Optimal Sample Sizes and Statistical Decision Rules*

Sanket Patil and Yuval Salant†

March 30, 2023

Abstract

A statistical decision rule is a mapping from data to actions induced by statistical inference on the data. We characterize these rules for data that are chosen strategically in persuasion environments. A designer wishes to persuade a decision maker (DM) to take a particular action and decides how many Bernoulli experiments about a parameter of interest the DM can obtain. After obtaining these data and estimating the parameter value, the DM chooses to take the action if the estimated value exceeds some threshold. We establish that as the threshold changes, the resulting statistical decision rules in many environments are either simple majority or reverse unanimity.

Keywords: Statistical inference, statistical decision rule, sample size, persuasion.

JEL classification: D81, D83, C90.

*For helpful comments and discussions we thank seminar participants at Columbia University, the Econometric Society World Congress, the joint economic theory workshop to Hebrew University and Tel-Aviv University, Northwestern University, University of Michigan, Washington University in St. Louis, and the Western Economic Association International Conference.

†Patil: Indian Institute of Management Bangalore (sanket.patil@iimb.ac.in). Salant: Kellogg School of Management, Northwestern University (y-salant@kellogg.northwestern.edu).

1 Introduction

Interested parties often give decision makers (DMs) access to data in order to persuade them to take a particular action. For example, sellers may allow buyers to experiment with products prior to making a purchase. Similarly, lobbyists and think tanks often commission public-opinion surveys in order to convince politicians to support certain policy proposals. In both cases, the interested party decides how much data about a parameter of interest — product quality or public support — the DM can obtain, and the parameter value governs the distribution from which the data are generated. This paper studies persuasion with these features.

An interested party (henceforth a designer) wishes to convince a DM to take a particular action. The value of the action to the DM is governed by a parameter with an unknown value. The designer has a prior belief on the parameter value and decides how large a sample about it to provide to the DM. Each data point in the sample is an independent Bernoulli experiment governed by the parameter value. For example, in public-opinion surveys, the parameter may be the underlying public support for a policy, and each experiment may correspond to the opinion of a single survey respondent.

By choosing a sample size, the designer chooses an information structure, that is, a collection of signal distributions corresponding to different parameter values. This information structure satisfies what we call *responsiveness*: for any two parameter values, the signal distribution for the DM-preferred parameter value first-order-stochastically dominates the distribution for the second parameter value, but not the other way around. This property differs from the standard assumption in Bayesian Persuasion (Kamenica (2019)) whereby the designer can assign the *same* signal distribution to any two parameter values.

Responsiveness naturally arises in real-life settings. Public-opinion surveys aim to be responsive to changes in the opinion of the public. Product experimentation is sensitive to product quality. And in clinical trials, outcomes are governed by the effectiveness of the drug being tested. Surveys and clinical trials are used precisely because they are responsive to the parameter of interest.

Unlike the designer, the DM in our model wishes to take the action only if its value, which is governed by the payoff-relevant parameter, exceeds the value of the status quo. The DM obtains the sample, estimates the parameter value using statistical inference, and takes the action if the estimated value is sufficiently high.

We consider two approaches to statistical inference: the frequentist approach that relies only on the sample and the Bayesian approach that starts with a prior belief on the parameter value and Bayes-update it based on the sample. An advantage of the frequentist approach is that it does not require prior knowledge and uses only objective data. It thus fits environments in which the DM is less experienced, less knowledgeable about fundamentals, or does not want prior beliefs to influence decision making.[1] An important subclass of frequentist inference, which we focus on, is the class of unbiased inference procedures whereby the expected value of an estimate (which is a distribution over parameter values) is identical to the sample mean. Leading examples are maximum likelihood estimation and beta estimation.

Our analysis focuses on choice regularities that emerge as the status quo changes, either across time for a single individual or across individuals in a population. To summarize these regularities, we follow Wald (1949) and use the notion of a *statistical decision rule*: A direct mapping from samples (n, k), where n is a sample size chosen by the designer for some status-quo value and k is the number of successful experiments, to binary choices between taking the action and keeping the status quo.

We identify two statistical decision rules that arise endogenously for unbiased statistical inference. The first is the *reverse-unanimity* rule, whereby the DM takes the action unless all the experiments fail. This rule summarizes choice behavior when the designer's prior on the parameter value is decreasing. To be sure, the DM is a statistician who decides whether to take the action based on the estimated parameter

[1]The statistics literature has debated the merits of the two approaches for over a century. Efron (2005) is a nice recent discussion.

value. But the designer's sample-size selection results in a rule that exhibits what may seem like a "bias" in favor of taking the action. The second rule is *simple majority*, whereby the DM takes the action if and only if a simple majority of experiments are successful. This rule summarizes choice behavior when the designer's prior is either increasing and concave or symmetric, and the status-quo value exceeds $1/2$.

The majority rule also summarizes the choice behavior of Bayesian DMs who share the designer's symmetric prior, although for a "compressed" interval of status-quo values relative to unbiased inference. Figure 4 provides a graphical illustration. The compression of the interval is perhaps expected because, for Bayesian DMs, estimation gravitates toward the mean parameter value according to the prior. As the strength of the prior decreases, Bayesian DMs rely increasingly on the data, and thus their behavior converges to that of frequentist DMs.

We proceed as follows. Section 2 presents the model. Sections 3 and 4 analyze unbiased and Bayesian inference, respectively. Section 5 discusses the related literature. Section 6 concludes with a discussion of costly data provision. The Appendix contains proofs that do not appear in the main text.

2 Model

A decision maker (DM) has to decide whether to take an action or keep the status quo. The value $t \in (0, 1]$ of keeping the status quo is known to the DM and the value $q \in [0, 1]$ of taking the action is not. To make a decision, the DM estimates q using data and statistical inference, and takes the action if the estimated value of q is weakly larger than t.

The data are independent Bernoulli experiments with success probability q. That is, the data are governed by the value of taking the action. A successful experimental realization, or simply a success, is interpreted as a data point in favor of taking the action. An unsuccessful realization is called a failure.

The size of the data $n \in \mathcal{N}$ is decided by a designer who obtains a payoff of 1 if the DM takes the action and 0 otherwise. The designer thus wishes to persuade the DM to take the action regardless of the values of t and q. The designer knows t and believes that q is drawn from an absolutely continuous distribution with density f. The prior belief f is *decreasing (increasing)* if it (1) weakly decreases (increases) in q on $[0, 1]$ and (2) differs from the uniform prior on a non-zero measure. A prior f is *symmetric* if $f(1/2 - q) = f(1/2 + q)$ for $q \in [0, 1/2]$. Data provision is costless.[2]

After the designer decides the data size, the Bernoulli experiments are carried out by the DM or a third party, and the DM obtains their realizations. The DM's sample is the pair (n, k) where n is the number of experiments and k is the number of successes. We will refer to n as the sample size and to k/n as the sample mean.

2.1 Statistical inference

An inference procedure describes how the DM makes inferences from samples.

Definition 1. *(Salant and Cherry (2020)) An* ***inference procedure*** $G = \{G_{n,k}\}$ *assigns a cumulative distribution function (CDF) $G_{n,k}$, called an estimate, to every sample (n, k) such that:*

(i) the estimate $G_{n,k'}$ weakly first-order stochastically dominates the estimate $G_{n,k}$ when $k' > k$, and

(ii) the estimate $G_{n,n}$ first-order stochastically dominates the estimate $G_{n,0}$.[3]

An inference procedure is the analogue of an estimator in the statistics literature. It can be used to describe many forms of statistical inference. A focal example is Bayesian inference.

Example 1 (Bayesian Inference)**.** *The DM has a non-degenerate prior belief on q and uses Bayes rule to update it based on the sample.*

[2]Section 6 discusses costly data provision.

[3]This definition is weaker than in Salant and Cherry (2020) as we require dominance in the strict sense only for $G_{n,n}$ and $G_{n,0}$.

In addition to Bayesian inference, an inference procedure can also be used to model frequentist inference procedures that do not rely on prior beliefs. Here are three examples.

