Persuading Statisticians

Sanket Patil and Yuval Salant

Northwestern University

October 19, 2020

Abstract

A decision maker (DM) contemplates whether to take a costly action. The DM does not know the action's value and relies on data and unbiased statistical inference to estimate it. The data are Bernoulli experiments governed by the action's value. A designer, who wishes the DM to take the action, controls the size of the data, i.e., the sample size, available to the DM. We establish that in many environments the designer's optimal sample size is the largest one satisfying that either a single — or a simple majority — of favorable realizations would persuade the DM to take the action.

1 Introduction

Decision makers (DMs) often rely on data and statistical inference to make choices. For example, buyers may experiment with a product in order to estimate its value and then make a purchase decision based on the estimate. Likewise, politicians use various statistics of polls and public opinion surveys in their decision making. In both cases, an interested designer may control the size of the data obtained by the DM but not necessarily its distribution. In the context of product experimentation, it is the seller who decides how much experimentation to allow potential buyers, and the buyers conduct the experiments themselves by interacting with the product. And in the context of polls and surveys, some interested party (e.g., lobbyist, think tank, or news agency) often decides how extensive a poll to commission, but exogenous factors such as the politician's own interests dictate the question and the subject population. The goal of this paper is to study the designer's optimal sample size in such settings.

We consider a designer who decides how large a sample about a payoff-relevant parameter to provide to a decision maker (DM). The designer wishes to persuade the DM to take an action. The DM is a statistician who uses the data and statistical inference to estimate the value of the parameter, and takes the action if the parameter estimate is larger than his outside option.

The data are Bernoulli experiments governed by the underlying value of the parameter. For example, in surveys, the parameter may be the proportion of individuals supporting a particular issue, and each Bernoulli experiment corresponds to the opinion of a single survey respondent. A similar setting is meetings where the DM relies on the opinions of participants to arrive at a decision. Bernoulli experiments may also fit product experimentation when the product is a platform with many small brands, the parameter is the fraction of brands that fit the buyer's taste, and each experiment corresponds to the DM experimenting with a single brand. Of course, there are many other settings where an experiment can reveal more information about the parameter value.

As for statistical inference, there are roughly speaking two classes of inference procedures: those that rely only on the data (frequentist inference), and those that start with a prior belief on the parameter value and Bayes-update it based on the sample (Bayesian inference). The advantages of the frequentist approach are that it is prior-free, does not require knowledge of conditional probabilities, and uses only objective data.[1] As such, the frequentist approach fits

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

environments in which the DM is less experienced, less knowledgeable about fundamentals, or does not want prior beliefs to influence its decision making.

Among frequentist inference procedures, an important subclass is the class of unbiased inference procedures whereby the expected value of an estimate (which we allow to be a distribution over possible parameter values) is identical to the proportion of successes in the sample. Leading examples include maximum likelihood estimation, maximum entropy estimation, and Beta estimation. Bayesian inference is not unbiased because the expected value of an estimate depends on the prior. With the exception of Section 5.2 that discusses Bayesian inference, we focus on unbiased inference procedures throughout the paper.

Our first main result is about environments in which the designer believes the distribution of the payoff-relevant parameter is decreasing. We establish that the designer's optimal sample size is the *largest* one satisfying that even a single favorable realization would persuade the DM to take the action. In other words, for every integer n, sample size n is optimal when the DM's outside option is between $1/(n+1)$ and $1/n$.

This result has several implications. First, any possible sample size is optimal for a non-trivial interval of the DM's outside option. Second, as the DM's outside option becomes more attractive, so that the DM's tendency to take the action decreases for any sample size, the designer's optimal sample size decreases. Third, for an outside observer, who has access to the DM's choices as a function of the data realizations, the DM's choice behavior may seem to exhibit "strong" preference for taking the action. This is because the DM chooses to take the action unless there is unanimity against doing so. To be sure, the DM is an unbiased statistician whose preference for taking the action is neither strong nor weak. Still, the designer's optimal choice of sample size creates the impression that the DM has a strong preference for taking the action.

In the second type of environments we consider, the designer believes the distribution of the payoff-relevant parameter is increasing and concave. We establish that when the value of the outside option is not small (formally, larger than $1/2$), the designer's optimal sample size is the *largest* odd sample size satisfying that a simple majority of successes would persuade the DM to take the action. More specifically, for every integer m, sample size $2m-1$ is optimal when the DM's outside option is between $(m+1)/(2m+1)$ and $m/(2m-1)$ so that $2m-1$ is the

nice recent discussion.

largest sample size satisfying the above condition.

Thus, similarly to the case of decreasing beliefs, any odd sample size is optimal for some interval of the DM's outside option, and the optimal sample size decreases as the value of the DM's outside option increases. In addition, an outside observer with access to the DM's choices may conclude that the DM uses a simple majority decision rule when choosing whether to take the action. Put differently, simple majority arises endogenously in the setting with increasing and concave beliefs.

A potential issue with our results is that a sophisticated DM may be able to strategically infer something about the designer's beliefs regarding the parameter from the designer's choice of sample size. The DM may then consider this information, together with statistical inference from the data, when making a choice. We establish that a designer with such a concern would choose some sample size satisfying that a simple majority of successes would persuade the DM to take the action, thus shutting down this channel for strategic inference.

In the third environment we consider, there is a seller and a buyer. The seller is the owner of a platform that offers many brands, some match a given buyer's taste and some do not. The seller decides how much experimentation with the platform to allow buyers prior to joining the platform, a platform membership fee, and a platform usage fee that depends on the number of brands consumed by the buyer after joining. Buyers use an unbiased inference procedure to estimate the proportion of brands on the platform that match their taste, and then decide whether to join the platform and pay the membership and usage fees.

We establish that the seller's joint choice of experimentation amount and fees is to provide the *smallest* possible amount of product experimentation to potential buyers, charge a monopoly membership fee against the resulting demand, and charge no usage fee. Put differently, when the designer can extract some of the DM's estimated value from taking the action through transfers, margin considerations dominate the volume considerations identified above.

This paper is related to two literatures. The first is the small game-theoretic literature on sampling (Osborne and Rubinstein (1998, 2003)) and statistical inference (Salant and Cherry (2020)). In Osborne and Rubinstein (1998)'s $S(k)$-equilibrium, players who 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) consider players who obtain data on other players' actions, use statistical inference to form an estimate about these actions, and best respond to the estimate. Sethi (2000, 2019) and Mantilla, Sethi and Cárdenas (2020) study the dynamic stability properties of $S(1)$-equilibria. Spiegler (2006a,b) studies competition between firms that face consumers who sample their prices or another payoff-relevant parameter once.

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 a different question. We treat the sample size as a design parameter and solve for the optimal sample size.

The second related literature is the literature on persuasion, which studies receiver-optimal persuasion (Glazer and Rubinstein (2004)) and sender-optimal persuasion (Kamenica and Gentzkow (2011)). We consider sender-optimal persuasion. A key assumption in models of sender-optimal persuasion, or Bayesian persuasion, is that the set of tools available to the designer is rich: the designer can choose any data-generating process that results in a distribution of posteriors whose expectation equals the prior distribution. In contrast, the set of tools available to the designer in our setup is very limited. Rather than choosing the data-generating process, the designer only chooses how much data the DM obtains. As discussed above, this assumption seems plausible when the experiments are conducted by the DM (product experimentation) or a third party, or when the experimental question and subject population cannot be easily changed (polls and surveys). On the technical side, the departure from the rich signal structure assumption requires developing new tools and solution techniques.

