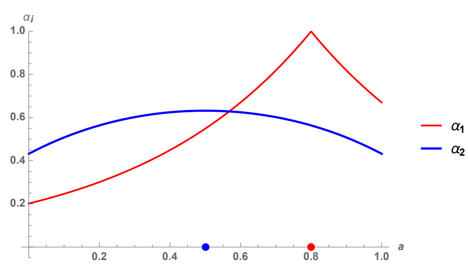
$$\int_0^1 \omega_P(a)\,\mathrm{d}a = 1, \qquad \int_0^1 f(a)\omega_P(a)\,\mathrm{d}a = f(a_P).$$

One can interpret attribute weight ω_P as the limit of (i) the density of $\mathcal{N}(a_P, \sigma^2)$ as $\sigma \to 0$, or alternatively (ii) $\tilde{\omega}_P(a) = \varepsilon$ for $a \in [a_P - \varepsilon/2, a_P + \varepsilon/2]$ and $\tilde{\omega}_P(a) = 0$ otherwise as $\epsilon \to 0$.

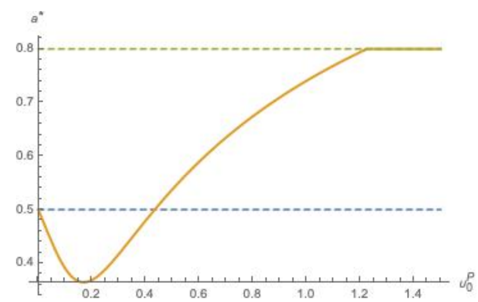
[33] This also relates to the *test bore approach* to site selection proposed in the New Jersey Income Maintenance Experiment in the 1970s: "examining a discrete number of purposely chosen and distinctive samples from which a complete composite can eventually be formed" (Watts, Peck and Taussig, 1977).

as the ratio of posterior variance to ex ante uncertainty about the project $\psi^2(\mathbf{a})/\text{Var}(v)$.[34] Third, the theorem implies that the researcher does not benefit from running pilot studies sequentially. But even if run sequentially, the outcome observed at a given pilot site is interpreted in the same way no matter whether it is an early or a late pilot.

Strategic site selection. For sharper insight we focus on the strategic choice of a single pilot site (i.e., $k = 1$). This capacity would suffice for the evaluator to fully learn her value: in her single-player benchmark, she would sample $a_P^* = a_P$. The researcher, on the other hand, would sample $a_A^* = 1/2$. Because any site correlates the players' posterior values perfectly, the sufficient statistics pair corresponding to $a \in [0, 1]$ simplifies to $(\alpha_1(a), \alpha_2(a)) = (\psi_P(a), \psi_A(a))$, as depicted in Figure 6a. Sites in the interval $[1/2, a_P]$ are



(a) The plot shows $(\alpha_1(\cdot), \alpha_2(\cdot))$ for $a_P = 4/5$ and $\ell = 1/2$. The blue and red dots mark the single-player optima.



(b) The plot depicts the optimal site as a function of $\mu(a_P)$ given $\bar{\mu} = 1/2$. The dashed lines depict $a_A^* = 1/2$ and $a_P^* = 4/5$.

Figure 6: Sufficient statistics (a) and optimal selection of a peripheral site (b).

compromise sites: slightly shifting the pilot site strictly increases informativeness for one player and decreases it for the other.

Three observations follow as corollaries of our analysis, and are expanded in appendix G.1. First, prior disagreement between the researcher and the evaluator always leads to an optimal site that is a compromise: $a^* \in [1/2, a_P]$. The exact tradeoff between ψ_A and ψ_P is pinned down by the relative magnitude of prior values $\mu(a_P)$ and $\bar{\mu}$. The more extreme $\mu(a_P)$ is, the more the researcher has to cater to the evaluator and therefore the closer is the optimal site to a_P.

Second, the possibility of an optimal site that is not a compromise – and hence socially inefficient – arises only if players are in prior agreement. Figure 6b plots the unique optimal site as a function of μ_P. For $|\mu_P|$ sufficiently close to zero relative to $\bar{\mu}$, the optimal site is

[34] For a related notion of generalizability, see Vivalt (Forthcoming).

even further away from a_P than the median site. The researcher undershoots his preferred site to reduce the uncertainty in the evaluator's adoption.

Third, such distortion can arise even if $a_P = 1/2$ – that is, when players rank sites in the same away and the median site is the most informative for both. The researcher influences by sampling a site other than the median site. We show that moderate correlation can even lead to selection of the most peripheral sites $a = 0$ and $a = 1$, whereas distortions vanish as sites become perfectly independent or perfectly correlated.

Random site selection. An empirically common alternative to purposive site selection is random selection of pilot sites, with uniform and stratified uniform selection being the most pervasive formats.[35] Online appendix G.2 offers a comparison of purposive site selection to the expected sites from each of these formats. Strikingly, we show that purposive site selection converges to one of these formats as site outcomes become perfectly independent $\ell \to 0$ or perfectly correlated $\ell \to +\infty$.

6 Sample centrality: A graph-theoretic approach

For the distance-based covariance of section 3.2, the single-player sample is strikingly central in the attribute space: sample attributes are equally spaced and symmetric around the median attribute. This extension formalizes the centrality of the single-player sample for any arbitrary covariance. Taking a graph-theoretic approach to the attribute problem, we construct and interpret an appropriate centrality measure in the attribute graph, *sample centrality*. This section heuristically argues that any single-player sample attains maximal sample centrality in the attribute graph.[36]

We consider an infinite weighted attribute graph $\mathcal{G} = (\mathcal{A}, E)$, where $\mathcal{A}$ is the set of attribute-nodes and E is the set of weighted and undirected edges (Figure 7a). The weight of an edge $aa' \in E$ joining nodes $a, a' \in \mathcal{A}$ is equal to $\sigma(a, a')$.[37] Within $\mathcal{G}$, we let $\mathcal{G}_{\mathbf{a}}$ denote the subgraph consisting only of nodes in sample $\mathbf{a}$ and edges joining them. A key object for our analysis is the *walk product* between any two nodes.

Definition 3 (Alternating walk product)**.** A walk w of length $\ell \geq 0$ is a sequence of attribute-nodes $w = (a^1, \ldots, a^\ell, a^{\ell+1})$ such that $a^m a^{m+1} \in E$ for $m \in \{1, \ldots, \ell\}$. The alternating walk product for walk w is defined as

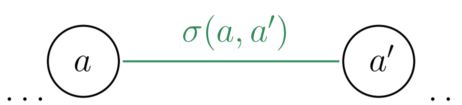
[35] For instance, the evaluation of the Job Training Partnership Act of 1982 initially aimed to select sites in which to conduct the study through either random or stratified random selection "in order to obtain nationally representative results." See Hotz (1992).

[36] The technical core of this section is developed in online appendix H.

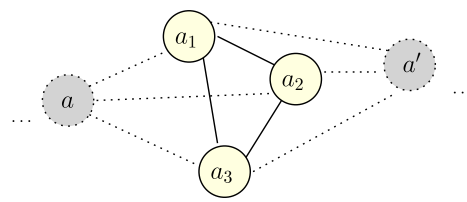
[37] Without loss, attribute weights are normalized so that $\sigma(a, a) = 1$ for any $a \in \mathcal{A}$. Hence, the edge weight equals the correlation between the two attribute-nodes that it joins.

$$\kappa(w) := (-1)^{\ell} \prod_{m=1}^{\ell} \sigma(a^m, a^{m+1}).$$

Heuristically, a walk corresponds to one particular inference channel from an attribute to another: the alternating walk product gives us the strength and sign of this inference. If one were to sum up the alternating product for all possible walks from node a to node a', one would know how much can she infer about $f(a')$ from $f(a)$. But we need only a subset of such inference channels. As depicted in Figure 7b, we need only sum up the alternating product for walks from a to a' that go exclusively through $\mathcal{G}_{\mathbf{a}}$. This way we obtain the *alternating walk sum* denoted by $(a \xrightarrow{\mathbf{a}} a')$. The alternating walk sum measures how well a sample connects any two nodes.



(a) Two attribute-nodes.



(b) Sample nodes $\mathbf{a} = (a_1, a_2, a_3)$ *are in yellow, edges of* $\mathcal{G}_{\mathbf{a}}$ *in solid lines, and two out-of-sample nodes* a, a' *in grey.*

Figure 7: Attribute graph $\mathcal{G} = (A, E)$

Hereafter, we assume that the alternating walk sum is well-defined for any sample $\mathbf{a} \in \mathcal{A}_k$, in the sense that for any $a_i, a_j \in \mathbf{a}$ the walk sum $(a_i \xrightarrow{\mathbf{a}} a_j)$ converges to the same value for every possible summation order of walks.[38]

We are now ready to introduce a centrality measure $\gamma : \mathcal{A}_k \to \mathbb{R}$ over finite samples of nodes. For any given sample, this measure quantifies how strongly the sample connects any two randomly drawn nodes in $\mathcal{A}$, where the probability of each node being drawn is given by ω. For a simplistic illustration, suppose $\mathcal{A} = \{a_1, a_2, a_3\}$ and $\mathbf{a} = (a_1)$. Up to order, the possible pairs for a first node and a final node are $(a_2 \xrightarrow{a_1} a_3)$, $(a_2 \xrightarrow{a_1} a_2)$, $(a_3 \xrightarrow{a_1} a_3)$, $(a_1 \xrightarrow{a_1} a_1)$, $(a_1 \xrightarrow{a_1} a_2)$, and $(a_1 \xrightarrow{a_1} a_3)$. Thus $\gamma(\mathbf{a})$ consists of the average across all such possible alternating walk sums.