Example 2 (Maximum Likelihood Estimation (MLE))**.** *The DM calculates the most likely parameter q to have generated the sample. It is easy to verify that this parameter is the sample mean. Thus, the resulting estimate is $G_{n,k}(q) = \mathbb{1}_{\{q \geq k/n\}}$.*

Example 3 (Beta Estimation)**.** *The DM wishes to conduct Bayesian updating relying as little as possible on prior beliefs. The DM starts with Haldane's "prior" (Haldane (1932)), which is the limit of the Beta(ϵ, ϵ) distribution as $\epsilon \to 0$. The DM's posterior belief after obtaining the sample (n, k) is the limit of the corresponding Bayesian posteriors Beta$(\epsilon + k, \epsilon + (n-k))$, i.e., it is the Beta$(k, n-k)$ distribution. Thus, the DM's estimates rely only on the sample.*[4]

Example 4 (Dogmatic Views)**.** *The DM believes q is distributed either according to the CDF F_0 or the CDF F_1 that first-order stochastically dominates F_0. The DM decides which distribution to use in decision making based on the sample. If $k(n)$ or more realizations are successes, the DM uses F_1. Otherwise, the DM uses F_0.*

Frequentist inference procedures can be classified according to whether they are unbiased.

Definition 2. *An inference procedure G is **unbiased** if the expected value $\int_0^1 q dG_{n,k}$ of any estimate $G_{n,k}$ is equal to the sample mean k/n.*

The MLE and Beta Estimation procedures are unbiased. The dogmatic views procedure is not. Bayesian inference is also not unbiased because the expected value of an estimate depends on both the prior and sample mean.

Following the estimation of q, the DM uses the estimate to calculate the expected value of q, and takes the action if this value is weakly larger than t. For completeness, we assume that a frequentist DM does not take the action if no data is provided.

[4] When the sample contains only failures or successes, the DM concentrates the estimate on 0 or 1 respectively.

2.2 The persuasion probability

The designer's payoff is equal to the expected probability that the DM takes the action. To calculate this probability, we first denote by $b(n, j, q) = \binom{n}{j} q^j (1-q)^{n-j}$ the probability of obtaining j successes in n Bernoulli experiments with success probability q and by

$$P(n, k, q) = \sum_{j=k}^{n} b(n, j, q)$$

the probability of obtaining k or more successes. We denote by

$$P(n, k) = \int_0^1 P(n, k, q) f(q) dq$$

the expected value of the probability $P(n, k, q)$ according to the designer's prior f.

By the first-order stochastic dominance property of an inference procedure, if the DM takes the action after obtaining the sample (n, k), then the DM also takes the action after obtaining a sample (n, j) with $j > k$. Thus the probability that a DM with inference procedure G takes the action is $P(n, k(n, G, t), q)$ where $k(n, G, t)$ denotes the smallest number of successes after which the DM takes the action.[5] We call $P(n, k(n, G, t), q)$ the *persuasion probability*, and refer to its expected value according to the designer's prior, $P(n, k(n, G, t))$, as the *expected persuasion probability.* When clear from the context, we omit the dependence of k on its arguments.

For unbiased inference procedures, the focus of our analysis in Section 3, the smallest number of successes that trigger the DM to take the action is $k(n, G, t) = \lceil tn \rceil$. This number does not depend on the specifics of the inference procedure, and so we can write the persuasion probability as:

$$P(n, \lceil tn \rceil, q) = \sum_{j=\lceil tn \rceil}^{n} b(n, j, q). \tag{1}$$

[5]There may be sample sizes for which the DM does not take the action regardless of the number of successes. Such sample sizes are never optimal, and so we ignore them.

2.3 The designer's objective

The objective of the designer is to choose, for any value of the status quo t, a sample size that maximizes the expected persuasion probability. That is, the designer solves for

$$S(t) \equiv \underset{n \in \mathcal{N} \cup \{0\}}{\arg\max}\, P(n, k(n, G, t)) \tag{2}$$

where $S(t)$ may include more than one sample size and sample size 0 corresponds to the designer not providing any data to the DM.[6]

2.4 Statistical decision rules

The optimal sample size may change as the value of the status quo changes. As the sample size and sample realizations change, so does the DM's choice behavior. The following definition, based on Wald (1949), is about the mapping from samples to choices.

Definition 3. *For an interval $I \subset (0, 1]$ of status-quo values, a* ***statistical decision rule*** *C_I assigns an action $C_I(n, k) \in \{a, s\}$, where a denotes taking the action and s keeping the status quo, to any sample (n, k) where $n \in S(t)$ for some $t \in I$ and $0 \leq k \leq n$.*

The domain of a statistical decision rule is affected by the designer's choice of sample sizes. It is therefore a subset of the set of all possible samples. This domain may still be quite large (often countably infinite as we show below) because a statistical decision rule allows for variation in the status-quo value.

Two ubiquitous statistical decision rules are discussed in the analysis to follow. The first is *reverse unanimity*, whereby the DM takes the action unless all sample realizations are failures. That is, $C_{(0,1]}(n, k) = a$ if and only if $k > 0$. The second statistical decision rule, for $t > 1/2$, is the *simple-majority* rule, whereby the DM

[6] Sample size 0 is only relevant for the analysis of Bayesian inference because a frequentist DM would not take the action without obtaining data.

takes the action if and only if more than half of the sample realizations are successes. That is, $C_{(\frac{1}{2},1]}(n,k) = a$ if and only if $k > n/2$.

3 Unbiased inference

We proceed to study the designer's optimal sample sizes and the DM's statistical decision rules for unbiased statistical inference. The following property of the persuasion probability (proved in the Appendix) will be useful in the analysis.

Single crossing. *For any two pairs (n', k') and (n, k) such that $n' < n$ and $0 < k' < k$, there exists a cutoff $q^* > 0$ such that $P(n', k', q) > P(n, k, q)$ for $q \in (0, q^*)$. If $n' - k' < n - k$ then $q^* < 1$ and $P(n, k, q) > P(n', k', q)$ for $q \in (q^*, 1)$.*

The single-crossing property can be used to rank persuasion probabilities. For a given t, let k' and k be the smallest number of successes after which the DM takes the action for sample sizes n' and n respectively. If $k' < k$ and $n' - k' < n - k$ then the persuasion probability $P(n', k', q)$ crosses $P(n, k, q)$ exactly once and from above in $(0, 1)$. Consequently, sample size n' dominates sample size n for priors that put a "large weight" on small values of q whereas sample size n dominates n' for priors that put a large weight on large values of q.

Figure 1 provides a graphical illustration. Suppose that $t \in (1/2, 2/3]$. For sample size 1, a single success suffices for taking the action, and for sample size 3, at least two successes are required. The conditions for an interior single crossing are thus met, and $P(1, 1, q)$ crosses $P(3, 2, q)$ exactly once and from above in $(0, 1)$.

3.1 Decreasing priors and reverse unanimity

The first class of designer priors we analyze is the class of decreasing priors. Recall that a prior is *decreasing* if it weakly decreases in q on $[0, 1]$ and differs from the uniform prior on a non-zero measure.

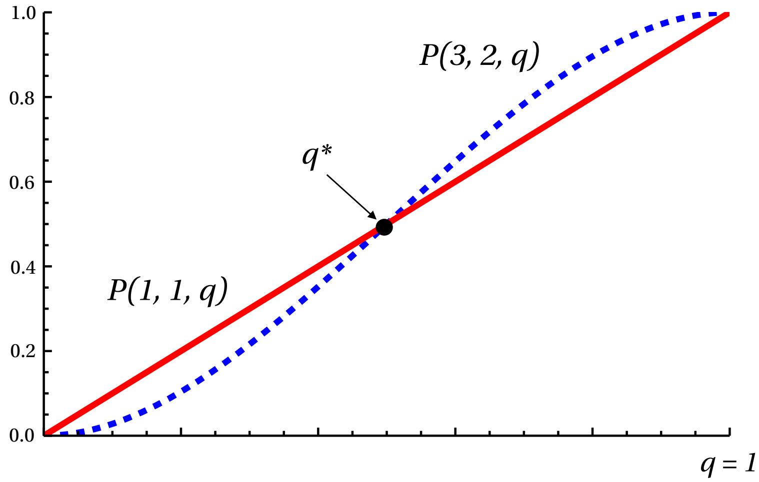

Figure 1: Persuasion probabilities $P(1,1,q)$ (solid) and $P(3,2,q)$ (dashed).

Theorem 1. *For any decreasing prior and any value t of the status quo, the uniquely optimal sample size is $\lfloor t^{-1} \rfloor$. Consequently, the DM's statistical decision rule is reverse unanimity.*

Theorem 1 says that any positive integer is a sample size chosen by the designer for some interval of status-quo values. Consequently, with sufficient variation in the status-quo, choice behavior can be observed for every feasible sample. This behavior, although generated by unbiased statistical inference, is consistent with the reverse-unanimity rule: the DM takes the action unless all experiments fail.