We also depart from another important assumption of the Bayesian persuasion literature whereby the designer and the DM are Bayesian and share a common prior. For most of the analysis, we consider a designer who merely considers the prior to be monotone and a frequentist DM.[2] In this respect, our paper is related to a small literature on persuasion with non-Bayesian DMs. Glazer and Rubinstein (2012) consider the receiver's optimal persuasion protocol against a sender who has limited abilities to find a persuasive message. In Levy, de Barreda and Razin (2018), the DM makes updating mistakes because he fails to account for the correlation between different information sources. In Eliaz, Spiegler and Thysen (2020), DMs who obtain a multi-

[2] Section 5.2 considers a Bayesian environment with a uniform prior and compares the optimal sample sizes to those obtained for a frequentist DM.

dimensional signal focus only on a subset of its dimensions, as directed by an informed designer, and do not make inferences on other dimensions from the designer's choice of dimensions. And in Galperti (2019)'s model of changing world views, the DM rejects his original prior in favor of a new prior when observing evidence that is inconsistent with the original prior.

We proceed as follows. Section 2 presents the model. Section 3 analyzes the optimal sample size. Section 4 considers the application to platform experimentation and pricing. Section 5 concludes with (i) extending the analysis to include upper and lower bounds on the feasible sample sizes, and (ii) comparing the predictions of unbiased inference to Bayesian inference from a uniform prior. Appendix A contains proofs that do not appear in the main text. Appendix B contains technical combinatorial Lemmas that are used in the proofs.

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, where a successful experimental realization, or simply a success, is interpreted as a data point in favor of taking the action. The size of the data is decided by a designer who wishes the DM to take the action. The designer knows t and has a prior f on q.[3] To avoid technical issues, we assume that the designer can choose any data size or ∞, which is interpreted as fully revealing the value of q. We consider lower and upper bounds on the size of the data in Section 5.1.

After the designer decides the data size, the Bernoulli experiments are carried out, 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 successful realizations. We will refer to n as the sample size and to k/n as the sample mean.

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

Definition. *An* ***inference procedure*** $G = \{G_{n,k}\}$ *assigns a cumulative distribution function*

[3]For most of the analysis, the designer need not have an exact prior on q, but rather believe the prior satisfies some regularity condition.

$G_{n,k}$, *called an estimate, to every sample* (n, k) *such that:*

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

(ii) the estimate $G_{n,n}$ *strictly first-order stochastically dominates the estimate* $G_{n,0}$*.*[4]

An inference procedure is the analogue of an estimator in the statistics literature. It can be used to describe a rich class of inferences from data. Here are a few examples.

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

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.*

Example 3 (Beta Estimation)**.** *The DM wishes to conduct Bayesian updating relying as little as possible on a prior belief. The DM starts with Haldane's "prior" (Haldane (1932)), which is not a proper prior, and updates it to the Beta*$(k, n-k)$ *distribution after obtaining the sample* (n, k)*.*[5]

To understand this procedure better, recall that the Beta distribution is a conjugate prior for Binomial distributions: a DM with a Beta(a, b) *prior on* q*, who obtains the sample* (n, k) *and uses Bayesian updating, would arrive at the posterior Beta*$(a + k, b + (n - k))$*. One may think about* $(a + b)$ *as measuring the weight the DM puts on the prior and about* n *as the weight the DM puts on the sample. Haldane's "prior" (Haldane (1932)) is the limit of the Beta*(ϵ, ϵ) *distribution as* $\epsilon \to 0$*, and the Beta*$(k, n-k)$ *distribution is the limit of the resulting posteriors after obtaining the sample* (n, k)*. As* $\epsilon \to 0$*, the DM puts no weight on the prior and bases the estimate only on the sample.*

Example 4 (Maximum Entropy)**.** *The DM follows the Principle of Maximum Entropy. From among all distributions with expected value equal to the sample mean, the DM searches for the one with the maximal uncertainty in terms of entropy. By Conrad (2004), when the sample mean is in* $(0, 1)$*, this distribution is a truncated exponential distribution with density* $g_{n,k}(q) = Ce^{\alpha q}$*. The values of* α *and* C *are determined uniquely by the constraints that (1)* $g_{n,k}$ *is a density*

---

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

[5] When the sample mean is 0 or 1, the DM concentrates the estimate on 0 or 1 respectively.

function (i.e., $\int_0^1 g_{n,k}(q)dq = 1$), and (2) the expected value of the estimate is equal to the sample mean (i.e., $\int_0^1 q g_{n,k}(q)dq = k/n$). For example, when the sample mean is 1/2, the unique α is 0*, the unique C is $C = 1$, and the resulting estimate is the uniform distribution over* $[0, 1]$.[6]

Example 5 (Dogmatic Views). *The DM believes q is distributed either according to the CDF F_0 or F_1 that strictly first-order stochastically dominates F_0. He uses the sample to decide which distribution should be used in decision making. If $k(n)$ or more realizations are successes, he uses F_1. Otherwise he uses F_0.*

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. Because of 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 $(n, \hat{k})$ for $\hat{k} > k$. Let

$$P(n, k, q) = \sum_{j=k}^{n} \binom{n}{j} q^j (1-q)^{n-j} = \sum_{j=k}^{n} b(n, j, q)$$

denote the corresponding probability of taking the action after obtaining $\geq k$ successes where $b(n, j, q)$ denotes the probability of obtaining j successes in n Bernoulli experiments governed by q.

Let $k(n, G, t)$ denote the smallest number of successes satisfying that a DM who uses the inference procedure G takes the action after observing the sample $(n, k(n, G, t))$.[7] The probability that the DM takes the action is:

$$P(n, G, t, q) = P(n, k(n, G, t), q).$$

The objective of the designer is to maximize the expected value, according to f, of the probability of taking the action, i.e. solve

$$\arg\max_{n \in \mathcal{N} \cup \infty} \int_0^1 P(n, G, t, q) f(q) dq.$$

With the exception of Section 5.2, our analysis focuses on unbiased inference procedures.

[6]When the sample mean is 0 or 1, the unique distribution with expected value equal to the sample mean puts a unit mass on the sample mean.

[7]It is possible that there are sample sizes n for which such $k(n, G, t) \leq n$ does not exist. Such sample sizes are never optimal, and so we ignore them.

Definition. *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 procedure is unbiased. The Beta Estimation and Maximum Entropy procedures are also unbiased. Bayesian inference is not unbiased because the expected value of an estimate depends on both the prior and the sample, whereas unbiasedness requires that the expected value depends only of the sample.

For an unbiased inference procedure G, the smallest number of successes leading the DM to take the action is $k(n,t) = \lceil tn \rceil$. Because this number does not depend on the specifics of G, we can write

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

for any unbiased inference procedure.

3 Optimal sample size

This section studies the designer's optimal sample size for unbiased inference procedures. Section 3.1 establishes a general lower bound on the optimal sample size. Section 3.2 analyzes the optimal sample size when the designer's prior is degenerate, i.e., assigns probability 1 to a particular q. This case highlights many of the forces that would determine the optimal sample size when the designer's prior is non-degenerate. Section 3.3 studies the designer's optimal sample size for priors that are either monotone or symmetric.

3.1 Lower bound

To obtain a lower bound on the optimal sample size for any prior on q, it is helpful to consider the case in which designer assigns probability 1 to some fixed q, which is smaller than t. In this case, it may not be optimal for the designer to provide the least amount of information about q, i.e., sample size 1, to the DM. This is because a DM, who knows more about q in the sense of obtaining a larger sample, may take the action with a higher probability than a DM who knows less about q. For example, if $t < 1/2$, then the probability of taking the action after obtaining two data points $P(2, t, q) = q^2 + 2q(1 - q)$ is larger than the probability of taking the action after obtaining a single data point $P(1, t, q) = q$ for any q, implying that for $t < 1/2$ sample size

1 would not be optimal. Observation 1 builds on this example to establish a lower bound on the optimal sample size.

Observation 1. *For any t and any prior on q, the optimal sample size is weakly larger than the largest sample size satisfying that even a single success would trigger the DM to take the action.*

In other words, for any integer n and any prior on q, the optimal sample size is weakly larger than n when the value t of the outside option is in the interval $(0, 1/n]$.