Definition 4 (Sample centrality)**.** For any $\mathbf{a} \in \mathcal{A}_k$, its *sample centrality* $\gamma(\mathbf{a})$ is defined as the expected alternating walk sum $\mathbb{E}_{a,a'}\left[(a \xrightarrow{\mathbf{a}} a')\right]$ from a random first node $a \in \mathcal{A}$ to a random last node $a' \in \mathcal{A}$ such that:

[38] We relax this assumption in online appendix H.3. This allows us to have a notion of sample centrality that is well-defined for any positive definite covariance function.

(i) the ordered pair (a, a') is drawn according to density $g(a, a') := \omega(a)\omega(a')$;

(ii) for each walk $w = (w_0, \dots, w_\ell)$, $w_0 = a$, $w_\ell = a'$, and $w_j \in \mathbf{a}$ for $j = 1, \dots, \ell - 1$.

Theorem 6.1 establishes that posterior variance $\psi^2(\mathbf{a})$ corresponds exactly to the sample centrality of $\mathbf{a}$. Hence, the single-player sample is the most central sample in $\mathcal{G}$. As a stepping stone, we first identify a graph representation of sample weights (Lemma H.3). The weight $\tau_j(\mathbf{a})$ corresponds to the expected alternating walk sum from a randomly drawn node in $\mathcal{A}$ to a_j. Because the random node is generically not in $\mathbf{a}$, $\tau_j(\mathbf{a})$ quantifies the average "graph distance" from out-of-sample nodes to a_j.[39] Therefore, terms of the form $\tau_j(\mathbf{a})\tau_m(\mathbf{a})\sigma(a_j, a_m)$ in the expression for posterior variance capture inference paths that go from a to a_j to a_m to a'.

Theorem 6.1 (Sample centrality of a single-player sample)**.**

(i) For any sample $\mathbf{a}$*, its sample centrality is equal to the posterior variance that the sample induces, i.e.* $\gamma(\mathbf{a}) = \psi^2(\mathbf{a})$*.*

(ii) Fix capacity k*. Any single-player sample attains maximal sample centrality.*

Sample centrality differs from most standard centrality measures insofar as it is well defined over finite subsets of nodes. The two existing measures most closely related to it are *betweenness centrality* and *Bonacich centrality* (Freeman (1977), Katz (1953), Bonacich (1987)). Akin to betweenness centrality, sample centrality measures the extent to which a subset of nodes connects any two arbitrary nodes in the graph. But it differs from betweenness centrality in that it counts walks of arbitrary length rather than geodesic paths. On the other hand, its walk-based construction relates it to Bonacich centrality.[40] In this way, sample centrality offers a natural walk-based generalization of betweenness centrality to non-singleton subsets of nodes.

Lastly, note that even though this extension focused on single-player sampling, principal-agent sampling has a graph-theoretic interpretation too. It is immediate that the underlying attribute graph $\mathcal{G}$ is the same for both players. The covariance between $\nu^P(\mathbf{a})$ and $\nu^A(\mathbf{a})$, which is given by

$$\text{Cov}\left(\nu^A(\mathbf{a}), \nu^P(\mathbf{a})\right) = \sum_{j,m=1}^{k} \tau_j^P(\mathbf{a})\tau_m^A(\mathbf{a})\sigma(a_j, a_m).$$

[39] The random node is generically not in $\mathbf{a}$ if the support of ω is infinite.

[40] Bonacich centrality discounts walks of length ℓ by β^ℓ, where $\beta \in (0, 1)$ is the discount factor. In contrast, sample centrality discounts walks of length ℓ by $(-1)^\ell$. The sign-alternating Bonacich centrality appears also in network games of pure substitutes (Bramoulle, Kranton and D'Amours (2014)). See section H.4.

corresponds to the expected alternating walk sum through $\mathbf{a}$ between any two random attributes, where one is drawn according to density ω_P and the other according to ω_A. That is, $\text{Cov}\left(\nu^A(\mathbf{a}), \nu^P(\mathbf{a})\right)$ captures how strongly the sample connects attributes of greatest interest for the agent to those of greatest interest for the principal.

7 Concluding remarks

Attribute sampling is central to a variety of economic applications in which attributes are inherently correlated. This paper offered a flexible theoretical framework for addressing the optimal learning of attributes. Our analysis established a benchmark for attribute sampling and studied distortions that arise when agency conflict is present.

We recognize the variety of extensions that follow from our analysis but are beyond the scope of this paper. First, our model abstractly covers many Gaussian processes that have so far been unexplored in learning models. Identifying natural and tractable applications for particular covariance classes (especially for multi-dimensional processes) is an immediate next step. Relatedly, the careful study of non-Markov covariance functions, which entail non-local patterns of extrapolation, is an exciting prospect for models of complex experimentation as in Callander (2011) and Garfagnini and Strulovici (2016).

Another important direction for future work is the possibility of partial adoption. The attribute problem is defined by the fact that the agent is forced to consume all attributes upon adoption. A natural relaxation would be to allow the agent to flexibly choose the subset of attributes he would like to adopt upon inspection. This direction conceptually bridges the attribute problem to traditional models of search, as well as it is related to the A/B testing problem in Azevedo et al. (2018).

Our framework is also promising for richer models of agency conflict. One such direction building on section 5 is to allow target sites to independently run their own pilot studies so as to influence the adoption decision of a social planner (e.g., a federal government). The question of how gradual discovery of site outcomes unfolds across sites is of particular interest. We leave these questions to future work.

References

Aghion, Philippe, Patrick Bolton, Christopher Harris, and Bruno Jullien. 1991. "Optimal Learning by Experimentation." *The Review of Economic Studies*, 58(4): 621–654. 6

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics*, 130(3): 1117–1165. 4, 24

Azevedo, Eduardo M., Alex Deng, José Luis Montiel Olea, Justin Rao, and E. Glen Weyl. 2018. "A/B Testing." 29

Banerjee, Abhijit, Sylvain Chassang, Sergio Montero, and Erik Snowberg. 2017. "A Theory of Experimenters." 7

Bardhi, Arjada. 2018. "Optimal Discovery and Influence Through Selective Sampling." 6

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, AliceNg'ang'a, and Justin Sandefur. 2018. "Experimental evidence on scaling up education reforms in Kenya." *Journal of Public Economics*, 168: 1–20. 4

Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology*, 92: 1170–1182. 28

Bramoulle, Yann, Rachel Kranton, and Martin D'Amours. 2014. "Strategic Interaction and Networks." *American Economic Review*, 104(3): 898–930. 28

Brocas, Isabelle, and Juan D. Carrillo. 2007. "Influence Through Ignorance." *The RAND Journal of Economics*, 38(4): 931–947. 6

Callander, Steven. 2011. "Searching and Learning by Trial and Error." *American Economic Review*, 101: 2277–2308. 5, 6, 10, 29

Callander, Steven, and Patrick Hummel. 2014. "Preemptive Policy Experimentation." *Econometrica*, 82(4): 1509–1528. 6

Callander, Steven, and Tom S. Clark. 2017. "Precedent and Doctrine in a Complicated World." *American Political Science Review*, 111(1): 184–203. 6

Che, Yeon-Koo, and Konrad Mierendorff. 2019. "Optimal Dynamic Allocation of Attention." *American Economic Review*, 109(8): 2993–3029. 6

Chilés, Jean-Paul, and Pierre Delfiner. 2012. *Geostatistics: Modeling Spatial Uncertainty. Wiley Series in Probability and Statistics*. 2nd ed., John Wiley & Sons. 5

Ding, Liang, and Xiaowei Zhang. 2018. "Scalable Stochastic Kriging with Markovian Covariances." 10, 37

Di Tillio, Alfredo, Marco Ottaviani, and Peter Norman Sørensen. 2017*a*. "Persuasion Bias in Science: Can Economics Help?" *The Economic Journal*, 127: 266–304. 7

Di Tillio, Alfredo, Marco Ottaviani, and Peter Norman Sørensen. 2017*b*. "Strategic Sample Selection." 7

Everett, M.G., and S.P. Borgatti. 1999. "The Centrality of Groups and Classes." *Journal of Mathematical Sociology*, 23(3): 181–201. 5

Freeman, L. P. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry*, 40: 35–41. 5, 28

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *American Economic Review*, 108(12): 3651–3684. 6

Garfagnini, Umberto, and Bruno Strulovici. 2016. "Social Experimentation with Interdependent and Expanding Technologies." *The Review of Economic Studies*, 83(4): 1579–1613. 6, 29

Geng, Sen, Leonardo Pejsachowicz, and Michael Richter. 2017. "Breadth versus Depth." 5