Theorem 1 also says that the optimal sample size increases as the status-quo value decreases. This is perhaps expected because the status-quo value reflects the intensity of the conflict between the DM and the designer. The DM wishes to take the action only if its value exceeds that of the status quo whereas the designer wishes the DM to take the action regardless of the status quo. The persuasion probability corresponding to the optimal sample size, which is given by $1-(1-q)^n$, also increases as the status quo decreases implying that persuasion is more likely when there is less conflict.

Proving Theorem 1 amounts to showing that sample size n' is uniquely optimal for the interval $I_{n'} = (\frac{1}{n'+1}, \frac{1}{n'}]$ of status-quo values. Let's fix this interval and focus on samples of the form $(n(k), k)$ with $n(k) = k(n'+1) - 1$. That is, $n(k)$ is the largest sample size n such that the sample (n, k) triggers taking the action for some status-quo value in $I_{n'}$.[7] By definition, $n(1) = n'$, and the sample $(n', 1)$ triggers taking the action for any status-quo value in $I_{n'}$.

The first step in the proof of Theorem 1 below is to establish that expected persuasion probabilities in the interval $I_{n'}$ are bounded from above by

$$\max_{k \in \mathcal{N}} P(n(k), k).$$

Therefore, it suffices to show that $P(n', 1)$ dominates all other $P(n(k), k)$'s. In the second step, we calculate the probabilities $P(n(k), k)$ under the uniform prior and show that they are equal. Finally, we observe that the transition from the uniform prior to a decreasing one is an operation that "shifts mass" from higher values of q to lower ones. Thus, by the single-crossing property, mass is shifted to values in which $P(n(1), 1, q)$ dominates all other $P(n(k), k, q)$'s, establishing the result.

Proof of Theorem 1. We show that sample size n' is uniquely optimal for $t \in I_{n'} = (\frac{1}{n'+1}, \frac{1}{n'}]$.

We first assign to every positive integer k the sample size $n(k) = k(n'+1) - 1$. Note that $n(1) = n'$.

Lemma 1. *For any integer $k \geq 1$, $P(n(k), k) > P(n, k)$ for $n < n(k)$.*

Lemma 1 implies that it suffices to compare $P(n', 1)$ to other probabilities of the form $P(n(k), k)$ in order to prove the optimality of n'. Indeed, consider some sample size n and let k be the smallest number of successes triggering the DM to take the action for this sample size. That is, $(k-1)/n < t \leq k/n$. Coupled with the inequality $t > 1/(n'+1)$, we have that $n \leq k(n'+1) - 1 = n(k)$. By Lemma 1,

[7] Note that the sample $(n(k), k)$ does not necessarily trigger taking the action for all status-quo values in $I_{n'}$.

the expected persuasion probability $P(n,k)$ is bounded from above by $P(n(k),k)$. Thus, if the expected persuasion probability $P(n',1)$ exceeds $P(n(k),k)$ for all $k \geq 2$ then $P(n',1)$ exceeds all the expected persuasion probabilities of the form $P(n,k)$ satisfying $\frac{k}{n} \geq t$. Consequently, sample size n' is optimal.

For the uniform prior, the probabilities $P(n(k),k)$ are equal. This is because the probability $P(n,k,q)$ is the incomplete regularized Beta function, which in turn is the CDF of the $Beta(k, n-k+1)$ distribution with mean $\frac{k}{n+1}$. Therefore, the probability $P(n,k)$, which for the uniform prior is the integral of the corresponding CDF, is equal to $1 - \frac{k}{n+1}$. Thus, for the uniform prior:

$$P(n(k),k) = 1 - k/(n(k)+1) = 1 - 1/(n'+1) = P(n',1).$$

For decreasing priors, $P(n(1),1)$ exceeds $P(n(k),k)$ for every $k \geq 2$. This is an implication of the single-crossing property because a decreasing prior intuitively "shifts mass" from high values of q to lower ones relative to the uniform prior. The following lemma formalizes this intuition and thus concludes the proof.

Lemma 2. *If $P(n',1) \geq P(n(k),k)$ for the uniform prior, then $P(n',1) > P(n(k),k)$ for any decreasing prior.*

□

3.2 Increasing or symmetric priors and simple majority

For increasing priors, we can obtain a lower bound on the optimal sample size.

Proposition 1. *For any increasing prior and any value t of the status quo, the optimal sample size is weakly larger than $n(k') = k'(n'+1) - 1$ where $n' = \lfloor t^{-1} \rfloor$ and k' is the unique integer satisfying $\frac{k'+1}{n(k'+1)} < t \leq \frac{k'}{n(k')}$.*

The proof of Proposition 1 follows similar ideas to those of Theorem 1 with two modifications. First, the transition from a uniform prior to an increasing one shifts mass to higher values of q, which favor larger sample sizes of the form $n(k)$. Second, for a given status quo, the largest sample of the form $(n(k), k)$ that triggers taking the action is $(n(k'), k')$ where k' is defined as in the statement of the proposition. Sample size $n(k')$ is thus a lower bound on the optimal sample size.

For $t \in (1/2, 1] = I_1$, we have that $n' = 1$ in the statement of Proposition 1. The smallest candidate for optimality is therefore the largest odd sample size $2k - 1$ satisfying that a simple majority k of successes triggers taking the action. Theorem 2 shows that this candidate is indeed optimal when the prior is also concave.

Theorem 2. *For any increasing and concave prior and any value $t > 1/2$ of the status quo, the uniquely optimal sample size is the largest odd sample size weakly below $\lfloor (2t-1)^{-1} \rfloor$. Consequently, the DM's statistical decision rule is simple majority.*

Theorem 2 says that for $t > 1/2$ any odd integer n is a uniquely optimal sample size for some interval of status-quo values. The optimal sample size and the corresponding persuasion probability decrease in the status-quo value, i.e., as the intensity of conflict between the DM and the designer increases. Variation in the status quo reveals a familiar choice pattern: behavior is consistent with the simple majority rule.

Theorem 2 can be extended to any prior that increases in q and decreases in the ratio $f(q)/q$.[8] Intuitively, when the ratio $f(q)/q$ decreases in q, the prior assigns sufficiently large mass to small values of q so that the sample sizes identified in Proposition 1 as a lower bound for $t > 1/2$ are optimal. When this is not the case, large sample sizes may dominate smaller ones and an optimal sample size may not exist. To illustrate, consider a prior that assigns positive mass only to q's above some fixed $q' > 0$. This prior may reflect the belief of a designer who is confident that the DM's value of taking the action is at least q'. For such an increasing prior, if the status-quo value is below q', the designer has an incentive to provide unlimited data to the DM as providing an additional data point further increases the persuasion probability for large sample sizes, and large sample sizes dominate small ones.

[8]The proof is essentially identical to the proof of Theorem 2 in the Appendix.

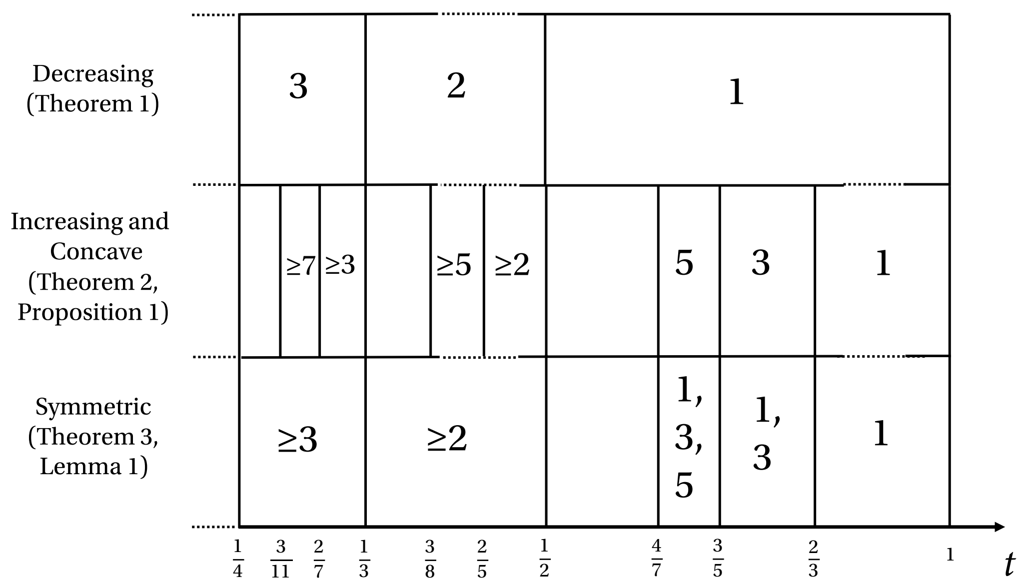

Figure 2: Optimal sample sizes for monotone or symmetric priors.