To obtain the lower bound in Observation 1, let n' be the largest sample size that triggers the DM to take the action after observing the sample $(n', 1)$. Then $n' = \lfloor t^{-1} \rfloor$. For any sample size $n \leq n'$, a single success would also trigger the DM to take the action because $1/n \geq 1/n' \geq t$. Thus, the probability of taking the action is $1 - (1-q)^n$ for $n \leq n'$. This probability increases in n, and so sample size n' dominates smaller sample sizes for any fixed q and, therefore, for any prior on q.

3.2 Degenerate priors

When the designer assigns probability 1 to some fixed q, full revelation is optimal if and only if $q \geq t$. Indeed, with full revelation, the DM takes the action with probability 1 when $q \geq t$ and with probability 0 when $q < t$, whereas the probability of taking the action is strictly between 0 and 1 for any finite sample size.

The more interesting case is $q < t$. Observation 1 implies that the optimal sample size is weakly larger than $\lfloor t^{-1} \rfloor$ in this case. To see that larger sample sizes have the potential be be optimal, let us consider sample sizes 1 and 3. When $t \leq 1/3$, the discussion following the statement of Observation 1 implies that sample size 3 dominates sample size 1. When t exceeds the "critical point" 1/3, the probability $P(3, t, q)$ "loses" the term $b(3, 1, q)$ in the sum on the right-hand side of Equation (1) because a single success no longer triggers the DM to take the action for $t > 1/3$. This leads to a discrete drop in $P(3, t, q)$. Note, however, that the marginal increase in $P(3, t, q)$ is larger than the marginal increase in $P(1, t, q)$. For $q < 1/2$, sample size 1, which did not lose a term, dominates sample size 3. At $q = 1/2$, sample size 3 catches up with sample size 1, and it dominates sample size 1 from that point until the next critical point at 2/3, i.e., for $1/2 < q < t \leq 2/3$. When t exceeds 2/3, $P(3, t, q)$ loses another term, $b(3, 2, q)$, and sample size 1 dominates it again. The left Panel in Figure 1 provides a graphical illustration. The right panel in Figure 1 illustrates a similar phenomenon for sample sizes 3 and 5.

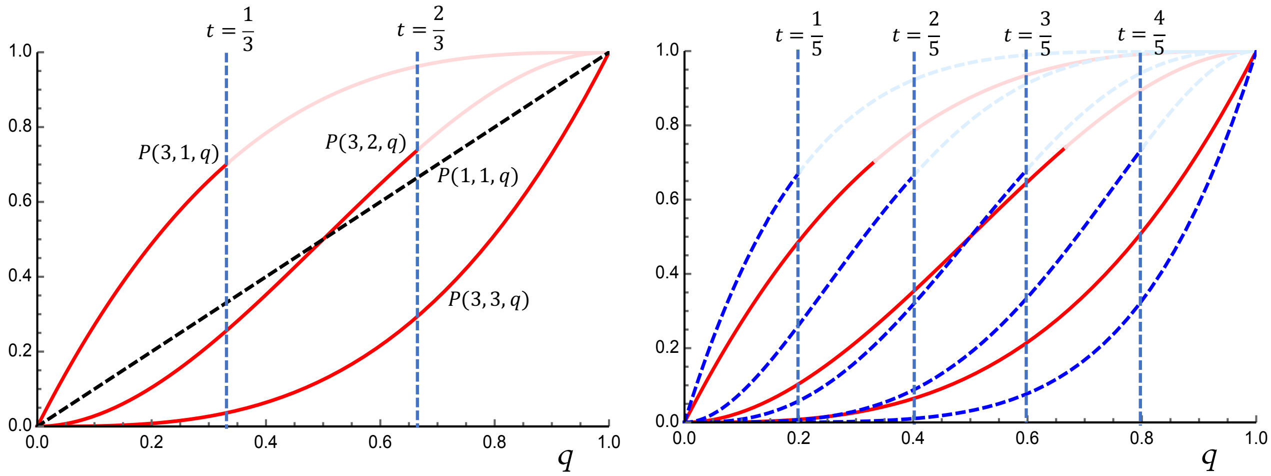

Figure 1: Probability of taking the action. Left panel depicts $P(1, t, q)$ (dashed) and $P(3, t, q)$ (solid). Right panel depicts $P(3, t, q)$ (solid) and $P(5, t, q)$ (dashed).

More generally, the discrete drop in $P(n, t, q)$ at critical points of the form k/n followed by a continuous increase of $P(n, t, q)$ as q moves toward t implies that larger sample sizes than those identified in Observation 1 may be optimal. Theorem 1 establishes that this does not happen when q is not too close to t. And Theorem 2 establishes that it does when q is close to t.

Theorem 1. *For any t and $q < (1 + \lfloor t^{-1} \rfloor)^{-1}$, the optimal sample size is the largest one satisfying that even a single success would trigger the DM to take the action.*

Put differently, for any integer n, sample size n is optimal when $q < 1/(n + 1) < t \leq 1/n$. The shaded rectangles in Figure 2 provide a graphical illustration.

Theorem 1 has several implications. First, for any sample size, there are non-trivial intervals of parameter values for which this sample size is optimal. Clearly, providing the least amount of information to the DM need not be optimal. Second, as Figure 2 illustrates, fixing the value q of taking the action, as the status quo becomes more attractive, the optimal sample size gradually decreases. In other words, the designer chooses to provide less information to the DM as the DM's ex-ante tendency to take the action becomes smaller. Third, consider an analyst who observes the DM's choices as a function of the data the DM obtains. The analyst will conclude that the DM's choice behavior is consistent with a very strong preference for taking the action: The DM chooses to take the action unless all realizations are in favor of keeping the status quo. These implications extend to the case in which the designer's prior is decreasing in q, as Theorem 3 will establish.

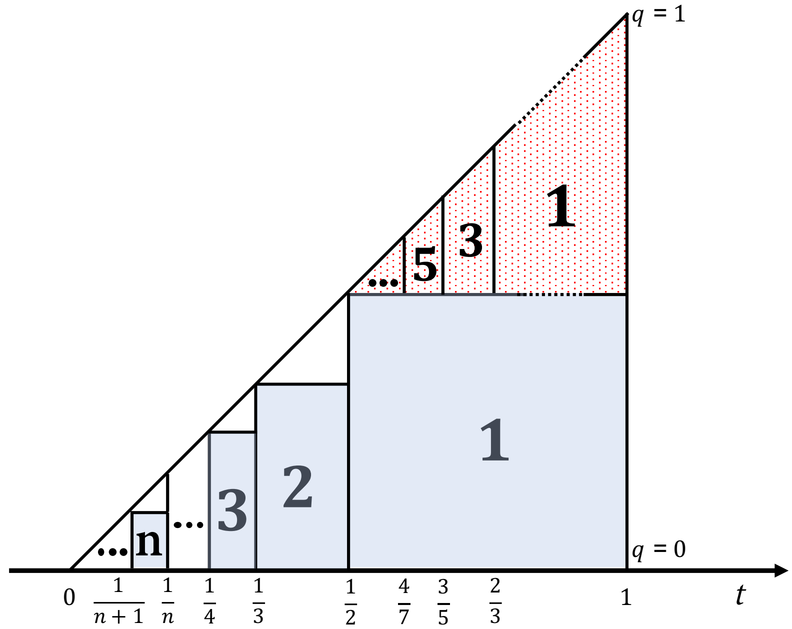

Figure 2: Optimal sample sizes for degenerate priors

Proof of Theorem 1. Fix an integer n' and let $q < 1/(n'+1) < t \leq 1/n'$. We establish the optimality of sample size n' in two steps. The first step identifies a collection of sample sizes $\mathcal{D}$ (including sample size n') that are candidates for optimality.

Step 1. For every $k \geq 1$, let $n(k)$ be the largest integer such that $k/n(k) > 1/(n'+1)$. Then $n(k) = (n'+1)k - 1$. Let $\mathcal{D} = \{n(k) \mid k \geq 1\}$. Then, any sample size $n \notin \mathcal{D}$ is dominated by a sample size in $\mathcal{D}$.