Glazer, Jacob, and Ariel Rubinstein. 2004. "On Optimal Rules of Persuasion." *Econometrica*, 72(6): 1715–1736. 6

Henry, Emeric, and Marco Ottaviani. 2019. "Research and the Approval Process: The Organization of Persuasion." *American Economic Review*, 109(3): 911–55. 6

Hirsch, Alexander V. 2016. "Experimentation and Persuasion in Political Organizations." *American Political Science Review*, 110(1): 68–84. 7

Hofmann, Thomas, Bernhard Schölkopf, and Alexander J. Smola. 2008. "Kernel Methods in Machine Learning." *The Annals of Statistics*, 36(3): 1171–1220. 5

Hotz, Joseph. 1992. "Designing Experimental Evaluations of Social Programs: The Case of the U.S. National JTPA Study." University of Chicago Harris School of Public Policy Working Paper 9203. 26

Ilut, Cosmin, and Rosen Vlachev. 2017. "Economic Agents as Imperfect Problem Solvers." 6

Jovanovic, Boyan, and Rafael Rob. 1990. "Long Waves and Short Waves: Growth Through Intensive and Extensive Search." *Econometrica*, 58(6): 1391–1409. 5, 6

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review*, 101(6): 2590–2615. 6

Katz, L. 1953. "A New Status Index Derived From Sociometric Analysis." *Psychometrika*, 18: 39–43. 28

Keeney, Ralph L., and Howard Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* New York:Wiley. 2

Ke, T. Tony, and J. Miguel Villas-Boas. 2019. "Optimal Learning Before Choice." *Journal of Economic Theory*, 180: 383–437. 6

Klabjan, Diego, Wojciech Olszewski, and Asher Wolinsky. 2014. "Attributes." *Games and Economic Behavior*, 88: 190–206. 5

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." *Journal of Political Economy*, 74: 132–157. 2

Liang, Annie, Xiaosheng Mu, and Vasilis Syrgkanis. 2020. "Dynamically Aggregating Diverse Information." 5

MacCrimmon, K. R. 1968. "Decisionmaking Among Multiple-Attribute Alternatives: A Survey and Consolidated Approach." RAND Corporation Memorandum RM-4823-ARPA. 2

Matheron, Georges. 1963. "Principles of Geostatistics." *Economic Geology*, 58(5): 1246–1266. 5

Mayskaya, Tatiana. 2019. "Dynamic Choice of Information Sources." 6

Moscarini, Giuseppe, and Lones Smith. 2001. "The Optimal Level of Experimentation." *Econometrica*, 69(6): 1629–1644. 6

Neeman, Zvika. 1995. "On Determining the Importance of Attributes with a Stopping Problem." *Mathematical Social Sciences*, 29: 195–212. 5

Olsen, Robert B., Larry L. Orr, Stephen H. Bell, and Elizabeth A. Stuart. 2013. "External Validity in Policy Evaluations that Choose Sites Purposively." *Journal of Policy Analysis and Management*, 32(1): 107–121. 4, 24

Olszewski, Wojciech, and Asher Wolinsky. 2016. "Search for an Object with Two Attributes." *Journal of Economic Theory*, 161: 145–160. 5

Owen, Donald Bruce. 1980. "A Table of Normal Integrals." *Communications in Statistics - Simulation and Computation*, 9(4): 389–419. 41

Patton, Michael Quinn. 2014. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice.* SAGE Publications. 24

Rasmussen, Carl Edward, and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning.* Cambridge, Massachusetts:The MIT Press. 5, 7

Rayo, Luis, and Ilya Segal. 2010. "Optimal Information Disclosure." *Journal of Political Economy*, 118(5): 949–987. 6

Rogers-Dillon, Robin H. 2004. *The Welfare Experiments: Politics and Policy Evaluation.* Stanford University Press. 24

Sanjurjo, Adam. 2017. "Search with Multiple Attributes: Theory and Empirics." 5

Sher, Itai. 2011. "Credibility and determinism in a game of persuasion." *Games and Economic Behavior*, 71: 409–419. 6

Vivalt, Eva. Forthcoming. ""How Much Can We Generalize from Impact Evaluations?"." *Journal of the European Economics Association.* 4, 25

Watts, H. W., J. K. Peck, and M. Taussig. 1977. "Site Selection, Representativeness of the Sample, and Possible Attrition Bias." In *The New Jersey Income Maintenance Experiment.* Vol. III, , ed. Harold W. Watts and Albert Rees. New York: Academic Press. 24

Weiss, Carol H. 1993. "Where Politics and Evaluation Research Meet." *Evaluation Practice*, 14(1): 93–106. 24

A Preliminaries

A.1 Sample-path continuity of Gaussian processes

Definition 5. The covariance function $\sigma(a, a')$ is *symmetric* if $\sigma(a, a') = \sigma(a', a)$ for any $(a, a') \in \mathcal{A}^2$. It is *positive semidefinite* if for any $f \in \mathcal{L}^2(\mathcal{A})$,

$$\int_{\mathcal{A}} \int_{\mathcal{A}} \sigma(a, a') f(a) f(a') \, \mathrm{d}a \, \mathrm{d}a' \geq 0.$$

Alternatively, σ is positive semidefinite if and only if for any $n \in \mathbb{N}$, any collection $(a_1, \ldots, a_n) \in \mathcal{A}^n$ and any vector $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$,

$$\sum_{i=1}^{n} \sum_{i=1}^{n} x_i x_j \sigma(a_i, a_j) \geq 0.$$

Definition 6 (Sample-path continuity)**.** Let Ω be an outcome space. A process f is *sample-path continuous* at $a_0 \in \mathcal{A}$ if, for almost all $\omega \in \Omega$, $a \to a_0$ implies $f(\omega, a) \to f(\omega, a_0)$. The process is sample-path continuous if it is sample-path continuous at any $a_0 \in \mathcal{A}$.

It is straightforward that the continuity of μ is necessary for sample path continuity. Hence without loss the following propositions normalize μ to zero.

Proposition A.1 (Sufficient conditions for sample-path continuity)**.** *Let $\mathcal{A} = [0,1]^d, d \geq 1$ and f be a zero-mean Gaussian process with covariance σ. If there exist $\beta, K > 0$ such that $\sigma(a, a) + \sigma(a', a') - 2\sigma(a, a') \leq K|a - a'|^{d+\beta}$ for all $a, a' \in \mathcal{A}$ then f has a modification on $\mathcal{A}$ that is sample-path continuous.*

Proof. By Kolmorogov's Continuity Theorem, such a continuous modification exists if there exist $\alpha, \beta, K > 0$ such that $\mathbb{E}[|f(a) - f(a')|^{\alpha}] \leq K|a - a'|^{d+\beta}$ $\forall a, a' \in \mathcal{A}$. Letting $\alpha = 2$ and using the fact that $\mu(a) = 0$ for any $a \in \mathcal{A}$, the LHS becomes

$$\mathbb{E}[|f(a) - f(a')|^2] = \mathbb{E}[f(a)^2] + \mathbb{E}[f(a')^2] - 2\mathbb{E}[f(a)f(a')] = \sigma(a, a) + \sigma(a', a') - 2\sigma(a, a').$$

If the RHS is less than $K|a - a'|^{d+\beta}$ for some $\beta, K > 0$ then a continuous modification of f exists. ☐

Proposition A.2 (Continuity of σ)**.** *Let $\mu(a) = 0$ for all $a \in \mathcal{A}$. If f is sample-path continuous at $a_1, a_2 \in \mathcal{A}$, then $\sigma(a, a')$ is continuous at $a = a_1, a' = a_2$.*

Proof for proposition A.2. First, if f is sample path continuous at some $a_1, a_2 \in \mathcal{A}$, then

$$\lim_{a \to a_1} \mathbb{E}\left[(f(a) - f(a_1))^2\right] = \lim_{a \to a_2} \mathbb{E}\left[(f(a) - f(a_2))^2\right] = 0.$$

Therefore, f is mean-square continuous at $a = a_1, a_2$. Also, note that $\sigma(a, a') - \sigma(a_1, a_2)$ $= (\sigma(a, a') - \sigma(a_1, a')) + (\sigma(a_1, a') - \sigma(a_1, a_2))$. But,

$$|\sigma(a, a') - \sigma(a_1, a')| = |\mathbb{E}\left[(f(a) - f(a_1))\, f(a')\right]| \leq \sqrt{\mathbb{E}\left[(f(a) - f(a_1))^2\right]}\sqrt{\mathbb{E}\left[f(a')^2\right]}$$

where the inequality follows from applying the Cauchy-Schwarz inequality for expectations. Because f is mean-square continuous at a_1, the first term of the RHS vanishes to zero as

$a \to a_1$. Hence, $\lim_{a\to a_1} |\sigma(a,a') - \sigma(a_1,a')| = 0$. By a similar logic, $\lim_{a'\to a_2} |\sigma(a_1,a') - \sigma(a_1,a_2)| = 0$. Combining these two observations, we conclude that $\sigma(a,a')$ is continuous at $a = a_1, a' = a_2$, i.e. $\lim_{a\to a_1} \lim_{a'\to a_2} \sigma(a,a') = \sigma(a_1,a_2)$. □