A modified version of Theorem 2 also holds for symmetric priors.

Theorem 3. *For any symmetric prior and any value $t > 1/2$ of the status quo, all odd sample sizes weakly below $\lfloor (2t-1)^{-1} \rfloor$ are optimal. Consequently, the DM's statistical decision rule is simple majority.*

To prove Theorem 3, we first observe that expected persuasion probabilities in I_1 are bounded from above by probabilities of the form $P(2k-1, k)$. The probabilities $P(2k-1, k)$ are in turn equal for symmetric priors. This is because for any two integers k and k', the difference function $\Delta(q) = P(2k-1, k, q) - P(2k'-1, k', q)$ satisfies $\Delta(q) = -\Delta(1-q)$. Thus, it suffices to focus on odd sample sizes satisfying that a simple majority of successes triggers taking the action. For a given $t > 1/2$, the largest sample size for which this happens is the one specified in the statement of Theorem 3.

Figure 2 summarizes the results of Section 3. Decreasing priors provide the lower envelope of the optimal sample sizes. For $t > 1/2$, increasing and concave priors provide the upper envelope, and symmetric priors span the range of all odd sample sizes between the lower and upper envelopes.[9]

4 Bayesian inference

When the DM is a Bayesian statistician who shares the designer's prior f rather than a frequentist statistician, simple majority continues to be a relevant choice regularity.

To illustrate this point, we consider $Beta(\alpha, \beta)$ priors with the density

$$f(\alpha, \beta) = \frac{q^{\alpha-1}(1-q)^{\beta-1}}{\int_0^1 q^{\alpha-1}(1-q)^{\beta-1}dq}$$

where α and β are real numbers. We may think about α as measuring the "prior number of successes", or how confident the DM is — based on prior knowledge — that q is large. Similarly, β is the "prior number of failures" measuring the DM's prior confidence that q is small.

Symmetric $Beta$ priors are characterized by $\alpha = \beta$. They have a variety of shapes as α changes. As α increases from just above 0 to 1, the $Beta(\alpha, \alpha)$ distribution has a U-shape that gradually flattens until the distribution becomes uniform for $\alpha = 1$. As α increases above 1, the distribution transitions to a reverse U-shape before taking a bell shape with an increasingly large mass around its mean. Figure 3 provides a graphical illustration.

A DM with a $Beta(\alpha, \alpha)$ prior who obtains the sample (n, k) Bayes-updates the prior belief to the $Beta(\alpha+k, \alpha+(n-k))$ distribution.[10] That is, the sample successes are added to the prior number of successes, and sample failures are added to the prior number of failures. The expected value of this Bayesian estimate is a weighted average of the prior mean 1/2 and the sample mean k/n with weights corresponding to the

[9] The lower bound on the optimal sample size for symmetric priors, and in fact any prior, follows from Lemma 1. By the Lemma, if $t \in I_{n'}$ then $P(n', 1) > P(n, 1)$ for $n < n'$.

[10] This is because the $Beta(\alpha, \beta)$ distribution is a conjugate prior for Bernoulli experiments.

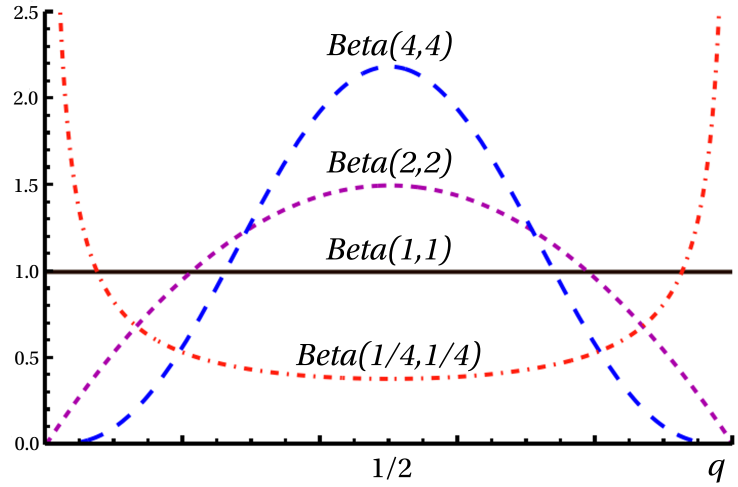

Figure 3: Symmetric *Beta* priors for $\alpha = 1/4, 1, 2$, and 4.

prior and sample relative strengths, 2α and n, respectively:

$$\underbrace{\frac{2\alpha}{2\alpha+n}}_{\text{prior weight}} \cdot \overbrace{\frac{1}{2}}^{\text{prior mean}} + \underbrace{\frac{n}{2\alpha+n}}_{\text{sample weight}} \cdot \overbrace{\frac{k}{n}}^{\text{sample mean}} .$$

Bayesian inference has two immediate implications for the analysis. First, if the status-quo value is smaller than the prior mean $1/2$, data provision is irrelevant: the DM takes the action with probability 1 without obtaining any additional data. Second, if the status-quo value is above $\frac{\alpha+1}{2\alpha+1}$, which is the posterior mean after observing the sample $(1,1)$, then sample size 1 cannot be optimal and the domain of relevant sample sizes is smaller than for unbiased inference. To facilitate the comparison of Bayesian inference to unbiased inference, we therefore focus on $t \in \left(1/2, \frac{\alpha+1}{2\alpha+1}\right] = I_\alpha$ where the domains of the designer's maximization problem for Bayesian and unbiased inference are identical.

Theorem 3B, where B stands for Bayesian, establishes that the DM's choice behavior is consistent with simple majority in I_α. In fact, the entire structure of the optimal sample sizes in I_α is identical to the structure characterized in Theorem 3 for unbiased inference in I_1.

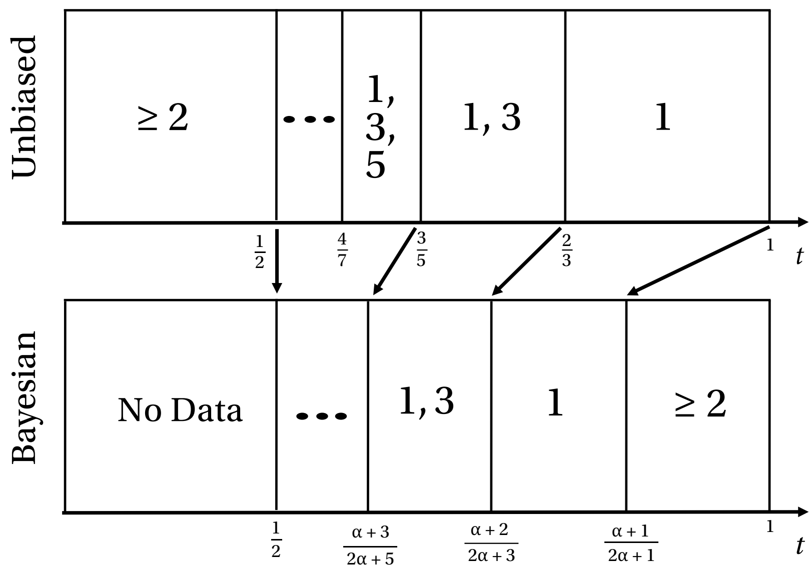

Figure 4: Optimal sample sizes for $Beta(\alpha, \alpha)$ priors.

Theorem 3B. *For any $Beta(\alpha, \alpha)$ prior and $t \in \left(1/2, \frac{\alpha+1}{2\alpha+1}\right]$, all odd sample sizes weakly below $\lfloor (2t-1)^{-1} - 2\alpha \rfloor$ are optimal. Consequently, the DM's statistical decision rule is simple majority.*

Figure 4 provides a graphical illustration of how the optimal sample sizes compare to those of unbiased inference. Sample size $2k-1$ is optimal for unbiased inference when $t \in (1/2, \frac{k}{2k-1}]$. The upper bound of this interval decreases to $\frac{\alpha+k}{2\alpha+(2k-1)}$ for Bayesian inference. As α decreases, the upper bound of the interval for Bayesian inference increases, and it converges to that for unbiased inference as α tends to 0. This convergence holds because as α decreases, the strength of prior 2α decreases, and Bayesian inference converges to the Beta estimation procedure in Example 3.