Proof. Fix a sample size $n \notin \mathcal{D}$ and let k be the minimal number of successes that trigger the DM to take the action if he obtains the sample (n, k). Then $k = \lceil tn \rceil$. By the definition of $n(k)$, we have that $n < n(k)$. Applying the combinatorial identity

$$P(m+1, l, q) = P(m, l, q) + qb(m, l-1, q) \tag{2}$$

multiple times establishes the desired domination. □

The second step establishes that sample size n' dominates all other sample sizes in the set $\mathcal{D}$ and concludes the proof.

Step 2. For $q < 1/(n'+1)$ and $k \geq 2$, sample size n' dominates sample size $n(k)$.

Proof. By Lemma 1 in Appendix B, there exists q^* such that $P(n', 1, q) > P(n(k), k, q)$ for $q \in (0, q^*)$ and a reverse inequality holds for $q \in (q^*, 1)$. By Lemma 2, $P(n', 1, 1/(n'+1)) \geq$

$P(n(k), k, 1/(n'+1))$ implying that $q^* \geq 1/(n'+1)$. Thus, sample size n' dominates any sample size $n(k)$ for $q < 1/(n'+1)$. $\square$

$\square$

Theorem 2 studies the optimal sample size when the value q of taking the action is close to the value t of the outside option.

Theorem 2. *For $\frac{1}{2} < q < t$, the optimal sample size is the largest odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.*

The dotted trapezoids in Figure 2 illustrate the optimal sample sizes identified in Theorem 2.

Thus, similarly to Theorem 1, any odd sample size is optimal for some interval of parameter values, and the optimal sample size decreases as the value of the status quo increases. In addition, an analyst who observes the DM's choices will conclude that the DM's choice procedure is consistent with simple majority: the DM takes the action when a majority of the realizations are in favor of doing so. These implications continue to hold when the designer's prior is increasing and concave, as Theorem 4 will show.

The intuition for why the largest odd sample size satisfying the conditions of Theorem 2 dominates smaller sample sizes is as follows. Fix a sample size $n' = 2m-1$ and assume $\frac{m+1}{2m+1} < t \leq \frac{m}{2m-1}$, i.e., the value of the status quo is in the range in which sample size n' should be optimal. For any sample size $n \leq 2m-1$, the DM takes the action when obtaining a simple majority of successes. Since this number is the same for the even sample size $2l$ and the odd sample size $2l+1$, the combinatorial identity (2) implies that even sample sizes are dominated by odd sample sizes. Thus, it suffices to focus on odd sample sizes.

For any odd sample size n, we have $P(n, 1/2, 1/2) = 1/2$ because the binomial distribution is symmetric around $1/2$ when $q = 1/2$. We can thus write

$$P(n, t, q) = 1/2 + \int_{1/2}^{q} P'(n, t, s)ds. \tag{3}$$

where P' denotes the marginal increase in the probability of taking the action, and is equal to $k\binom{n}{k}q^{(k-1)}(1-q)^{(n-k)}$ where $k = \lceil (n+1)/2 \rceil$. This marginal increase increases in n for $n \leq 2m-1$ in the relevant range of t and q implying that sample size $2m-1$ dominates smaller odd sample sizes.

3.3 Non-degenerate priors

We now extend the analysis to monotone or symmetric priors.

Our first result establishes that the optimal sample size identified in Theorem 1 is also optimal for decreasing priors where a prior is *decreasing (increasing)* if it weakly decreases (increases) in q on $[0,1]$ and differs from the uniform prior on a non-zero measure.

Theorem 3. *For any decreasing prior and any t, the optimal sample size is the largest one satisfying that even a single success would trigger the DM to take the action.*

Proof of Theorem 3. We first consider finite sample sizes. By Observation 1 and Step 1 in the proof of Theorem 1, it suffices to show that sample size $n' = \lfloor t^{-1} \rfloor$ dominates any sample size of the form $n = k(n'+1) - 1$ where $k \geq 2$, or, more specifically, that the expected probability $P(n', 1) = \int_0^1 P(n', 1, q) f(q) dq$ is strictly larger than $P(k(n'+1)-1, k) = \int_0^1 P(k(n'+1) - 1, k, q) f(q) dq$ where f is the designer's prior. We show this domination in two steps. The first step establishes that it suffices to examine the expected probabilities with respect to the uniform prior.

Step 1. If $P(n', 1) \geq P(k(n'+1)-1, k)$ for the uniform prior, then $P(n', 1) > P(k(n'+1)-1, k)$ for a decreasing prior.

Proof. By Lemma 1, $P(n', 1, q) > P(k(n'+1) - 1, k, q)$ for $q \in (0, q^*)$ and a reverse inequality holds for $q \in (q^*, 1)$ if $q^* < 1$. The result follows because for a decreasing prior, the mass on values in the interval $(0, q^*)$ is strictly larger than under the uniform prior. □

The second step examines the expected probabilities according to the uniform prior.

Step 2. For the uniform prior, $P(n', 1) = P(k(n'+1) - 1, k)$.

Proof. For the uniform prior, $P(n, k) = 1 - \frac{k}{n+1}$ because (i) $P(n, k, q)$ is equal to the incomplete regularized Beta function $I_q(k, n-k+1)$ and (ii) $I_q(k, n-k+1)$ is the CDF of the $Beta(k, n-k+1)$ distribution with mean $\frac{k}{n+1}$. Thus,

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

□

As for full revelation, by using the reasoning of Step 1 above, it suffices to show that sample size n' dominates full revelation for the uniform prior. This is true because the expected value of full revelation in this case is $1-t$ and $t > 1/(n'+1)$. $\square$

We proceed to analyze increasing priors. We first identify a lower bound on the optimal sample size that is tighter than the one obtained in Observation 1. We then show this bound achieves optimality when the prior is also concave.

For a fixed t, let $n' = \lfloor t^{-1} \rfloor$ and let $a_k = k/(k(n'+1)-1)$ where $k \geq 1$. Because the sequence $\{a_k\}$ decreases in k and converges to $1/(n'+1) < t$ as k tends to infinity, there exists a unique k' such that $a_{k'+1} < t \leq a_{k'}$.

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

Proof. By definition, after obtaining the sample $(k'(n'+1)-1, k')$, the DM takes the action. Fix a sample size $n < k'(n'+1)-1$. The minimal number of successes that trigger the DM to take the action is $k \leq k'$. By Step 1 in Theorem 1, sample size n is a candidate for optimality only if has the form $n = k(n'+1)-1$. An analogous argument to the one made is Step 1 of Theorem 3 can be used to show that if $k < k'$ and $P(k'(n'+1)-1, k') \geq P(k(n'+1)-1, k)$ for the uniform prior, then $P(k'(n'+1)-1, k') > P(k(n'+1)-1, k)$ for increasing priors. The inequality $P(k'(n'+1)-1, k') \geq P(k(n'+1)-1, k)$ holds by Step 2 in Theorem 3. $\square$

Note that the lower bound identified in Proposition 1 is identical to the one identified in Observation 1 for $t \in (1/(n'+1/2), 1/n']$. However, for $t \in (1/(n'+1), 1/(n'+1/2)]$, the lower bound identified in Proposition 1 is tighter. For example, if $t > 1/2$, then the bound of Observation 1 implies that any sample size can be optimal whereas Proposition 1 implies that the smallest candidate for optimaility is the largest sample size satisfying that a simple majority of successes would trigger the DM to take the action. The following result shows that this candidate is indeed optimal when the prior is also concave.

Theorem 4. *For any increasing and concave prior and any $t > 1/2$, the optimal sample size is the largest one satisfying that a simple majority of successes would trigger the DM to take the action.*

Concavity is important for establishing Theorem 4. It guarantees that the prior does not assign "too large a mass" to large values of q. When it does, very large sample sizes or full revelation may be optimal. To illustrate this point, fix some q' and consider a prior that assigns probability 0 to all q below q'. For this prior, full revelation is optimal for any $t \leq q'$ because the probability of taking the action for full revelation is 1 whereas it is strictly smaller than 1 for any finite sample size. Concavity precludes such priors.