A.2 Proofs for section 2.2

Proof for lemma 2.1. (i) The joint distribution is given by

$$\begin{pmatrix} f(\hat{a}) \\ f(\mathbf{a}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu(\hat{a}) \\ \mu(\mathbf{a}) \end{pmatrix}, \begin{pmatrix} \sigma(\hat{a},\hat{a}) & \Sigma(\hat{a},\mathbf{a}) \\ \Sigma(\hat{a},\mathbf{a})^\top & \Sigma(\mathbf{a}) \end{pmatrix}\right)$$

where $\Sigma(\hat{a},\mathbf{a}) = \begin{pmatrix} \sigma(a_1,\hat{a}) & \dots & \sigma(a_n,\hat{a}) \end{pmatrix}$ and $\Sigma(\mathbf{a})$ is the sample covariance matrix. Hence, $\mathbb{E}[f(\hat{a}) \mid \mathbf{a}, f(\mathbf{a})] = \mu(\hat{a}) + \Sigma(\hat{a},\mathbf{a})[\Sigma(\mathbf{a})]^{-1}\left(f(\mathbf{a}) - \mu(\mathbf{a})\right)$.
(ii) Sample realizations are observed perfectly. Hence for any $(\mathbf{a}, f(\mathbf{a}))$ and any $a_j \in \mathbf{a}$, $\mathbb{E}[f(a_j) \mid \mathbf{a}, f(\mathbf{a})] = f(a_j)$. Therefore, $\tau_j(a_j;\mathbf{a}) = 1$ and $\tau_m(a_j;\mathbf{a}) = 0$ for $m \neq j$. □

Proof for lemma 2.2. Let $\mu(a) = 0$ for all $a \in \mathcal{A}$. Applying lemma 2.1, we obtain

$$\begin{aligned} \nu^i(\mathbf{a}, f(\mathbf{a})) &= \int_{\mathcal{A}} \mathbb{E}[f(a) \mid \mathbf{a}, f(\mathbf{a})]\omega_i(a)\,\mathrm{d}a \\ &= \int_{\mathcal{A}} \left(\tau_1(a;\mathbf{a})f(a_1) + \dots \tau_n(a;\mathbf{a})f(a_n)\right)\omega_i(a)\,\mathrm{d}a = \sum_{j=1}^{n} f(a_j)\left(\int_{\mathcal{A}} \tau_j(a;\mathbf{a})\omega_i(a)\,\mathrm{d}a\right). \end{aligned}$$

Because for any $j \in \{1,\dots,k\}$, $\nu^i(\mathbf{a}, f(\mathbf{a})) - \nu_0^i$ and $f(a_j) - \mu(a_j)$ are centered at zero, for any arbitrary mean $\mu : [0,1] \to \mathbb{R}$

$$\nu^i(\mathbf{a}, f(\mathbf{a})) - \nu_0^i = \sum_{j=1}^{n}\left(\int_{\mathcal{A}} \tau_j(a;\mathbf{a})\omega_i(a)\,\mathrm{d}a\right)(f(a_j) - \mu(a_j)).$$

B Proofs for section 3

□

B.1 Section 3.1: Characterization

Proof for theorem 3.1. (i) Suppose the single-player sample $\mathbf{a}^* \in \mathcal{A}_{k-1}$. By assumption 2, $\mathrm{Var}(v \mid \mathbf{a}^*, f(\mathbf{a}^*)) > 0$. Hence, there exists $a \in \mathcal{A}$ such that $a \notin \mathbf{a}^*$ and $\mathrm{Cov}(v, f(a) \mid f(\mathbf{a}^*)) > 0$. Consider the new sample $\tilde{\mathbf{a}} = \mathbf{a}^* \cup a$. It is immediate that $\psi(\tilde{\mathbf{a}}) > \psi(\mathbf{a}^*)$, which contradicts the optimality of $\mathbf{a}^*$.

(ii) Fix $\mathbf{a} \in \mathcal{A}_k$. Given $(\mathbf{a}, f(\mathbf{a}))$, lemma 2.2 provides the expression for $\nu(\mathbf{a}, f(\mathbf{a}))$. Prior to observing $f(\mathbf{a})$, the distribution of $\nu(\mathbf{a}, f(\mathbf{a}))$ is Gaussian. Note that $\mathbb{E}[\nu(\mathbf{a}, f(\mathbf{a}))] =$

ν_0^i as $\mathbb{E}[f(a_j)] = \mu(a_j)$ for any $a_j \in \mathbf{a}$. The posterior variance is given by

$$\begin{aligned}
\psi^2(\mathbf{a}) &:= \mathrm{Var}\left[\nu(\mathbf{a}, f(\mathbf{a}))\right] \\
&= \mathrm{Cov}\left[\sum_{j=1}^{k} \tau_j(\mathbf{a})\left(f(a_j) - \mu(a_j)\right), \sum_{m=1}^{k} \tau_m(\mathbf{a})\left(f(a_m) - \mu(a_m)\right)\right] \\
&= \sum_{j=1}^{k} \tau_j(\mathbf{a}) \left(\sum_{m=1}^{k} \tau_m(\mathbf{a}) \mathrm{Cov}\left(f(a_j) - \mu(a_j), f(a_m) - \mu(a_m)\right) \right) \\
&= \sum_{j=1}^{k} \tau_j(\mathbf{a}) \left(\sum_{m=1}^{k} \tau_m(\mathbf{a}) \sigma(a_j, a_m) \right).
\end{aligned}$$

The player adopts if and only if $\nu(\mathbf{a}, f(\mathbf{a})) \geq r$. His value from sampling $\mathbf{a}$ is

$$\begin{aligned}
V(\mathbf{a}) &= \Pr\left[\nu(\mathbf{a}, f(\mathbf{a}))) \geq r\right] \mathbb{E}\left[\nu(\mathbf{a}, f(\mathbf{a}))\right] \mid \nu(\mathbf{a}, f(\mathbf{a}))) \geq r) + r \Pr\left[\nu(\mathbf{a}, f(\mathbf{a}))) < r\right] \\
&= \Phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right) \left(\nu_0 + \psi(\mathbf{a}) \frac{\phi\left(\frac{r-\nu_0}{\psi(\mathbf{a})}\right)}{\Phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right)} \right) + r\left(1 - \Phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right)\right) \\
&= r + (\nu_0 - r)\Phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right) + \psi(\mathbf{a})\phi\left(\frac{\nu_0 - r}{\psi(\mathbf{a})}\right).
\end{aligned}$$

But $V(\mathbf{a})$ is strictly increasing in $\psi(\mathbf{a})$ for any ν_0 and r. Hence, any single-player sample maximizes $\psi(\mathbf{a})$.

(iii) For any $\mathbf{a} \in \mathcal{A}_k$, $\psi(\mathbf{a})$ is independent of μ, ν_0 and r. Hence, part (ii) implies that the set of single-player samples is the same for (μ, ν_0, r).

□

Proof for Proposition 3.2. Consider sequential sampling in $m \geq 1$ batches of respective sizes $(q_1, \ldots, q_m)$. First, note that the posterior variance attained by an optimal sequential sample is at least as high as that of an optimal simultaneous sample. Fix an optimal sequential sample $\mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$ where $\mathbf{a}_j = (a_j^1, \ldots, a_j^{q_j})$. Let ν_j denote the posterior after $\mathbf{a}^j := (\mathbf{a}_1, \ldots, \mathbf{a}_j)$ have been sampled.

Claim 1. *For any j, $(\mathbf{a}_{j+1}, \ldots, \mathbf{a}_m)$ is independent of the sequence of posteriors $\nu_1, \ldots, \nu_j$.*

Proof. The proof proceeds by induction. First, let $j = m - 1$. As in theorem 3.1(ii), agent chooses $\mathbf{a}_m$ to maximize

$$V(\mathbf{a}_m) = r + (\nu_{m-1} - r)\Phi\left(\frac{\nu_{m-1} - r}{\psi(\mathbf{a}_m)}\right) + \psi(\mathbf{a}_m)\phi\left(\frac{(\nu_{m-1} - r}{\psi(\mathbf{a}_m)}\right)$$

So, given $\mathbf{a}^{m-1}$ the optimal continuation sample $\mathbf{a}_m$ is independent of ν_{m-1} and therefore of $f(\mathbf{a}^{m-1})$. Next, suppose that for some $j < m-1$, the sequence of optimal continuation samples $\mathbf{a}_{j+1}, \ldots, \mathbf{a}_m$ depends on $\mathbf{a}^j$ but not on $f(\mathbf{a}^j)$. Consider now the choice of $\mathbf{a}_j$. For any $\mathbf{a}_j$, there exists optimal continuation subsamples $\mathbf{a}_{j+1}(\mathbf{a}_j), \ldots, \mathbf{a}_m(\mathbf{a}_j)$ that –by the inductive step – do not depend on $f(\mathbf{a}_j)$. With some abuse of notation, let the posterior variance attained by $\mathbf{a}^j \cup \mathbf{a}_j \cup \mathbf{a}_{j+1}(\mathbf{a}_j) \cup \ldots \cup \mathbf{a}_m(\mathbf{a}_j)$ be denoted by $\psi(\mathbf{a}_j)$. The agent chooses a_j so as to maximize

$$V(\mathbf{a}_m) = r + (\nu_{j-1} - r)\Phi\left(\frac{\nu_{j-1} - r}{\psi(\mathbf{a}_j)}\right) + \psi(\mathbf{a}_j)\phi\left(\frac{(\nu_{j-1} - r}{\psi(\mathbf{a}_j)}\right).$$