A focal example of a symmetric prior is the uniform distribution, which is the $Beta(1, 1)$ distribution. Because we can derive closed-form expressions of the expected persuasion probabilities in this case, we can solve for the optimal sample size directly and derive the DM's statistical decision rule for the entire interval of status-quo values.

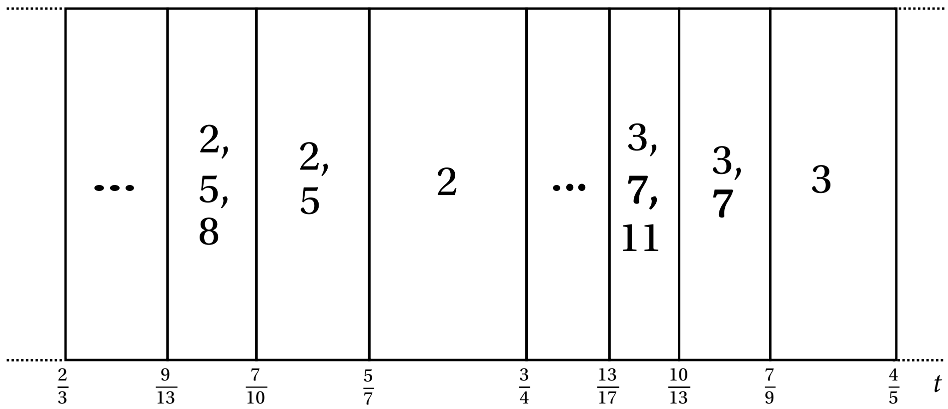

Figure 5: Optimal sample sizes for Bayesian inference from a uniform prior and $t \in (2/3, 4/5]$.

Proposition 2. *Let* $t \in \left(\frac{n'}{n'+1}, \frac{n'+1}{n'+2}\right]$ *for some positive integer* n'. *For the uniform prior, sample size* n *is optimal if and only if* $n = k(n'+1) - 1$ *for some positive integer* k *and the sample* (n, kn') *triggers taking the action. Consequently, the DM's statistical decision rule is the* $\frac{n'}{n'+1}$*-super majority rule.*

Thus, for $t \in (1/2, 2/3]$, the optimal sample sizes and the DM's statistical decision rule coincide with those of Theorem 3B: Simple majority is the required majority for taking the action. For $t \in (2/3, 3/4]$ the required super-majority is two thirds, for $t \in (3/4, 4/5]$ it is three quarters, and so on. Figure 5 depicts the corresponding collection of optimal sample sizes.

5 Related literature

This paper is related to the literature that incorporates sampling (Osborne and Rubinstein (1998, 2003)) and statistical inference (Salant and Cherry (2020)) into games. In Osborne and Rubinstein (1998)'s $S(k)$-equilibrium, players do not know the mapping from own actions to payoffs, sample the payoff of each action k times, and choose the action with the highest sampled payoff. Osborne and Rubinstein (2003) is a subsequent contribution in which players sample other players' actions instead of own payoffs and best respond to sample averages. Salant and Cherry (2020) enrich this framework by considering players who use statistical inference to estimate other players' actions and best respond to their estimates. Spiegler (2006a,b) studies competition between firms that face consumers who sample a payoff-relevant parameter once, and Sethi (2000, 2021) and Mantilla, Sethi and Cárdenas (2020) study stability properties of $S(1)$-equilibria. This literature treats players' sample size as a primitive of their decision making procedure or a component of the solution concept. And with the exception of Salant and Cherry (2020), the literature solves models with players who obtain very small samples. The focus of the current paper is on an orthogonal question. We treat the sample size as a design parameter and solve for the optimal sample size.

As such, our paper contributes to the small literature on the design of experiments in strategic environments. In Di Tillio, Ottaviani and Sørensen (2021), the DM decides (1) the minimal sample size the designer collects on behalf of the DM, and (2) whether to allow the designer to collect a larger sample at a cost and provide the DM with a non-random selection from it. We consider a complementary setting in which the designer decides the sample size without any constraints imposed by the DM, and the DM obtains the entire sample. Other contributions to this literature include Chassang, Padró i Miquel and Snowberg (2012) and Banerjee et al. (2020).

A third related literature is the literature on Bayesian persuasion (Kamenica and Gentzkow (2011)). We depart from this literature in two ways. First, as discussed in the Introduction, feasible information structures in our setup are responsive to changes in the value of the relevant parameter. Second, we consider frequentist inference procedures in addition to Bayesian inference. In this respect, the current paper is also related to contributions on persuasion with non-Bayesian DMs (Glazer and Rubinstein (2012), Galperti (2019), Eliaz, Spiegler and Thysen (2021), and Levy, Moreno de Barreda and Razin (2022)).

6 Conclusion

We considered a designer who wishes to persuade a DM to make a particular choice and who controls how large a sample about a parameter of interest the DM can obtain before choosing. Selecting a sample size results in an information structure that is responsive to changes in the parameter value: the signal distribution increases in first-order stochastic dominance sense as the parameter value increases.

The DM in our model is a statistician who uses data and statistical inference to estimate the parameter. A large part of the analysis focused on frequentist statistical inference that does not rely on prior beliefs. In addition to analytical tractability, incorporating prior-free statistical inference into the study of strategic interactions is relevant because of the increasing use of prior-free estimation methods and machine learning techniques in real-life strategic settings.[11]

After estimating the parameter, the DM makes a choice based on the estimated parameter value. A main focus of our analysis was to examine the resulting statistical decision rules. We identified environments in which commonly observed decision rules — simply majority and reverse unanimity — summarize the DM's choice behavior when sample sizes are chosen strategically by a designer to maximize the persuasion probability.

In real-life settings, data provision may be costly to the designer. For example, allowing consumers more experimentation with a product may carry an opportunity cost or lead to wear and tear. The designer may take such a cost into account when deciding how much experimentation to allow. We conclude by illustrating that although the designer's optimal sample sizes may change with costly data provision, the DM's statistical decision rule may not.

Consider a designer who faces a small cost $c < P(1,1) = \mathbb{E}_f(q)$ of providing an additional experiment to an unbiased statistician. Theorem 1C, where C stands

[11] Earlier contributions that incorporate the frequentist approach into economic theory include Al-Najjar (2009) in the context of individual decision making and Salant and Cherry (2020) in the context of multi-person decision making.

for costly, establishes that a modified Theorem 1 continues to hold with costly data provision.

Theorem 1C. *For any decreasing prior f and a positive cost $c < \mathbb{E}_f(q)$, there exists $\overline{n}(c) \in \mathcal{N}$ such that for any t, the optimal sample size is min*$\{\overline{n}(c), \lfloor t^{-1} \rfloor\}$. *Consequently, the DM's statistical decision rule is reverse unanimity.*

Thus, for status-quo values above $1/(\overline{n}(c)+1)$ the optimal sample sizes are identical to those with costless data provision whereas for smaller status-quo values, the cost binds and the designer chooses sample size $\overline{n}(c)$. The domain of the statistical decision rule is therefore finite and it increases (in the set inclusion sense) as c decreases. While the domain changes relative to costless experimentation, choice behavior does not: It continues to be consistent with reverse unanimity.

An analogous result holds for increasing and concave priors.[12]

Theorem 2C. *For any increasing and concave prior f and a positive cost $c < \mathbb{E}_f(q)$, there exists $\overline{n}(c) \in \mathcal{N}$ such that for any $t > 1/2$, the designer's optimal sample size is the largest odd sample size weakly below min*$\{2\overline{n}(c) - 1, \lfloor (2t-1)^{-1} \rfloor\}$. *Consequently, the DM's statistical decision rule is simple majority.*

Thus, similarly to decreasing priors, the domain of the statistical decision rule changes with the cost, but the decision rule itself remains identical to the one obtained with costless data provision.

A Proofs

Proof of the single-crossing property. Consider the ratio function

$$r(q) \equiv \frac{P'(n,k,q)}{P'(n',k',q)} = \frac{k\binom{n}{k}}{k'\binom{n'}{k'}} q^{(k-k')}(1-q)^{(n-k-n'+k')}$$

[12] For symmetric priors and $t > 1/2$, the designer always chooses sample size 1 with costly data provision.

where P' is the derivative of P. The function $r(q)$ is continuous, positive on $(0,1)$, and, because $k > k'$, approaches 0 as q approaches 0. Thus, $P(n', k', q)$ is above $P(n, k, q)$ in a neighborhood of 0.