Some lower bound on the value of the status quo is also important. (Note the difference from Theorem 3 which holds for any value of t.) To see why, consider the probability of taking the action for the linear prior $h(q) = 2q$. It can be shown that as t decreases, the probability of taking the action for full revelation increases sufficiently fast so that full revelation becomes uniquely optimal. Because the expectation operator is continuous, full revelation continues to be optimal for nearby concave priors.

For increasing and convex priors, a stronger result on the optimaility of full revelation can be obtained.

Proposition 2. *For all increasing and convex priors with $f(0) = 0$, there exists a single $t' > 0$ such that full revelation is optimal for all $t \leq t'$.*

The last class of priors that we consider is the class of symmetric priors.

Proposition 3. *For any symmetric prior with a non-zero mass around 1/2 and any $t > 1/2$, all odd sample sizes satisfying that a simple majority of successes would trigger the DM to take the action are optimal.*

Proof. We first observe that any even sample size $2n$ satisfies the combinatorial identity $P(2n+1, n+1, q) = P(2n, n+1, q) + qb(2n, n, q)$ and is thus dominated by sample size $2n + 1$.

Any odd sample size $2n+1$ satisfies the equality $P(2n+1, n+1, q) = 1-P(2n+1, n+1, 1-q)$. Thus, for any symmetric prior f, the expected probability $P(2n+1, n+1)$ is equal to $F(1/2) = \int_0^{1/2} f(q)dq$ and is independent of n. This is because:

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

where the first equality holds because $P(2n+1, n+1, q) = 1 - P(2n+1, n+1, 1-q)$ and the second equality holds by the symmetry of f. Moreover, $P(2n+1, n+1, q)$ is strictly larger than $P(2n+1, k, q)$ for $k > n+1$. Because the probability of taking the action is equal to

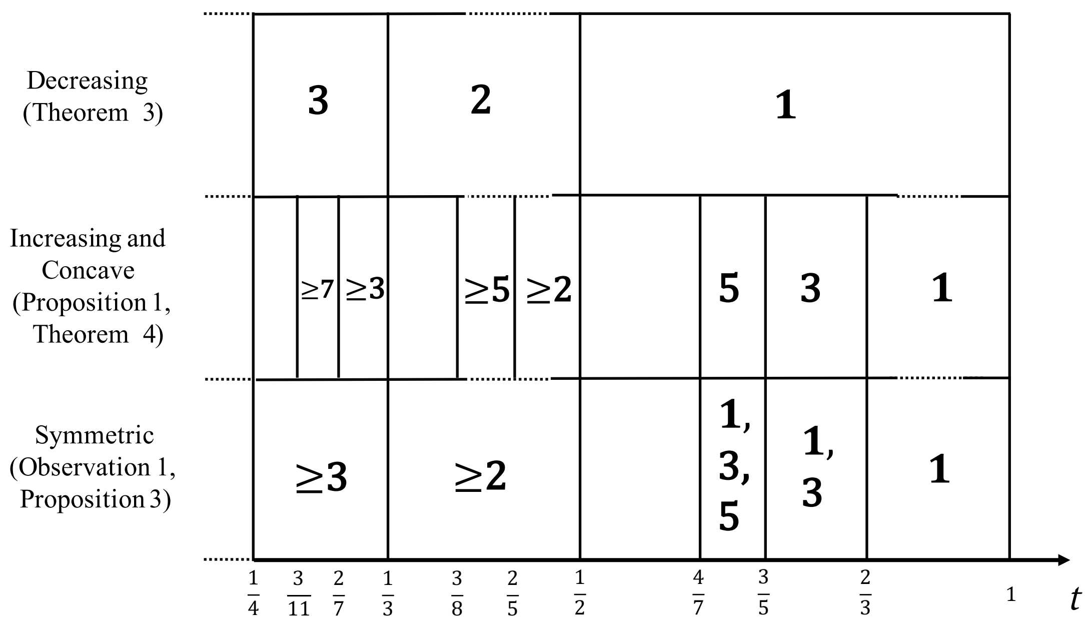

Figure 3: Optimal sample sizes for non-degenerate priors

$P(2n+1, n+1)$ only for n's such that a simple majority of successes triggers the DM to take the action, we obtain that the corresponding sample sizes are the only candidates for optimality.

Finally, full revelation is dominated by these sample sizes. This is because (i) the symmetry of the prior implies that the expected probability of taking the action for full revelation is $\int_t^1 f(q)dq = \int_0^{1-t} f(q)dq = F(1-t)$, and (ii) $F(1-t) < F(1/2)$ because $t > 1/2$ and the prior has a positive mass in the neighborhood of $1/2$. □

Figure 3 summarizes the optimal sample sizes for monotone or symmetric priors. Decreasing priors provide the lower envelope of the optimal sample sizes. Increasing and concave priors provide the upper envelope for $t > 1/2$, and symmetric priors span the range between the lower and upper envelopes for $t > 1/2$.

Figure 3 has two implications. First, it highlights that in many cases the designer does not need to have an exact prior in order to identify the optimal sample size. Rather, knowledge of more general properties such as monotonicity, concavity, or symmetry suffices. Second, for $t > 1/2$, a designer, who does not want the sample size choice to reveal any information about the prior, would choose a sample size satisfying that a simple majority of successes would trigger the DM to take the action. To illustrate the second point more formally, consider a designer who does not know whether — or does not want to reveal by the choice of sample size that — the prior is decreasing, increasing and concave, or symmetric. This designer would choose a sample size as follows.

Corollary 1. *For any $t > 1/2$, the optimal sample size is an odd sample size satisfying that a simple majority of successes would trigger the DM to take the action.*

To verify Corollary 1, let k' be the largest integer such that $t \leq k'/(2k'-1)$. Then, the set of candidates for optimaility according to Corollary 1 is $\mathcal{D} = \{1, 3, \cdots, 2k'-1\}$. Consider some sample size $n \notin \mathcal{D}$. If n is an even integer smaller than $2k'-1$ then it is dominated by the odd sample size just above it, which is in $\mathcal{D}$. And if n is larger than $2k'-1$ or $n = \infty$ then it is dominated by some sample size in $\mathcal{D}$ as the proofs of Theorem 3, Theorem 4, and Proposition 3 establish.

4 An application

This section studies an application to a buyer-seller interaction in which transfers are allowed.

The seller's product is a *platform* with a unit mass of brands. Each brand either matches a given buyer's taste in which case it generates a value of 1 to this buyer, or it does not in which case it generates a value of 0 to the buyer.

There is a unit mass of risk-neutral buyers and each of them has to decide whether to join the platform. Each buyer is characterized by two parameters. The first is the value of the buyer's outside option denoted by $t \in [0, 1]$. The value of the outside option is drawn from a CDF $H(t)$ and is known to the buyer. The second parameter is the taste parameter q, which is the proportion of platform brands that match the buyer's taste. The parameter q is drawn from a CDF $F(q)$. The buyer does not know q and uses an unbiased inference procedure to estimate it. Absent any fees, a buyer joins the platform if the estimate of q is weakly larger than t.

The platform owner is a risk-neutral and profit-maximizing seller. She knows F and H and chooses two objects. The first is how much experimentation n to allow buyers prior to deciding whether to join the platform. The second is the post-experimentation two-part tariff $P = (P_M, P_U)$ where P_M is a *membership fee* paid independently of platform usage and P_U is a *usage fee* paid every time the buyer uses a brand on the platform.

We make the following assumption on the distribution H of buyers' outside options.

Assumption 1. The ratio $H(t)/t$ increases on $(0, 1]$.

Clearly, any convex H satisfies Assumption 1. However, H need not be convex. For example, Any S-shaped CDF such that the line segment connecting the points $(0,0)$ and $(t, H(t))$ lies above the graph of H for any t satisfies the assumption.