Hence, $\mathbf{a}_j$ maximizes $\psi(\mathbf{a}_j)$. □

By claim 1, any optimal sequential sample can be implemented under simultaneous sampling: sequential sampling cannot attain a strictly higher posterior variance. □

B.2 Section 3.2: Ornstein-Uhlenbeck covariance

Lemma B.1 (Nearest-neighbor extrapolation)**.** *Fix sample* $\mathbf{a} = (a_1, \ldots, a_k)$ *where* $a_1 < \ldots < a_k$*. For* $a \in (0, a_1)$*,* $\tau_1(a, \mathbf{a}) = e^{-(a_1 - a)/\ell}$ *and* $\tau_j(a, \mathbf{a}) = 0$ *for* $j \geq 2$*. For* $a \in (a_k, 1)$*,* $\tau_k(a, \mathbf{a}) = e^{-(a - a_k)/\ell}$ *and* $\tau_j(a, \mathbf{a}) = 0$ *for* $j \leq k-1$*. For* $a \in (a_j, a_{j+1})$ *where* $j = 1, \ldots, k-1$*:*

$$\tau_n(a, \mathbf{a}) = \begin{cases} \operatorname{csch}\left(\frac{a_{j+1} - a_j}{\ell}\right) \sinh\left(\frac{a_{j+1} - a}{\ell}\right) & \text{if } n = j \\ \operatorname{csch}\left(\frac{a_{j+1} - a_j}{\ell}\right) \sinh\left(\frac{a - a_j}{\ell}\right) & \text{if } n = j+1 \\ 0 & \text{if } n \neq j, j+1. \end{cases} \tag{14}$$

Moreover, extrapolation between $f(a_j)$ *and* $f(a_{j+1})$ *is generically non-linear.*

Proof for lemma B.1. Let $p(a) := e^{a/\ell}$ and $q(a) := e^{-a/\ell}$. Covariance σ_{OU} can be decomposed into $\sigma_{OU}(a, a') = p(a)q(a')\mathbb{1}[a \leq a'] + p(a')q(a)\mathbb{1}[a' < a]$. Fix an arbitrary sample $\mathbf{a} = (a_1, \ldots, a_k)$ s.t. $a_1 < \ldots < a_k$. By theorem 3 in Ding and Zhang (2018), $\Sigma(\mathbf{a})$ is invertible because (1) $p(a)q(a') - p(a')q(a) < 0$ for any $a < a'$, and (2) $p(a)q(a') > 0$ for any $a, a' \in [0, 1]$. By theorem 1 in Ding and Zhang (2018), $\Sigma^{-1}(\mathbf{a})$ is symmetric tridiagonal. Applying theorem 2 and letting $d_j := a_{j+1} - a_j$, the diagonal entries of $\Sigma^{-1}(\mathbf{a})$ are

$$p_{jj} := \left(\Sigma^{-1}(\mathbf{a})\right)_{j,j} = \begin{cases} \dfrac{e^{d_1/\ell}}{2\sinh(d_1/\ell)} & \text{if } j = 1 \\ \dfrac{1}{2}\left(\coth(d_{j-1}/\ell) + \coth(d_j/\ell)\right) & \text{if } 2 \leq j \leq k-1 \\ \dfrac{e^{d_{k-1}/\ell}}{2\sinh(d_{k-1}/\ell)} & \text{if } j = k, \end{cases}$$

The off-diagonal entries of $\Sigma(\mathbf{a})$ are given by

$$p_{j-1,j} := \left(\Sigma^{-1}(\mathbf{a})\right)_{j-1,j} = \frac{-1}{2\sinh(d_{j-1}/\ell)}.$$

The tridiagonal nature of the precision matrix simplifies the expression for $\tau_j(a,\mathbf{a})$ to

$$\tau_j(a,\mathbf{a}) = \sigma(a,a_j)p_{jj} + \sigma(a,a_{j-1})p_{j-1,j} + \sigma(a,a_{j+1})p_{j,j+1}.$$

Replacing $\sigma(a,a') = e^{-|a-a'|/\ell}$ and the expressions for p_{ij}, straightforward algebra yields (14).

Note that for any $a \in (a_j, a_{j+1})$, $\tau_j(a,\mathbf{a}) + \tau_{j+1}(a,\mathbf{a}) < 1$. If extrapolation were linear, the coefficients would be $\tau_j(a,\mathbf{a}) = (a_{j+1}-a)/(a_{j+1}-a_j)$ and $\tau_{j+1}(a,\mathbf{a}) = (a-a_j)/(a_{j+1}-a_j)$. Therefore, extrapolation is non-linear. □

Proof for lemma 3.3. Let $a_0 = 0$ and $a_{k+1} = 1$, and suppose $\mathbf{a} = (a_1,\ldots,a_k) \in \mathcal{A}_k$ where $0 \le a_1 < \ldots < a_k \le 1$. For any $j = 1,\ldots,k$

$$\tau_j(\mathbf{a}) = \int_{a_{j-1}}^{a_j} \tau_j(a,\mathbf{a})\,\mathrm{d}a + \int_{a_j}^{a_{j+1}} \tau_j(a,\mathbf{a})\,\mathrm{d}a, \tag{15}$$

where $\tau_j(a,\mathbf{a})$ is derived in lemma B.1. Substituting (14) into (15) gives the result. □

Proof of proposition 3.4.
(i) We first argue that any optimal sample is interior, i.e. given $k \in \mathbb{N}$, $a_1^* > 0$ and $a_k^* < 1$. Suppose by way of contradiction that $a_1^* = 0$. For any sample $\mathbf{a} \in \mathcal{A}_k$, taking the partial derivative with respect to the leftmost attribute, we obtain

$$\left.\frac{\partial\psi^2(\mathbf{a})}{\partial a_1}\right|_{a_1=0} = 2\ell e^{-2a_1/\ell}\left(2e^{a_1/\ell}-1\right) - 2\ell\,\mathrm{sech}^2\left(\frac{a_2-a_1}{2\ell}\right)\Bigg|_{a_1=0} = 2\ell\left(1-\mathrm{sech}^2\left(\frac{a_2}{2\ell}\right)\right) > 0$$

for any $a_2 > 0$ and $\ell > 0$. Hence, the posterior variance strictly increases if a_1^* increases incrementally to a strictly positive level, which contradicts the optimality of a_1^*. Therefore $a_1^* > 0$. By a similar argument, $a_k^* < 1$.

Second we show that for any $j \in \{2,\ldots,k\}$ distance $a_j^* - a_{j-1}^*$ is constant. For any $a_2^*,\ldots,a_{k-1}^*$ the first-order condition is given by

$$\frac{\partial\psi^2(\mathbf{a})}{\partial a_j^*} = 2\ell\left(\mathrm{sech}^2\left(\frac{a_j^*-a_{j-1}^*}{2\ell}\right) - \mathrm{sech}^2\left(\frac{a_{j+1}^*-a_j^*}{2\ell}\right)\right) = 0.$$

Hence $a_j^* - a_{j-1}^* = a_{j+1}^* - a_j^*$ for any $j = 2,\ldots,k-1$. This implies that $a_j^* - a_{j-1}^* =$

$(1 - a_1^* - a_k^*)/n$ for any such j and therefore $\tau_2^* = \ldots = \tau_{k-1}^*$.

Suppose, by way of contradiction, that $\tau_1^* > \tau_k^*$. By the fact that $a_2^* - a_1^* = a_k^* - a_{k-1}^*$, it must be that $a_1^* > 1 - a_k^*$. Consider a slight modification of the sample from $\mathbf{a}^*$ to $\tilde{\mathbf{a}}$ so that $\tilde{a}_j = a_j^* - \epsilon$ for any $j = 1, \ldots, k$ and for ϵ sufficiently small. By construction $\tilde{a}_j - \tilde{a}_{j-1} = a_j^* - a_{j-1}^*$ for all $j = 2, \ldots, k$, hence $\tilde{\tau}_j = \tau_j^*$ for any such j. Let $\tilde{\tau}_1 = \tau_1^* - \epsilon_1$ and $\tilde{\tau}_k = \tau_k^* + \epsilon_k$, where $\epsilon_1 < \epsilon_k$ by the concavity of $1 - e^{-d/\ell}$ in d and the fact that $a_1^* > 1 - a_k^*$. Then,

$$\tilde{\tau}_1 \sum_{j=2}^{k-1} \tilde{\tau}_j + \tilde{\tau}_k \sum_{j=2}^{k-1} \tilde{\tau}_j > \tau_1^* \sum_{j=2}^{k-1} \tau_j^* + \tau_k^* \sum_{j=2}^{k-1} \tau_j^*$$

and

$$\begin{aligned}(\tau_1^* - \epsilon_1)^2 + (\tau_k^* + \epsilon_k)^2 + 2(\tau_1^* - \epsilon_1)(\tau_k^* + \epsilon_k)& \\ = \tau_1^* + \tau_k^* + 2\tau_1^*\tau_k^* + 2(\tau_1^* + \tau_k^*)(\epsilon_k - \epsilon_1) + (\epsilon_1 - \epsilon_k)^2 &> 0.\end{aligned}$$

The rest of the terms are shared between $\psi^2(\tilde{\mathbf{a}})$ and $\psi^2(\mathbf{a}^*)$. Hence, $\psi^2(\tilde{\mathbf{a}}) > \psi^2(\mathbf{a}^*)$. This contradicts the optimality of $\mathbf{a}^*$; hence $\tau_1^* \leq \tau_k^*$. By a similar argument we establish that $\tau_1^* \geq \tau_k^*$. Hence $\tau_1^* = \tau_k^*$.