To verify that $P(n, k, q)$ crosses $P(n', k', q)$ at most once in $(0,1)$, assume to the contrary that there are two or more such crossings. Then, the difference function $\Delta(q) = P(n', k', q) - P(n, k, q)$ has at least three interior extremum points with $\Delta'(q) = 0$ because $\Delta(0) = \Delta(1) = 0$. Because $\Delta'(q) = 0$ for $q \in (0,1)$ if and only if $r(q) = 1$, there are at least three interior points in which $r(q) = 1$. But $r(q)$ has at most two such points as we show below, implying that $P(n, k, q)$ crosses $P(n', k', q)$ at most once in $(0,1)$.

Observe that $r(q)$ has at least one point in which it is equal to 1. Otherwise, $r(q) < 1$, and thus $\Delta'(q) < 0$, for all $q \in (0,1)$. The latter inequality implies that $\Delta(1) < 0$ in contradiction to $\Delta(1) = 0$. To establish that there are at most two such points, consider first the case in which $n' - k' \geq n - k$. In this case, $r(q)$ increases in q. Therefore there is exactly one point in which $r(q)$ is equal to 1 and $q^* = 1$. If, however, $n - k > n' - k'$, then $r(q)$ increases up to $\frac{k-k'}{n-n'} < 1$ and then decreases. Thus, there are exactly two interior points in which $r(q)$ is equal to 1 implying that $q^* < 1$. □

Proof of Lemma 1. Fix an integer $k \geq 1$. It suffices to show that $P(n(k), k, q) > P(n, k, q)$ for $q \in (0,1)$. Using the binomial identity $b(m, l, q) = qb(m-1, l-1, q) + (1-q)b(m-1, l, q)$ multiple times, we obtain that

$$P(n+1, k, q) = P(n, k, q) + qb(n, k-1, q) > P(n, k, q) \tag{1}$$

which in turn implies the desired inequality because $n(k) > n$. □

Proof of Lemma 2. By single crossing, $P(n', 1, q)$ crosses $P(n(k), k, q)$ from above at some $q^* \in (0,1)$. If $f(q^*) = 0$, then because f is decreasing, its support is nested in $[0, q^*]$ and $P(n', 1, q)$ dominates $P(n(k), k, q)$ on the entire support of f.

Suppose $f(q^*) > 0$ and consider the constant function $h(q) = f(q^*)$. Because f is decreasing, the function h either reduces the mass on q's smaller than q^* relative to f, i.e., $\int_0^{q^*} f(q)dq > \int_0^{q^*} h(q)dq$, or h increases the mass on q's larger than q^*. Thus, by single crossing, if $P(n', 1) \geq P(n(k), k)$ where the expectation is taken with respect to h then a strict inequality holds with respect to f. Since h is a re-scaling of the uniform distribution, $P(n', 1) \geq P(n(k), k)$ for h if and only if $P(n', 1) \geq P(n(k), k)$ for the uniform prior. The result follows. □

Proof of Proposition 1. Fix $t \in \left(\frac{k'+1}{n(k'+1)}, \frac{k'}{n(k')}\right]$. For sample size $n(k)$ with $k \leq k'$, the smallest number of successes that trigger taking the action is k. Therefore, the corresponding expected persuasion probability is $P(n(k), k)$. As in the proof of Theorem 1, the probabilities $P(n(k), k)$ are equal under the uniform prior, and an argument mirroring Lemma 2 implies that $P(n(k'), k')$ dominates $P(n(k), k)$ for $k < k'$.

As for the remaining sample sizes, fix $n < n(k')$ and let k be the smallest number of successes that trigger taking the action for this sample size. Since $\frac{k}{n} \geq t > \frac{1}{n'+1}$, we have that $n < n(k)$. Thus, by Lemma 1, sample size $n(k)$, which is strictly dominated by sample size $n(k')$, dominates sample size n. □

Proof of Theorem 2. Fix an integer $m \geq 1$ and let $t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right]$. By Proposition 1, it suffices to show that sample size $2m - 1$ dominates larger sample sizes. For any sample size $n > 2m - 1$, let

$$k[n] = \left\{ \lceil nt \rceil \mid t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1} \right] \right\} \tag{2}$$

denote the set of all integers k satisfying that there exists t in the relevant interval for which k is the smallest integer such that the DM takes the action after obtaining the sample (n, k). Let $\kappa(n)$ denote the smallest integer in $k[n]$. By definition, $k[2m - 1]$ is a singleton with $\kappa(2m - 1) = m$, and $\kappa(2m) = m + 1$.

Figure 6 illustrates construction of the set $k[n]$ for $n = 3, \ldots, 13$ and $m = 2$.

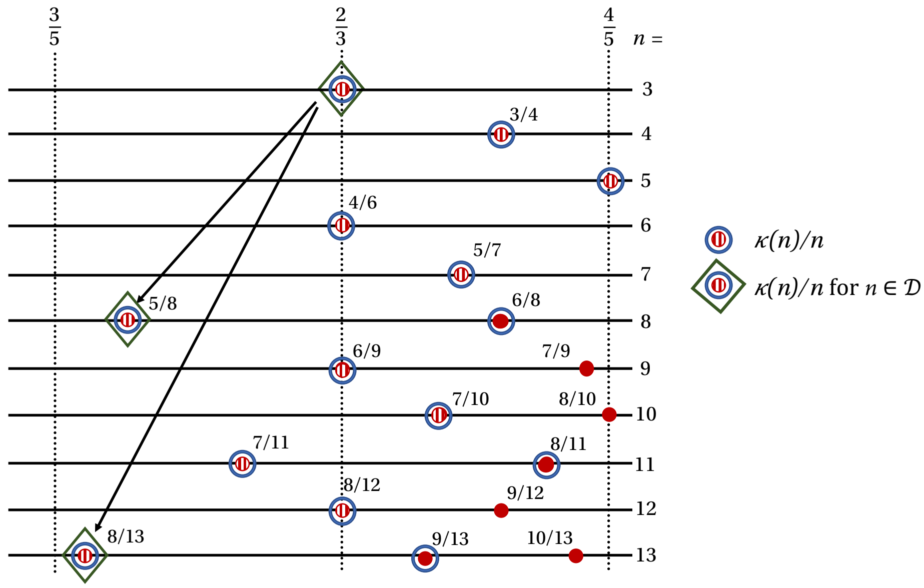

Figure 6: Construction of $k[n]$ and $\mathcal{D}$ in Theorem 2 for $m = 2$.
Note: Dots correspond to k/n's. Circled dots correspond to $\frac{k}{n}$ for $k \in k[n]$. Circled-striped dots correspond $\frac{\kappa(n)}{n}$. Dots circumscribed by diamonds correspond to $\frac{\kappa(n)}{n}$ for $n \in \mathcal{D}$. Arrows point in the direction of domination in the set $\mathcal{D}$.

For $m = 2$, the relevant interval of t's is $(3/5, 2/3]$. Dots in the figure (solid and striped) correspond to k/n's and they are circled when $k \in k[n]$. For example, $k[8] = \{5, 6\}$ and therefore the corresponding dots are circled, while $7 \notin k[9]$ and therefore the corresponding dot is not circled. Circled-striped dots correspond to $\frac{\kappa(n)}{n}$. For example, $\kappa(8) = 5$ and therefore the corresponding dot is striped.

We need to show that $P(2m-1, m) > P(n, k)$ for $n > 2m - 1$ and $k \in k[n]$. We do so in two steps. The first step establishes that it suffices to examine the expected persuasion probabilities with respect to the linear prior $h(q) = 2q$.

Step 1. If $P(2m-1, m) > P(n, k)$ for the linear prior then $P(2m-1, m) > P(n, k)$ for an increasing and concave prior.

Proof. If $m - 1 \geq n - k$, then single crossing implies the desired inequality holds regardless of the prior because $m < k$. Otherwise, $P(2m-1, m, q) > P(n, k, q)$ for $q \in (0, q^*)$ and a reverse inequality holds for $q \in (q^*, 1)$, where $q^* \in (0, 1)$. Consider the linear function $f_h(q) = \frac{f(q^*)}{q^*} q = \frac{f(q^*)}{2q^*} h(q)$ created by extending the secant between $(0,0)$ and $(q^*, f(q^*))$ until $q = 1$. This function either reduces the mass on $q < q^*$ or increases the mass on $q > q^*$ implying that if $P(2m-1, m) > P(n,k)$ for f_h (which is not necessarily a density), then $P(2m-1,m) > P(n,k)$ for f. Because f_h is a re-scaling of h, it suffices to prove the inequality for h. $\square$

The second step establishes the desired inequality with respect to h.