By choosing the sample size n, the seller chooses the demand system she faces. For a given q, this demand system is

$$D(n, P_M, P_U \mid q) = \sum_{j=0}^{n} H(j/n(1-P_U) - P_M) b(n, j, q). \tag{4}$$

Indeed, for a given q, a buyer observes the sample (n, j) with probability $b(n, j, q)$ and estimates the expected value of joining the platform to be $j/n - P_M - j/nP_U$. He joins the platform if this value is weakly above t. Because t is distributed according to H, the probability of joining the platform conditional on the sample (n, j) is $H(j/n(1-P_U)-P_M)$ where $H(x) = 0$ for $x \leq 0$. Summing over all possible j's gives the demand system.

The seller thus chooses (n, P_M, P_U) to solve

$$\max \int_0^1 (P_M + qP_U) D(n, P_M, P_U \mid q) dF(q).$$

Proposition 4. *The seller's optimal choice is to provide buyers access to a single experiment, charge no usage fee, and charge a profit-maximizing membership fee against the demand function* $H(1-p)$.

In other words, the "volume" considerations identified in Section 3 are dominated by "margin" considerations. The platform owner prefers to provide as little data as possible to buyers and target those who obtain a a favorable realization, rather than to provide more data to buyers and possibly capture a larger volume by reducing price.

Proof. We first show that the demand system for sample size 1 is strictly larger than the demand system for any larger sample size. To do so, fix q, P_M, and P_U. Let $D(n|P_M, P_U, q)$ be the corresponding demand for sample size n, and consider the function

$$\widehat{H}(x) = \begin{cases} 0 & \textit{if } x(1-P_U) \leq P_M \\ H(x(1-P_U)-P_M) & \textit{otherwise.} \end{cases}$$

By definition, $\widehat{H}(x)/x < \widehat{H}(1)$. When $0 < x(1-P_U) \leq P_M$, this is because $\widehat{H}(x) = 0$. And for larger x's, this is because if $x > x' \geq P_M/(1-P_U)$ then by Assumption 1, the inequality $\widehat{H}(x)/x \geq K(\widehat{H}(x')/x')$ holds where $K = \frac{(x(1-P_U)-P_M)x'}{(x'(1-P_U)-P_M)x} > 1$.

The demand system for sample size n is a convex combination of the values of $\widehat{H}$ at the points $\{0, 1/n, \ldots, 1\}$. Since $\widehat{H}(x) < x\widehat{H}(1)$ as we just showed, we can write:

$$\begin{aligned} D(n|P_M, P_U, q) &< \textstyle\sum_{j=0}^{n}(j/n)\widehat{H}(1)b(n,j,q) \\ &= H(1-P_U-P_M)(\textstyle\sum_{j=0}^{n}(j/n)b(n,j,q)) \\ &= H(1-P_U-P_M)q \\ &= D(1|P_M, P_U, q). \end{aligned}$$

Thus, the seller's optimal sample size choice is 1 and she chooses (P_M, P_U) to maximize

$$\int_0^1 (P_M + qP_U)H(1-P_U-P_M)q dF(q).$$

To obtain the optimal fees, consider a related problem in which the seller chooses prices for a known q. If P_U is positive, the seller can increase profit by reducing P_U by a small δ and increasing P_M by $q\delta$. Indeed, with these new prices, the overall fee $P_M + qP_U$ remains the same but demand increases to $H(1-P_M-P_U+(1-q)\delta)q$. Thus, the seller optimally sets a usage fee of 0 and a membership fee P_M that maximizes $P_M H(1-P_M)\int_0^1 qdF(q)$. □

5 Concluding comments

This paper incorporated statistical inference into persuasion environments. We considered a designer who wishes to persuade a DM to take an action and controls how much data about a payoff-relevant parameter the DM sees prior to making a choice. The DM is an statistician who uses the data and statistical inference to estimate the parameter and decide whether to take the action.

5.1 Bounds on the sample size

The analysis characterized the designer's optimal sample size when the set of feasible sample sizes is the set of all integers. It is possible, however, that the designer may face exogenous constraints when choosing the sample size.

One relevant constraint is an upper bound on the sample size. For example, in an experimental setting or a public opinion survey, the pool of potential participants may be small. Incorporating such an upper bound is relatively straightforward. We demonstrate for the setting of Theorem 3.

Observation 2. *For any decreasing prior and any upper bound $\overline{n}$ on the sample size, sample size $\overline{n}$ is optimal for $t \leq 1/\overline{n}$, and the sample size identified in Theorem 3 is uniquely optimal otherwise.*

Clearly, for $t > 1/\overline{n}$, the optimal sample size identified Theorem 3 is feasible and therefore optimal. For $t \leq 1/\overline{n}$ and sample size $n \leq \overline{n}$, even a single success would trigger the DM to take the action. Thus, the probability of taking the action is $1-(1-q)^n$. This probability increases in n, and the conclusion of Observation 2 for $t \leq 1/\overline{n}$ follows.

Another relevant constraint is a lower bound $\underline{n}$ on the sample size. For example, the Food and Drug Administration requires a minimal number of participants in clinical trials when a pharmaceutical company applies for drug approval. Incorporating such a bound into the analysis is more involved. We demonstrate again for the setting of Theorem 3.

Proposition 5. *For any decreasing prior, $t > 1/2$, and an odd lower bound $\underline{n}$ on the sample size, if simple majority triggers the DM to take the action for sample size $\underline{n}$, then this sample size is optimal. Otherwise, the optimal sample size is weakly smaller than $2\underline{n}+1$.*

Proof. For sample size $\underline{n}$, a simple majority triggers the DM to take the action if and only if $1/2 < t \leq \frac{\underline{n}+1}{2}/\underline{n}$. Following the proof of Theorem 3 where $(n', 1)$ is replaced by $(\underline{n}, (\underline{n}+1)/2)$ establishes the optimality of sample size $\underline{n}$ in this case

For $t > \frac{\underline{n}+1}{2}/\underline{n}$, it suffices to show that any sample size $n \geq 2\underline{n}+1$ is dominated by a sample size, which is "about half of it", in order to establish the upper bound on the optimal sample size. By Step 1 in Theorem 3, it suffice to establish this dominance for the uniform prior.

To establish the required dominance for the uniform prior, fix n and $k = \lceil tn \rceil$. The following table identifies, for every possible combination of even and odd n and k, a sample size n' (on left) and an integer k' (on right) such that

(i) n' is about half of n and $n' \geq \underline{n}$ (i.e., n' is feasible)

(ii) $k'/n' \geq k/n$ (i.e., k' successes trigger the DM to take the action for sample size n'), and

(iii) $P(n', k') = 1 - k'/(n'+1) \geq 1 - k/(n+1) = P(n, k)$:

	k even	k odd
n even	$n/2,\ k/2$	$n/2-1,\ (k+1)/2-1$
n odd $(k/n > 2/3)$	$(n-1)/2-1,\ k/2-1$	$(n+1)/2,\ (k+1)/2$
n odd $(k/n \leq 2/3)$	$(n-1)/2+2,\ k/2+1$	$(n+1)/2,\ (k+1)/2$

.

$\square$

Thus, for the decreasing prior, the optimal sample size cannot be more than about twice the lower bound.

5.2 Bayesian inference

Another assumption that we made is that the DM is an unbiased statistician. An alternative approach would be to consider a Bayesian DM who is endowed with a prior belief on the parameter value and uses the data to Bayes-update this prior. The optimal sample size would then depend on the prior belief because the prior belief, together with the data, determine the estimated value of the parameter. We conclude by extending the analysis to the case of Bayesian inference from a uniform prior and comparing the designer's optimal sample size to the case of an unbiased statistician.

Suppose the designer and the DM have a uniform prior on q. For $t \leq 1/2$, the DM would take the action with probability 1 without any additional data. The designer would therefore provide no data to the DM in this case. For $t > 1/2$, the DM would only take the action if the designer provides additional data to the DM. The following result identifies the designer's optimal sample size in this case.