(ii) By the fact that $\tau_j^* = \tau_i^*$ for any $i, j \in \{1, \ldots, k\}$ we observe that $a_1^* = 1 - a_k^*$. Hence $a_j^* - a_{j-1}^* = (1 - 2a_1^*)/(k-1)$ for any $j = 2, \ldots, k$. This implies that

$$a_j^* = a_1^* + \frac{j-1}{k-1}(1 - 2a_1^*)$$

and hence the optimal sample is symmetric around 1/2.

The first order condition with respect to a_1^* pins down the entire optimal sample. Hence,

$$2\ell e^{-2a_1^*/\ell}\left(2e^{a_1^*/\ell} - 1\right) - 2\ell \operatorname{sech}^2\left(\frac{1 - 2a_1^*}{2\ell(k-1)}\right) = 0 \quad \Leftrightarrow \quad 2e^{a_1^*/\ell} - 1 = e^{2a_1^*/\ell} \operatorname{sech}^2\left(\frac{1 - 2a_1^*}{2\ell(k-1)}\right).$$

By a known trigonometric identity, $\operatorname{sech}^2(x) = 1 - \tanh^2(x) = (1 - \tanh(x))(1 + \tanh(x))$. Moreover, $\tau_1^* = \tau_2^*$ gives us an expression for $1 - \tanh\left((1 - 2a_1^*)/(2\ell(k-1))\right)$. Substituting that into the FOC and simplifying, we obtain

$$1 - e^{-a_1^*/\ell} = \tanh\left(\frac{1 - 2a_1^*}{2\ell(k-1)}\right).$$

Note that this also implies the sufficiency of $\tau_1^* = \ldots = \tau_k^*$ for optimality: if a_1^* guarantees that all observations are weighted equally, then a_1^* is the optimal leftmost attribute. □

C Proofs for section 4

Proof for lemma 4.1. Letting $\nu^i(\mathbf{a}, f(\mathbf{a})) =: \nu^i(\mathbf{a})$ and $\rho(\mathbf{a})$ denote the correlation between $\nu_P(\mathbf{a})$ and $\nu_A(\mathbf{a})$, the joint distribution is Gaussian:

$$\begin{pmatrix} \nu^P(\mathbf{a}) \\ \nu^A(\mathbf{a}) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \nu_0^P \\ \nu_0^A \end{pmatrix}, \begin{pmatrix} \psi_P^2(\mathbf{a}) & \rho(\mathbf{a})\psi_A(\mathbf{a})\psi_P(\mathbf{a}) \\ \rho(\mathbf{a})\psi_A(\mathbf{a})\psi_P(\mathbf{a}) & \psi_P^2(\mathbf{a}) \end{pmatrix} \right).$$

Claim 2. *For any $r_P \in \mathbb{R}$,*

$$f(\nu^A(\mathbf{a}) \mid \nu^P(\mathbf{a}) \geq r_P) = \frac{\phi\left(\frac{\nu^A(\mathbf{a})-\nu_0^A}{\psi_A(\mathbf{a})}\right)}{\psi_A(\mathbf{a})\Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right)} \Phi\left(\frac{\nu_0^P + \rho(\mathbf{a})\frac{\psi_P(\mathbf{a})}{\psi_A(\mathbf{a})}(\nu^A(\mathbf{a}) - \nu_0^A) - r_P}{\psi_P(\mathbf{a})\sqrt{1-\rho(\mathbf{a})^2}}\right).$$

Proof. Let x_1, x_2 be jointly Gaussian with means μ_1, μ_2, variances σ_1^2, σ_2^2 and covariance σ_{12}. Let f_1, f_2 and F_1, F_2 denote their respective pdf and cdf. Then,

$$\begin{aligned} f(x_1 \mid x_2 \geq \bar{x}) &= \frac{1}{1 - F_2(\bar{x})} \Pr(x_2 \geq \bar{x}) f(x_1 \mid x_2 \geq \bar{x}) \\ &= \frac{1}{1 - F_2(\bar{x})} \int_{\bar{x}}^{\infty} f(x_2 \mid x_1) f_1(x_1) dx_2 \\ &= \frac{f_1(x_1)}{1 - F_2(\bar{x})}(1 - F_{x_2|x_1}(\bar{x})). \end{aligned}$$

The first line multiplies and divides by $\Pr(x_2 \geq \bar{x})$. The second line rewrites $\Pr(x_2 \geq \bar{x}) f(x_1 \mid x_2 \geq \bar{x})$ using the joint density and the observation that $f(x_1, x_2) = f(x_2 \mid x_1) f_1(x_1)$. The last two lines use the conditional distribution of $x_2 \mid x_1$. But,

$$x_2 \mid x_1 \sim \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1-\rho^2)\sigma_2^2\right)$$

and $\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$. Therefore, we can substitute in the expression for $F_{x_2|x_1}$ to obtain

$$f(x_1 \mid x_2 \geq \bar{x}) = \frac{f_1(x_1)}{1 - F_2(\bar{x})}\left(1 - \Phi\left(\frac{\bar{x} - \mu_2 - \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)}{\sigma_2\sqrt{1-\rho^2}}\right)\right).$$

Switching back to our variables of interest, let $x_1 := \nu^A(\mathbf{a}) \sim \mathcal{N}(\nu_0^A, \psi_A^2(\mathbf{a}))$, $x_2 := \nu^P(\mathbf{a}) \sim \mathcal{N}(\nu_0^P, \psi_P^2(\mathbf{a}))$ and $\bar{x} := r_P$. Therefore,

$$f(\nu^A(\mathbf{a})|\nu^P(\mathbf{a}) \geq r_P) = \frac{\phi\left(\frac{\nu^A(\mathbf{a})-\nu_0^A}{\psi_A(\mathbf{a})}\right)}{\psi_A(\mathbf{a})\left(1 - \Phi\left(\frac{r_P - \nu_0^P}{\psi_P(\mathbf{a})}\right)\right)}\left(1 - \Phi\left(\frac{r_P - \nu_0^P - \rho(\mathbf{a})\frac{\psi_P(\mathbf{a})}{\psi_A(\mathbf{a})}(\nu^A(\mathbf{a}) - \nu_0^A)}{\psi_P(\mathbf{a})\sqrt{1-\rho(\mathbf{a})^2}}\right)\right).$$

□

Using the claim, observe that:

$$\Pr(\nu^P(\mathbf{a}) \geq r_P)\mathbb{E}[\nu^A(\mathbf{a}) \mid \nu^P(\mathbf{a}) \geq r_P] = \Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) \int_{-\infty}^{\infty} \nu^A(\mathbf{a}) f(\nu^A(\mathbf{a})|\nu^P(\mathbf{a}) \geq r_P)\, \mathrm{d}\nu^A(\mathbf{a})$$
$$= \int_{-\infty}^{\infty} \frac{\nu^A(\mathbf{a})}{\psi_A(\mathbf{a})}\phi\left(\frac{\nu^A(\mathbf{a}) - \nu_0^A}{\psi_A(\mathbf{a})}\right)\Phi\left(\frac{\nu_0^P + \rho(\mathbf{a})\frac{\psi_P(\mathbf{a})}{\psi_A(\mathbf{a})}(\nu^A(\mathbf{a}) - \nu_0^A) - r_P}{\psi_P(\mathbf{a})\sqrt{1-\rho(\mathbf{a})^2}}\right) \mathrm{d}\nu^A(\mathbf{a})$$
$$= \int_{-\infty}^{\infty} \left(x\psi_A(\mathbf{a}) + \nu_0^A\right)\phi(x)\Phi\left(\frac{\nu_0^P + \rho(\mathbf{a})\psi_P(\mathbf{a})x - r_P}{\psi_P(\mathbf{a})\sqrt{1-\rho^2(\mathbf{a})}}\right) \mathrm{d}x,$$

where in the last line $x := \frac{\nu^A(\mathbf{a}) - \nu_0^A}{\psi_A(\mathbf{a})}$. From Owen (1980), we have the following Gaussian identities (respectively, numbered 10,010.8 and 10,011.1 in Owen (1980)):

$$\int_{-\infty}^{\infty} \phi(x)\Phi(a+bx)dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right), \quad \int_{-\infty}^{\infty} x\phi(x)\Phi(a+bx)dx = \frac{b}{\sqrt{1+b^2}}\phi\left(\frac{a}{\sqrt{1+b^2}}\right).$$