Step 2. For the linear prior, $P(2m-1, m) > P(n,k)$.

Proof. For any sample (n, k), we have:

$$\begin{aligned} P(n,k) &= [P(n,k,q)q^2]_{q=0}^1 - \int_0^1 P'(n,k,q)q^2 dq \\ &= 1 - \frac{k\binom{n}{k}}{(k+2)\binom{n+2}{k+2}} \int_0^1 P'(n+2,k+2,q)dq = 1 - \frac{k(k+1)}{(n+1)(n+2)} \end{aligned}$$

where the first equality follows from integration by parts, and the second equality follows from $P'(n,k,q) = k\binom{n}{k} q^{(k-1)}(1-q)^{(n-k)}$.

To show that $P(2m-1,m) > P(n,k)$, we thus need to show that $\frac{k(k+1)}{(n+1)(n+2)} > \frac{(m+1)}{2(2m+1)}$ which holds if (i) $\frac{k+1}{n+1} > \frac{m+1}{2m+1}$, and (ii) $\frac{k}{n+2} \geq \frac{1}{2}$.

Inequality (i) holds because $\frac{k+1}{n+1} \geq \frac{k}{n} \geq t > \frac{m+1}{2m+1}$. To prove inequality (ii), it suffices to consider whether (ii) holds for $\kappa(n)$, and to do that, we consider the set $\mathcal{D}$ of all sample sizes n satisfying that $\frac{\kappa(n)}{n} < \frac{\kappa(n_1)}{n_1}$ for every n_1 such that $2m-1 \leq n_1 < n$. In Figure 6, sample sizes $3, 8$ and 13 belong to set $\mathcal{D}$ and the corresponding dots are circumscribed by diamonds. If $n \in \mathcal{D} \backslash \{2m-1\}$, Lemma A1 (stated and proved below) implies that $n = (2m+1)l - 2$ for some integer $l \geq 2$. By definition, $\kappa(n) = (m+1)l - 1$. Thus, $\kappa(n)/(n+2) \geq 1/2$. If $n \notin \mathcal{D}$, Lemma A2 (stated and proved below) implies that there exists $n_1 \in \mathcal{D}$ that satisfies $n_1 \leq n$ and $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$. These two inequalities imply that $\kappa(n) \geq \kappa(n_1)$, which in turn implies that $\kappa(n)/(n+2) \geq \kappa(n_1)/(n_1+2) \geq 1/2$. $\square$

$\square$

Lemma A1. *Fix an integer $m \geq 1$. If $n \in \mathcal{D}$ then $n = (2m+1)l - 2$ for some $l \in \mathcal{N}$.*

Proof. Fix $l \in \mathcal{N}$ and consider sample sizes $n_1 = (2m+1)l - 2$ and $n_2 = (2m+1)(l+1) - 2$. By definition, $\kappa(n_1) = (m+1)l - 1$ and $\kappa(n_2) = (m+1)(l+1) - 1$. Thus, $\frac{\kappa(n_2)}{n_2} < \frac{\kappa(n_1)}{n_1}$. To complete the proof, it suffices to show that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$ for any $n_1 < n < n_2$.

Any such n satisfies $\kappa(n) = (m+1)l + j$ for some integer $0 \leq j \leq m$ because $\kappa(n) > \kappa(n_1)$ (this follows from $\frac{\kappa(n_1)}{n_1+1} \leq \frac{m+1}{2m+1}$) and $\kappa(n) \leq \kappa(n_2)$. Fix j and consider the set of all n's with $\kappa(n) = \kappa(j) = (m+1)l + j$. Since the ratio $\frac{\kappa(n)}{n}$ decreases in n, it suffices to verify that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$ for the largest n in the set. We denote this maximal n as $n(j)$. Then $n(j) = (2m+1)l + 2j - 1$ for $0 \leq j \leq m-1$ because it satisfies the inequality $\frac{\kappa(j)}{n(j)+1} \leq \frac{m+1}{2m+1} < \frac{\kappa(j)}{n(j)}$, and $n(m) = n_2 - 1$. Verifying that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(j)}{n(j)}$ completes the proof. □

Lemma A2. *If $n \notin \mathcal{D}$, then there exists $n_1 \in \mathcal{D}$ such that $n_1 \leq n$ and $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$.*

Proof. Fix $n \notin \mathcal{D}$ and let n_1 be the largest sample size in $\mathcal{D}$ that is smaller than n. The proof of Lemma A1 implies that $\frac{\kappa(n_1)}{n_1} \leq \frac{\kappa(n)}{n}$. □

Proof of Theorem 3. Fix an integer $m \geq 1$. It suffices to show that sample sizes in the set $\mathcal{D} = \{1, 3, ..., 2m-1\}$ are optimal for $t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right]$.

For sample size n, the sample (n, k) triggers the DM to take the action for t slightly above 1/2 if $\frac{k}{n} > \frac{1}{2}$. Thus, the DM needs to obtain at least a simple majority of successes in order to take the action in this case.

For an even sample size n and t in the relevant interval, $P(n, \lceil nt \rceil)$ is weakly smaller than $P(n, \frac{n+2}{2})$ because $\frac{n+2}{2} \leq \lceil nt \rceil$. The probability $P(n, \frac{n+2}{2})$ is in turn smaller than $P(n+1, \frac{n+2}{2})$ by identity (1) in the proof of Lemma 1. Thus, the expected persuasion probability $P(n, \lceil nt \rceil)$ of an even sample size n is dominated by the probability $P(n+1, \frac{n+2}{2})$.

For an odd sample size $n > 2m-1$ and t in the relevant interval, the expected persuasion probability $P(n, \lceil nt \rceil)$ is smaller than $P(n, \frac{n+1}{2})$ because $\frac{n+1}{2} < \lceil nt \rceil$ by the choice of n and t. Thus, to show that sample sizes in $\mathcal{D}$ are optimal, it suffices to show that $P(n, \frac{n+1}{2})$ is equal for all odd integers. This holds because for any $n = 2k-1$ and any symmetric distribution f, we have that

$$\begin{aligned} P(2k-1,k) &= \int_0^{1/2} P(2k-1,k,q)f(q)dq + \int_{1/2}^{1}(1-P(2k-1,k,1-q))f(q)dq \\ &= \int_0^{1/2} P(2k-1,k,q)f(q)dq + \int_0^{1/2}(1-P(2k-1,k,q))f(q)dq \\ &= \int_0^{1/2} f(q)dq = F(1/2) \end{aligned}$$

where the first equality holds because $P(2k-1,k,q) = 1 - P(2k-1,k,1-q)$ and the second equality holds by the symmetry of f. $\square$

Proof of Theorem 3B. Fix a positive integer m and $t \in \left(\frac{\alpha+m+1}{2\alpha+2m+1}, \frac{\alpha+m}{2\alpha+2m-1}\right]$. It suffices to show that sample sizes in the set $\mathcal{D} = \{1,3,...,2m-1\}$ are optimal.

For sample sizes in $\mathcal{D}$, a simple majority of successes suffices for taking the action as in unbiased inference. By the proof of Theorem 3, the corresponding expected persuasion probabilities are equal. The expected persuasion probability for an even sample size n is dominated by $P(n+1, \frac{n+2}{2})$ as in the proof of Theorem 3. Finally, for an odd sample size $n > 2m-1$, the smallest number of successes that triggers taking the action is weakly larger than in the case of unbiased inference. Thus, by the proof of Theorem 3, sample size n is dominated by sample sizes in the set $\mathcal{D}$. $\square$

Proof of Proposition 2. Fix an integer n' and $t \in \left(\frac{n'}{n'+1}, \frac{n'+1}{n'+2}\right]$. Let $n(k) = k(n'+1)-1$. The sample $(n(k), kn')$ triggers the DM to take the action if $\frac{kn'+1}{n(k)+1} \geq t$. The left-hand side of this inequality decreases in k and converges to $\frac{n'}{n'+1}$. Thus, there exists $\bar{k}(t)$ such that a sample $(n(k), kn')$ triggers the DM to take the action if and only if $k \leq \bar{k}(t)$. Let $\mathcal{D}$ be the corresponding set of sample sizes. We need to show that sample sizes in $\mathcal{D}$ are optimal.