Theorem 5. *Suppose $t > 1/2$. The smallest sample size n' such that only a unanimity of successes would trigger the DM to take the action is optimal. For $l \geq 2$, if the sample $(l(n'+1)-1, ln')$ triggers the DM to take the action, then sample size $l(n'+1)-1$ is also optimal.*

The left panel in Figure 4 provides a graphical illustration for $t \in (2/3, 3/4]$. The smallest sample size such that only a unanimity of successes would trigger the DM to take the action is 2 because the expected value of q after obtaining the sample $(1,1)$ is $E(1,1) = \int_0^1 q dG_{1,1} = 2/3$ and the corresponding expected value for the sample $(2,2)$ is $E(2,2) = 3/4$.[8] Thus, sample size 2 is optimal in this interval. The next candidate for optimality is sample size 5, and the relevant sample is $(5,4)$. Because $E(5,4) = 5/7$, sample size 5 is optimal for $t \leq 5/7$. Similarly, sample size 8 is optimal for $t \leq 7/10$ and so on.

[8] To see why this and the other calculations below hold, recall that the uniform prior is a $Beta(1,1)$ distribution and that the Beta distribution is a conjugate prior with respect to the Binomial distribution. Therefore, the DM's estimate after obtaining the sample (n,k) is $G_{n,k} = Beta(k+1, n-k+1)$ with mean $(k+1)/(n+2)$.

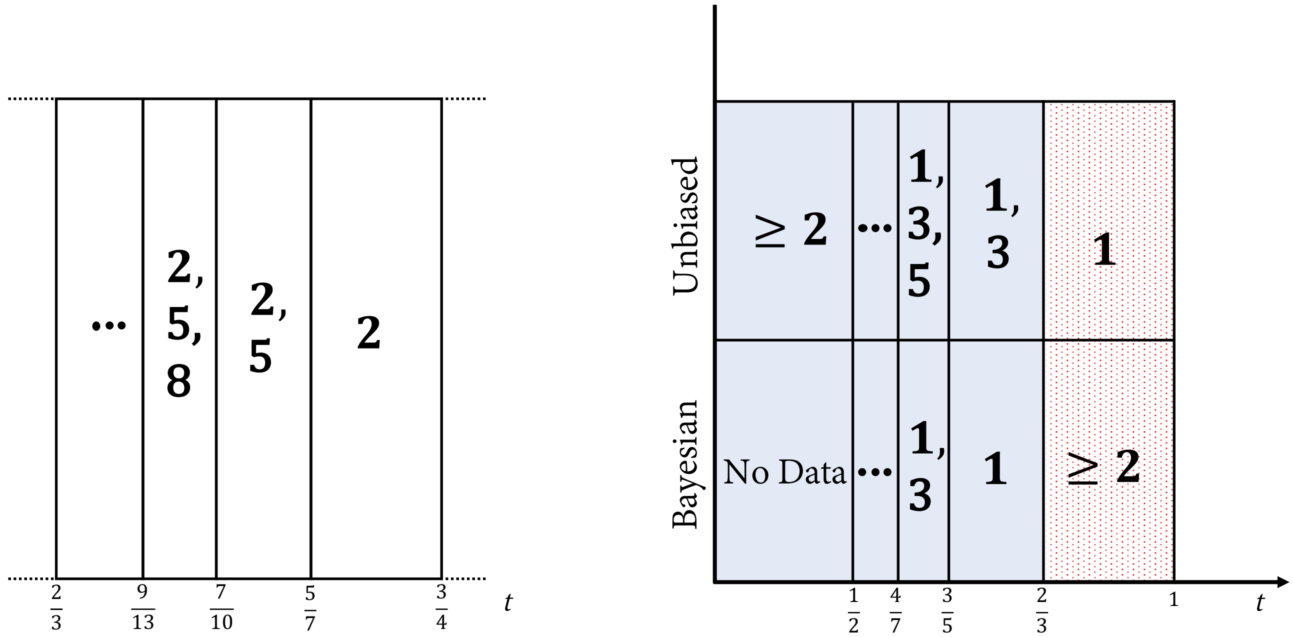

Figure 4: Bayesian inference with a uniform prior. Left panel depicts the optimal sample size between 2/3 and 3/4. Right panel depicts the comparison to unbiased inference.

The right panel in Figure 4 compares the optimal sample sizes for unbiased and Bayesian inference. For $t < 1/2$, the optimal sample size is larger for unbiased inference than for Bayesian inference. This is because the prior triggers the DM to take the action without obtaining any data. For $t > 2/3$, the reverse ranking holds. This is because the designer has to provide a Bayesian DM with additional data (relative to unbiased inference) in order to overcome the effect of the prior. For $1/2 < t \leq 2/3$, the set of optimal sample sizes when facing an unbiased statistician is "larger" than when facing a Bayesian statistician in the sense that the former set (characterized in Proposition 3) includes all sample sizes in the latter set (characterized in Theorem 5) plus one that is strictly larger. The intuition is that for both unbiased and Bayesian inference, any sample size that triggers the DM to take the action when a simple majority of the realizations are successes is optimal. For unbiased inference, sample sizes 1 and 3 would qualify, for example, when $3/5 < t \leq 2/3$. But for Bayesian inference, only sample size 1 would qualify for t in this interval because the prior "pushes" the posterior in the direction of 1/2.

References

Anderson, Theodore, and Stephen Samuels. 1967. "Some inequalities among binomial and Poisson probabilities." The Regents of the University of California.

Conrad, Keith. 2004. "Probability distributions and maximum entropy." *Unpublished Note.*

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. 2020. "Strategic Interpretations." *Working Paper.*

Galperti, Simone. 2019. "Persuasion: The art of changing worldviews." *American Economic Review*, 109(3): 996–1031.

Glazer, Jacob, and Ariel Rubinstein. 2004. "On optimal rules of persuasion." *Econometrica*, 72(6): 1715–1736.

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

Levy, Gilat, Inés Moreno de Barreda, and Ronny Razin. 2018. "Persuasion with Correlation Neglect." *Working Paper.*

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. 2019. "Procedural Rationality in Repeated Games." Working paper.

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.

A Proofs of theorems

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

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

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 minimal integer in $k[n]$. By definition, $k[2m-1]$ is a singleton with $\kappa(2m-1) = m$, and $\kappa(2m) = m+1$.

Figure 5 illustrates construction of the set $k[n]$ for $n = 3, \ldots, 13$ and $m = 2$. For $m = 2$, the relevant range of t 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 stripped dots correspond to $\frac{\kappa(n)}{n}$. For example, $\kappa(8) = 5$ and therefore the corresponding dot is striped.

Fix a sample size $n > 2m-1$ and $k \in k[n]$. Let $\Delta_{n,k}(q) = P(2m-1, m, q) - P(n,k,q)$. We need to show that $\Delta_{n,k}(q) > 0$ for $q \in (0, b(n,k)]$ where $b(n,k) = \min\left\{\frac{k}{n}, \frac{m}{2m-1}\right\}$. Lemma

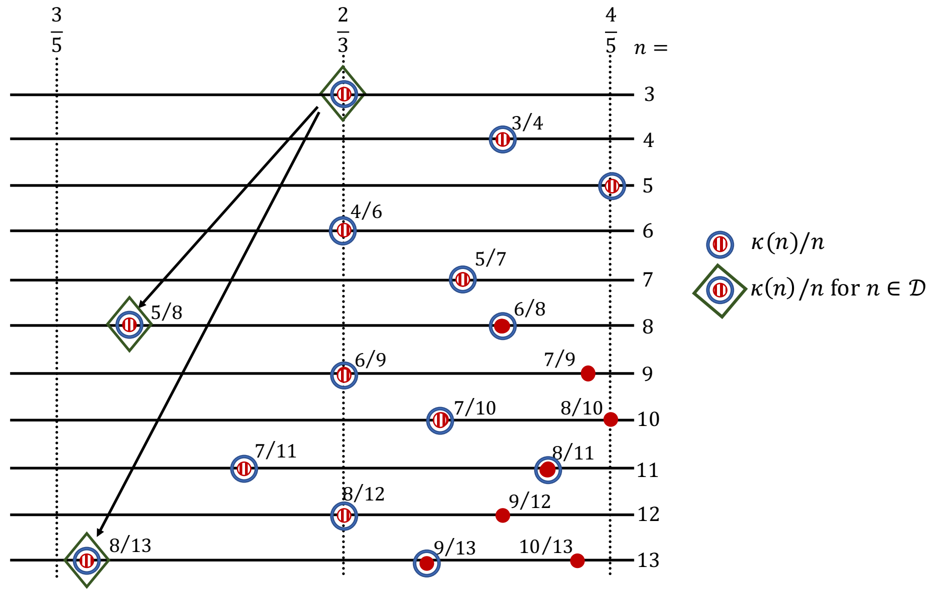

Figure 5: Construction of $k[n]$ and $\mathcal{D}$ in Theorem 2 for $m = 2$

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 $\mathcal{D}$.