Letting $a := (\nu_0^P - r_P)/(\psi_P(\mathbf{a})\sqrt{1-\rho^2(\mathbf{a})})$ and $b := \rho(\mathbf{a})/\sqrt{1-\rho^2(\mathbf{a})}$,

$$\Pr(\nu^P(\mathbf{a}) \geq r_P)\mathbb{E}[\nu^A(\mathbf{a}) \mid \nu^P(\mathbf{a}) \geq r_P] = \nu_0^A\Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) + \rho(\mathbf{a})\psi_A(\mathbf{a})\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right).$$

Therefore, the agent's payoff from sample $\mathbf{a}$ simplifies to

$$V_A(\mathbf{a}) = \Pr(\nu^P(\mathbf{a}) < r_P)r_A + \nu_0^A\Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) + \rho(\mathbf{a})\psi_A(\mathbf{a})\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right)$$
$$= r_A + (\nu_0^A - r_A)\Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) + \rho(\mathbf{a})\psi_A(\mathbf{a})\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right).$$

□

Proof for proposition 4.2. Because $\omega_A = \omega_P$, $\psi_i(\mathbf{a}) = \psi(\mathbf{a})$ and $\nu_0^i = \nu_0$ for $i = A, P$. From theorem 4.1, the agent's payoff from $\mathbf{a} \in \mathcal{A}_k$ simplifies to

$$V_A(\mathbf{a}) = r_A + (\nu_0 - r_A)\Phi\left(\frac{\nu_0 - r_P}{\psi(\mathbf{a})}\right) + \psi(\mathbf{a})\phi\left(\frac{\nu_0 - r_P}{\psi(\mathbf{a})}\right).$$

V_A is increasing in $\psi(\mathbf{a})$ if and only if

$$\psi^2(\mathbf{a}) \geq (r_P - r_A)(\nu_0 - r_P). \tag{16}$$

Hence for any (ν_0, r_P, r_A), V_A is either strictly increasing or single-troughed in ψ.

(i) Suppose first that $r_A < r_P$. Fix $k \in \mathbb{N}$ and let ψ_k^2 denote the posterior variance of a single-player sample. For $\nu_0 \leq r_P$, the RHS of (16) is strictly negative, hence V_A is strictly increasing in ψ. Therefore the optimal sample is the single-player sample. If $\nu_0 > r_P$, V_A is single-troughed in ψ, with the trough at $\psi^2 = (r_P - r_A)(\nu_0 - r_P)$.

Therefore, the agent chooses either $\psi = 0$ or $\psi = \psi_k$. The agent prefers $\psi = \psi_k$ to $\psi = 0$ if and only if

$$\lambda\left(\frac{\nu_0 - r_P}{\psi_k}\right) := \frac{\phi\left(\frac{\nu_0 - r_P}{\psi_k}\right)}{1 - \Phi\left(\frac{\nu_0 - r_P}{\psi_k}\right)} > \frac{\nu_0 - r_A}{\psi_k} = \frac{\nu_0 - r_P}{\psi_k} + \frac{r_P - r_A}{\psi_k} \tag{17}$$

where λ is the inverse Mill's ratio. By (16) at $\nu_0 = r_P$, agent prefers ψ_k: hence (17) implies $\lambda(0) > (r_P - r_A)/\psi_k$. But $\lambda'(\cdot) \in (0,1)$, strictly increasing, and $\lambda\left(\frac{\nu_0 - r_P}{\psi}\right) \to \frac{\nu_0 - r_P}{\psi}$ for ν_0 sufficiently high. Hence, there exists $\bar{\nu}_0 > r_P$ such that (17) holds for any $\nu_0 < \bar{\nu}_0$.

The argument for $r_P < r_A$ follows a similar argument and is therefore omitted.

(ii) First, observe that $\psi_{k'} \geq \psi_k$ for any $k' > k$. Let $\bar{\nu}_0(k)$ denote the corresponding cutoff from part (i). At $\nu_0 = \bar{\nu}_0(k)$ the agent is indifferent between providing $\psi = 0$ and $\psi = \psi_k$. Hence he strictly prefers higher posterior variance $\psi_{k'}$. This implies that $\bar{\nu}_0(k') \geq \bar{\nu}_0(k)$. Moreover, note that

$$\psi_k^2 \to \int_{\mathcal{A}}\int_{\mathcal{A}} \sigma(a,a')\omega_i(a)\omega_i(a')\,\mathrm{d}a\,\mathrm{d}a' := \psi_\infty^2 < \infty$$

as $k \to \infty$. For any ν_0 such that $\psi_\infty^2 < (r_P - r_A)(\nu_0 - r_P)$, $\psi(\mathbf{a}^*) = 0$. Hence, $\bar{\nu}_0(k)$ converges to a finite limit $\bar{\nu}_0$. A similar argument establishes that $\underline{\nu}_0(k)$ is decreasing in k and converges to a finite limit $\underline{\nu}_0 > -\infty$ as $k \to \infty$.

□

Proof for proposition 4.3. Let $\nu_0 := \nu_0^A$ and $\psi(\mathbf{a}) := \psi_P(\mathbf{a}) = \psi_A(\mathbf{a})$. From theorem 4.1, the agent's payoff simplifies to

$$V_A(\mathbf{a}) = r_A + (\nu_0 - r)\Phi\left(\frac{-\nu_0 - r}{\psi(\mathbf{a})}\right) - \psi(\mathbf{a})\phi\left(\frac{-\nu_0 - r}{\psi(\mathbf{a})}\right).$$

The agent's payoff is increasing in $\psi(\mathbf{a})$ if and only if

$$\psi(\mathbf{a})^2 \leq -2r(r + \nu_0). \tag{18}$$

Fixing ν_0 and r, V_A is single-peaked in ψ, with the peak at $\hat{\psi} = \sqrt{-2r(r+\nu_0)}$. Let ψ_k^2 denote the posterior variance of the single-player sample with capacity k.

(1) For $r = 0$, the RHS of (18) is zero, so the agent's payoff is strictly decreasing in posterior variance. Hence, for any k the agent prefers $\psi_0^2 = 0$ to any $\psi_k^2 > 0$.

(2) Fix $r > 0$. If $\nu_0 \geq -r$, the RHS of (18) is negative. Hence, the agent's payoff is strictly decreasing in posterior variance. No attributes are sampled optimally.

Consider $\nu_0 + r < 0$. If ν_0 is such that $\hat{\psi}^2 > \psi_k^2$, the agent's payoff is strictly increasing in ψ for $\psi^2 \in [0, \psi_k^2]$, hence the optimal sample consists of the single-player sample.

But $\hat{\psi}^2 > \psi_k^2 \Leftrightarrow \nu_0 < -r - \psi_k^2/(2r)$. If, on the other hand, $\hat{\psi}^2 < \psi_k^2$, the agent's payoff is maximized at posterior variance $\psi^2 = \hat{\psi}^2$. The argument for $r < 0$ is similar.

(3) Consider $r > 0$. For any $k' > k$, $\psi_{k'}^2 \geq \psi_k^2$. Fixing $\nu_0 < -r$, if the agent prefers $\hat{\psi}^2$ under k, he prefers it under $k' > k$ as well. Therefore, $\underline{\nu}_0(k)$ decreases with k. A similar argument shows that for $r > 0$, $\bar{\nu}_0(k)$ increases with k.

□

Lemma C.1 (Indifferent principal)**.** *If $\nu_0^P = r_P$ and the principal adopts with probability $\gamma > 1/2$ (resp., $\gamma < 1/2$) in the absence of any sampling, then the agent forgoes sampling for ν_0^A sufficiently high (resp., low) and maximizes $\alpha_2(\cdot)$ otherwise. If $\gamma = 1/2$, for any $\nu_0^A \in \mathbb{R}$ the agent forgoes sampling if and only if all feasible samples are controversial.*

Proof of lemma C.1. Suppose that if $\nu_0^P = r_P$ principal mixes between adoption and rejection with probabilities $(\gamma, 1-\gamma)$ where $\gamma \in [0,1]$. For any $\mathbf{a} \in \mathcal{A}_k$, $V_A(\mathbf{a}) = r_A + \frac{1}{2}(\nu_0^A - r_A) + \alpha_2(\mathbf{a})\phi(0)$. Therefore, the optimal sample solve

$$\mathbf{a}^* \in \arg\max_{\mathbf{a}\in\mathcal{A}_k} \quad \alpha_2(\mathbf{a}).$$

The agent prefers sampling to forgoing sampling if and only if

$$\begin{cases} (\nu_0^A - r_A)\,(\gamma - 1/2) \leq \max_{\mathbf{a}\in\mathcal{A}_k} \alpha_2(\mathbf{a})\phi(0) & \text{if } \gamma > 1/2 \\ (r_A - \nu_0^A)\,(1/2 - \gamma) \leq \max_{\mathbf{a}\in\mathcal{A}_k} \alpha_2(\mathbf{a})\phi(0) & \text{if } \gamma < 1/2. \end{cases}$$

If $\gamma = 1/2$, $V_A(\mathbf{a}) = V_A(\emptyset) + \alpha_2(\mathbf{a})\phi(0)$. This is strictly greater than $V_A(\emptyset)$ if and only if $\alpha_2(\mathbf{a}) > 0$. Hence, agent forgoes $\mathbf{a} \in \mathcal{A}_k$ if and only if $\alpha_2(\mathbf{a}) \leq 0$. □

Proof of proposition 4.4.