Because $t > \frac{n'}{n'+1}$, the expected persuasion probability for sample sizes smaller than n' is 0. Because the prior is uniform, the expected persuasion probability for sample sizes in $\mathcal{D}$ is $P(n(k), kn') = 1 - \frac{n'}{n'+1} > 0$. The expected persuasion probability for any sample size $n(k)$ with $k > \bar{k}(t)$ is smaller because more than kn' successes are required for taking the action. For the remaining sample sizes, fix some sample size n and let k' be the largest integer such that the DM takes the action after obtaining the sample $(n, n-k'+1)$, i.e., $\frac{n-k'+2}{n+2} \geq t$. Combined with the inequality $t > \frac{n'}{n'+1}$, we obtain that $n > n(k')$. Consequently, $\frac{n-k'+1}{n+1} > \frac{n(k')-k'+1}{n(k')+1} = \frac{n'}{n'+1}$ where the first inequality holds because $n > n(k')$. The persuasion probability $P(n, n-k'+1) = 1 - \frac{n-k'+1}{n+1}$ is therefore dominated by the corresponding persuasion probabilities for sample sizes in the set $\mathcal{D}$. $\square$

Proof of Theorem 1C. We first identify $\overline{n}(c)$. Consider the function $\pi(n) = P(n,1) - cn$. The first term increases in n, is smaller than 1, and is concave in n because $P(n,1,q) - P(n-1,1,q) = q(1-q)^{n-1}$ for all $q \in (0,1)$. The second term is linear and tends to infinity as n tends to infinity. Therefore, the function $\pi(n)$ obtains its global maximum at some finite n, which we denote by $\overline{n}(c)$.[13] It is larger than 0 because $c < P(1,1)$.

Fix t and let $n' = \lfloor t^{-1} \rfloor$. Theorem 1 states that sample sizes larger than n' are dominated by sample size n' when data provision is costless. They continue to be dominated with costly data provision. For sample sizes weakly smaller than n', even a single success triggers the DM to take the action, and thus the relevant function for optimization is $\pi(n)$. If $n' \leq \overline{n}(c)$, then $\pi(n)$ increases till n', which is the optimal sample size. Otherwise, $\pi(n)$ increases up to sample size $\overline{n}(c)$, after which it weakly decreases. Thus, sample size $\overline{n}(c)$ is optimal. $\square$

Proof of Theorem 2C. We first identify $\overline{n}(c)$. Consider the function $\pi(n) = P(2n-1, n) - (2n-1)c$. The first term is smaller than 1, and by Lemma A3 proved below, it is concave and increasing in n. The second term is linear and tends to infinity as n tends to infinity. Therefore, the function $\pi(n)$ obtains its global maximum at some positive n, which we denote by $\overline{n}(c)$.

[13] If the maximum is obtained in two adjacent n's, we use the smaller one.

Fix $t \in \left(\frac{m+1}{2m+1}, \frac{m}{2m-1}\right]$. Then $2m-1$ is the largest odd sample size weakly smaller than $\lfloor (2t-1)^{-1} \rfloor$. Theorem 2 implies that sample sizes larger than $2m-1$ are suboptimal. Any even sample size $n < 2m-1$ is dominated by sample size $n-1$ because the relevant persuasion probability for sample size n, $P(n, n/2+1, q)$, satisfies $P(n, n/2+1, q) = P(n-1, n/2, q) - (1-q)b(n-1, n/2, q)$ where $P(n-1, n/2, q)$ is the relevant probability for sample size $n-1$. For the remaining sample sizes, the proof is analogous to that of Theorem 1C. $\square$

Lemma A3. *For any increasing prior, $P(2n-1, n)$ is concave and increasing in n.*

Proof. Fix $n \in \mathcal{N}$ and let $\delta_n(q) = P(2n-1, n, q) - P(2n-3, n-1, q)$. Then

$$\delta_n(q) = \binom{2n-2}{n-1} q^{n-1}(1-q)^{n-1}\left(q - \frac{1}{2}\right).$$

To verify this equality, we observe that $P(2n-1, n, q) = P(2n-2, n, q) + qb(2n-2, n-1, q)$ by identity (1) in the proof of Lemma 1. Adding and subtracting $(1-q)b(2n-2, n-1, q)$ to the right-hand side, we obtain that $P(2n-1, n, q) = P(2n-2, n-1, q) - (1-q)b(2n-2, n-1, q)$. Identity (1) also implies that $P(2n-3, n-2, q) = P(2n-2, n-1, q) - qb(2n-3, n-2, q)$. Thus, $\delta_n(q) = qb(2n-3, n-2, q) - (1-q)b(2n-2, n-1, q)$ and the above equality follows.

The function $\delta_n(q)$ is negative on $(0, 1/2)$ and is an odd function around $1/2$. Therefore, $\int_0^1 \delta_n(q)dq = 0$. Because an increasing prior shifts mass, relative to the uniform prior, from $(0, 1/2)$ to $(1/2, 1)$, we obtain that $\Delta_n > 0$, where Δ_n denotes the expected value of δ_n with respect to the designer's increasing prior. The probability $P(2n-1, n)$ therefore increases in n.

The function $\delta_n(q)$ crosses the function $\delta_{n+1}(q)$ once and from below at $q = 1/2$. This follows from the fact that $\delta_{n+1}(q) = 2q(1-q)\delta_n(q)$. Since an increasing prior shifts mass to q's in which δ_n dominates δ_{n+1}, $\Delta_{n+1} < \Delta_n$ and the concavity of $P(2n-1, n)$ follows. $\square$

References

Al-Najjar, Nabil I. 2009. "Decision Makers as Statisticians: Diversity, Ambiguity, and Learning." *Econometrica*, 77(5): 1371–1401.

Banerjee, Abhijit V., Sylvain Chassang, Sergio Montero, and Erik Snowberg. 2020. "A Theory of Experimenters: Robustness, Randomization, and Balance." *American Economic Review*, 110(4): 1206–1230.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *American Economic Review*, 102(4): 1279–1309.

Di Tillio, Alfredo, Marco Ottaviani, and Peter Norman Sørensen. 2021. "Strategic Sample Selection." *Econometrica*, 89(2): 911–953.

Efron, Bradley. 2005. "Bayesians, Frequentists, and Scientists." *Journal of the American Statistical Association*, 100(469): 1–5.

Eliaz, Kfir, Ran Spiegler, and Heidi C. Thysen. 2021. "Strategic Interpretations." *Journal of Economic Theory*, 192: 105192.

Galperti, Simone. 2019. "Persuasion: The Art of Changing Worldviews." *American Economic Review*, 109(3): 996–1031.

Glazer, Jacob, and Ariel Rubinstein. 2012. "A Model of Persuasion with Boundedly Rational Agents." *Journal of Political Economy*, 120(6): 1057–1082.

Haldane, J. B. S. 1932. "A Note on Inverse Probability." *Mathematical Proceedings of the Cambridge Philosophical Society*, 28(1): 55–61.

Kamenica, Emir. 2019. "Bayesian Persuasion and Information Design." *Annual Review of Economics*, 11: 249–272.

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

Levy, Gilat, Inés Moreno de Barreda, and Ronny Razin. 2022. "Persuasion with Correlation Neglect: A Full Manipulation Result." *American Economic Review: Insights*, 4(1): 123–38.

Mantilla, César, Rajiv Sethi, and Juan Camilo Cárdenas. 2020. "Efficiency and Stability of Sampling Equilibrium in Public Goods Games." *Journal of Public Economic Theory*, 22(2): 355–370.

Osborne, Martin J., and Ariel Rubinstein. 1998. "Games with Procedurally Rational Players." *American Economic Review*, 88(4): 834–847.

Osborne, Martin J., and Ariel Rubinstein. 2003. "Sampling Equilibrium, with an Application to Strategic Voting." *Games and Economic Behavior*, 45(2): 434–441.

Salant, Yuval, and Josh Cherry. 2020. "Statistical Inference in Games." *Econometrica*, 88(4): 1725–1752.

Sethi, Rajiv. 2000. "Stability of Equilibria in Games with Procedurally Rational Players." *Games and Economic Behavior*, 32(1): 85–104.

Sethi, Rajiv. 2021. "Stable Sampling in Repeated Games." *Journal of Economic Theory*, 197: 105343.

Spiegler, Ran. 2006*a*. "Competition Over Agents with Boundedly Rational Expectations." *Theoretical Economics*, 1(2): 207–231.

Spiegler, Ran. 2006*b*. "The Market for Quacks." *Review of Economic Studies*, 73(4): 1113–1131.

Wald, Abraham. 1949. "Statistical Decision Functions." *The Annals of Mathematical Statistics*, 165–205.