(1) Suppose players are in prior disagreement and $\mathbf{a}^*$ is dominated, i.e. $\exists \tilde{\mathbf{a}} \neq \mathbf{a}^*$, $\tilde{\mathbf{a}} \in \mathcal{A}_k$ such that $\alpha_1(\tilde{\mathbf{a}}) \geq \alpha_1(\mathbf{a}^*)$ and $\alpha_2(\tilde{\mathbf{a}}) \geq \alpha_2(\mathbf{a}^*)$ with at least one strict inequality. Without loss, suppose both are strict and $\nu_0^P - r_P > 0 > \nu_0^A - r_A$. Then,

$$(\nu_0^A - r_A)\left(\Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\tilde{\mathbf{a}})}\right) - \Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}^*)}\right)\right) > 0$$

and

$$\alpha_2(\tilde{\mathbf{a}})\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\tilde{\mathbf{a}})}\right) > \alpha_2(\mathbf{a}^*)\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}^*)}\right).$$

This contradicts the optimality of $\mathbf{a}^*$.

(2) Without loss, suppose $\nu_0^A > r_A$, $\nu_0^P > r_P$, and $\gamma > 1/2$ for an ex ante indifferent principal. We first show that it is without loss to restrict attention to the set $\{\mathbf{a} \in \mathcal{A}_k : \alpha_2(\mathbf{a}) \geq 0\}$, which is non-empty by condition (ii).

Claim 3. *If players are in prior agreement, then $\rho(\mathbf{a}^*) > 0$ for any optimal sample $\mathbf{a}^* \in \mathcal{A}_k$.*

Proof. Without loss, consider $\nu_0^i > r_i$ for $i = A, P$. By contradiction suppose that $\rho(\mathbf{a}^*) \leq 0$ for an optimal $\mathbf{a}^*$. The agent prefers $\mathbf{a}^*$ to no sampling if and only if

$$\rho(\mathbf{a}^*)\psi_A(\mathbf{a}^*)\frac{\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}^*)}\right)}{1 - \Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}^*)}\right)} \geq \nu_0^A - r_A.$$

The LHS is weakly negative due to $\rho(\mathbf{a}^*) \leq 0$, whereas the RHS is strictly positive. This contradicts the optimality of $\mathbf{a}^*$. So for any optimal sample $\mathbf{a}^*$, $\rho(\mathbf{a}^*) > 0$. □

$V_A(\mathbf{a})$ is continuous in (ν_0^A, ν_0^P). Moreover the set of feasible pairs $\{(\alpha_1(\mathbf{a}), \alpha_2(\mathbf{a})) : \mathbf{a} \in \mathcal{A}^k\}$ is closed and bounded by the continuity of covariance σ on $\mathcal{A} = [0,1]^d$, as well as constant in (ν_0^A, ν_0^P). By the Maximum Theorem, $V_A(\mathbf{a})$ is continuous in (ν_0^A, ν_0^P) and the set of optimal samples is upper hemicontinuous and compact in (ν_0^A, ν_0^P).

By lemma C.1, the agent samples an α_2-maximal sample if $\nu_0^P = r_P$ and $(\nu_0^A - r_A)$ is sufficiently close to zero. For the rest of this argument, let $\mathbf{a}_+^*$ denote a α_2-maximal sample that attains the highest informativeness for the principal α_1.

Fix $\nu_0^P - r_P < \epsilon$ where $\epsilon > 0$ small. Then for any $\nu_0^A > r_A$ there exists $\delta > 0$ such that $\mathbf{a}^*$ is within Euclidean distance δ from $\mathbf{a}_+^*$ in the (α_1, α_2)-plane. The agent's payoff strictly increases from a slight modification of $\mathbf{a}_+^*$ in a direction at which $\mathrm{d}\alpha_1(\mathbf{a}_+^*) < 0$ iff

$$(\nu_0^P - r_P)\left((\nu_0^P - r_P)\alpha_2(\mathbf{a}_+^*) - (\nu_0^A - r_A)\alpha_1(\mathbf{a}_+^*)\right)\,\mathrm{d}\alpha_1(\mathbf{a}_+^*) + \alpha_1(\mathbf{a}_+^*)\,\mathrm{d}\alpha_2(\mathbf{a}_+^*) > 0$$

$$\Leftrightarrow \frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}_+^*)}\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}_+^*)} - \frac{\nu_0^A - r_A}{\alpha_2(\mathbf{a}_+^*)}\right) < -\frac{\mathrm{d}\alpha_2(\mathbf{a}_+^*)/\alpha_2(\mathbf{a}_+^*)}{\mathrm{d}\alpha_1(\mathbf{a}_+^*)/\alpha_1(\mathbf{a}_+^*)}. \tag{19}$$

On the other hand, for agent to strictly prefer $\mathbf{a}_+^*$ to no sampling it must be that

$$\frac{\nu_0^A - r_A}{\alpha_2(\mathbf{a}_+^*)} < \frac{\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}_+^*)}\right)}{1 - \Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}_+^*)}\right)} \tag{20}$$

Such a modification of $\mathbf{a}_+^*$ is feasible because of condition (iii). From any such modification, $\mathrm{d}\alpha_2(\mathbf{a}^*) \leq 0$ because $\mathbf{a}_+^*$ is α_2-maximal. For ν_0^A and ν_0^P to satisfy both (19) and (20), ν_0^P should be sufficiently close to r_P and

$$\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}_+^*)} < \frac{\nu_0^A - r_A}{\alpha_1(\mathbf{a}_+^*)} < \frac{\phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}_+^*)}\right)}{1 - \Phi\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a}_+^*)}\right)}.$$

For any ν_0^P there exists ν_0^A that satisfies these inequalities because for any $x \geq 0$ $\phi(x)/(1-\Phi(x)) > x$. Moreover, at any such (ν_0^P, ν_0^A) modifying $\mathbf{a}_+^*$ so that $\mathrm{d}\alpha_1(\mathbf{a}_+^*) > 0$ strictly decreases the principal's payoff. Therefore, $\alpha_2(\mathbf{a}^*) \leq \alpha_2(\mathbf{a}_+^*)$ and $\alpha_1(\mathbf{a}^*) \leq \alpha_1(\mathbf{a}_+^*)$, with at least one strict inequality. □

Proof for proposition 4.5. (i) By way of contradiction, let players be in prior agreement and $\rho(\mathbf{a}^*) < 0$ for an optimal $\mathbf{a}^*$. Suppose first $\nu_0^i - r_i > 0$ for $i = A, P$. The agent prefers sampling to no sampling iff

$$\max_{\mathbf{a}\in\mathcal{A}_k} \quad \rho(\mathbf{a})\psi_A(\mathbf{a})\lambda\left(\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) \geq \nu_0^A - r_A.$$

For any informative $\mathbf{a} \in \mathcal{A}_k$, $\psi_A(\mathbf{a}) > 0$. Hence, the agent prefers no sampling to $\mathbf{a}^*$: this contradicts the optimality of $\mathbf{a}^*$. Therefore, $\rho(\mathbf{a}^*) > 0$. If instead $\nu_0^i - r_i < 0$ for $i = A, P$, the agent prefers sampling to no sampling iff

$$\max_{\mathbf{a}\in\mathcal{A}_k} \quad \rho(\mathbf{a})\psi_A(\mathbf{a})\lambda\left(-\frac{\nu_0^P - r_P}{\psi_P(\mathbf{a})}\right) \geq r_A - \nu_0^A.$$

Because $r_A - \nu_0^A > 0$, $\rho(\mathbf{a}^*) > 0$ for any optimal $\mathbf{a}^*$.

(ii) This follows immediately from the proof of part (i).

(iii) Suppose there exist another sample $\tilde{\mathbf{a}} \in \mathcal{A}_k$ such that $\rho(\tilde{\mathbf{a}}) > 0$ and it dominates the optimal sample $\mathbf{a}^*$ in both sufficient statistics. Because $\alpha_1(\tilde{\mathbf{a}}) \geq \alpha_1(\mathbf{a}^*)$ and $\mathrm{sgn}(\nu_0^A - r_A) \neq \mathrm{sgn}(\nu_0^P - r_P)$,

$$(\nu_0^A - r_A)\left(\Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}^*)}\right) - \Phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\tilde{\mathbf{a}})}\right)\right) \leq 0.$$

Moreover, $\alpha_2(\tilde{\mathbf{a}}) > 0$ and $\alpha_2(\tilde{\mathbf{a}}) \geq \alpha_2(\mathbf{a}^*)$ implies that

$$\alpha_2(\mathbf{a}^*)\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\mathbf{a}^*)}\right) - \alpha_2(\tilde{\mathbf{a}})\phi\left(\frac{\nu_0^P - r_P}{\alpha_1(\tilde{\mathbf{a}})}\right) \leq 0.$$

If at least one of these inequalities holds strictly, $V_A(\tilde{\mathbf{a}}) > V_A(\mathbf{a}^*)$, which contradicts the optimality of $\mathbf{a}^*$.

□