1 implies that $\Delta_{n,k}(q) > 0$ in a neighbourhood of $q = 0$ and changes its sign at most once on $(0, 1)$. To complete the proof, it thus remains to show that $\Delta_{n,k}(b(n, k)) > 0$.

Let $\mathcal{D}$ be a set of all sample sizes n satisfying that $\frac{\kappa(n)}{n} < \frac{\kappa(n_1)}{n_1}$ for every n_1 with $2m - 1 \leq n_1 < n$. In Figure 5, sample sizes $3, 8$ and 13 belong to set $\mathcal{D}$ and the corresponding dots are circumscribed by diamonds. We first prove that for any sample size $n \in \mathcal{D}\backslash\{2m - 1\}$, $\Delta_{n,\kappa(n)}(b(n, \kappa(n))) > 0$. This domination is shown by solid arrows in Figure 5. We then prove that $\Delta_{n,k}(b(n, k)) > 0$ for all other pairs (n, k) with $n > 2m - 1$ and $k \in k[n]$.

Fix a sample size $n \in \mathcal{D}\backslash\{2m - 1\}$. Lemma 3 implies that $n = n(l) = (2m + 1)l - 2$ for some $l \geq 2$. By definition, $\kappa(n) = (m + 1)l - 1$. Since $b(n, \kappa(n)) = \kappa(n)/n$ in this case, we need to prove that $\Delta(l) = \Delta_{n(l),\kappa(n(l))} (\kappa(n(l))/n(l)) > 0$ for any $l \geq 2$. Lemma 5 establishes this inequality for $l = 2$, and Lemma 6 for $l \geq 3$.

Fix any of the remaining pairs (n, k) with $n > 2m - 1$ and $k \in k[n]$. By Lemma 4, there exists $n_1 \in \mathcal{D}$ that satisfies $n_1 \leq n$, $\frac{\kappa(n_1)}{n_1} \leq \frac{k}{n}$, and $n_1 - \kappa(n_1) \leq n - k$. The following series of inequalities implies that $\Delta_{n,k}(b(n, k)) > 0$:

$$P\left(n, k, \tfrac{k}{n}\right) < P\left(n_1, \kappa(n_1), \tfrac{\kappa(n_1)}{n_1}\right) < P\left(2m - 1, m, \tfrac{\kappa(n_1)}{n_1}\right) < P\left(2m - 1, m, \tfrac{k}{n}\right).$$

The left inequality holds by the Lemma below, the middle one was proved in the previous paragraph, and the right one follows from the monotonicity of $P(\cdot,\cdot,q)$ in q.

Lemma. *Suppose* $n_1 < n$, $n_1 - k_1 \leq n - k$ *and* $\frac{k_1}{n_1} \leq \frac{k}{n}$. *Then* $P(n,k,\frac{k_1}{n_1}) < P(n_1,k_1,\frac{k_1}{n_1})$ *and* $P(n,k,\frac{k}{n}) < P(n_1,k_1,\frac{k}{n})$.

Proof. It suffices to prove that $P(n,k,\frac{k}{n}) < P(n_1,k_1,\frac{k_1}{n_1})$ because $P(\cdot,\cdot,q)$ increases in q. This inequality follows from:

$$P\left(n,k,\tfrac{k}{n}\right) \leq P\left(n_1+k-k_1,k,\tfrac{k}{n_1+k-k_1}\right) < P\left(n_1,k_1,\tfrac{k_1}{n_1}\right).$$

The left inequality is obtained by applying $(n-n_1)-(k-k_1)$ times the inequality $P\left(n,k,\frac{k}{n}\right) \leq P\left(n-1,k,\frac{k}{n-1}\right)$ (Anderson and Samuels (1967), Theorem 2.3). The right inequality is obtained by applying $k-k_1 \geq 0$ times the inequality $P\left(n,k,\frac{k}{n}\right) < P\left(n-1,k-1,\frac{k-1}{n-1}\right)$ (Anderson and Samuels (1967), Theorem 2.2). $\square$

Proof of Theorem 4. Fix an integer $m \geq 1$ and let $t \in \left(\frac{m+1}{2m+1},\frac{m}{2m-1}\right]$. The largest sample size such that a simple majority of successes triggers the DM to take the action is $2m-1$.

We first consider finite sample sizes. By Proposition 1, it suffices to show that sample size $2m-1$ dominates larger sample sizes. So fix a sample size $n > 2m-1$ and $k \in k[n]$, where $k[n]$ is defined as in Equation (1). The following two steps establish that $P(2m-1,m) > P(n,k)$.

The first step establishes that it suffices to examine the expected 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. By Lemma 1, $P(2m-1,m,q) > P(n,k,q)$ for $q \in (0,q^*)$ and a reverse inequality holds for $q \in (q^*,1)$ if $q^* < 1$. Consider the linear function $f_h(q) = \frac{f(q^*)}{q^*}q$ 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 prior), then $P(2m-1,m) > P(n,k)$ for f. Because $f_h(q) = \frac{f(q^*)}{2q^*}h(q)$, it suffices to prove the inequality for h. $\square$

The second step examines the expected probabilities with respect to h.

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

Proof. For any n and 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 $P'(n,k,q) = k\binom{n}{k}q^{(k-1)}(1-q)^{(n-k)}$. Thus, $P(2m-1,m) = 1 - \frac{(m+1)}{2(2m+1)}$.

Thus, it suffices 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 $\frac{k+1}{n+1} \geq \frac{k}{n}$ and $t > \frac{m+1}{2m+1}$.

To prove (ii), it suffice to consider whether (ii) holds for $\kappa(n)$, and to do that, we define the set $\mathcal{D}$ as in Theorem 2. If $n \in \mathcal{D}\backslash\{2m-1\}$, Lemma 3 implies that $n = (2m+1)l - 2$ for some $l \geq 2$. By definition, $\kappa(n) = (m+1)l - 1$. Thus, $\kappa(n)/(n+2) \geq 1/2$. If $n \notin \mathcal{D}$, Lemma 4 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 $\kappa(n)/(n+2) > \kappa(n_1)/(n_1+2) \geq 1/2$. □

As for the full revelation, by Step 1, it suffices to show that sample size $2m-1$ dominates full revelation for the linear prior. This holds because the expected value of the full revelation is $1 - t^2$ and $t > \frac{m+1}{2m+1}$. □

Proof of Proposition 2. Let $P_F = \int_0^1 \mathbb{1}_{\{q \geq t\}} f(q)dq$ denote the expected probability of taking the action for full revelation. We want to show that there exists $t' > 0$ such that for any $t \leq t'$ the expected probability $P(n, \lceil tn \rceil)$ is smaller than P_F for any integer n.

We do so in two steps. The first establishes that it suffices to examine the expected probabilities with respect to the linear prior $h(q) = 2q$.

Step 1. If $P_F > P(n,k)$ for the linear prior, then $P_F > P(n,k)$ for a convex prior with $f(0) = 0$.

Proof. The inequality $P(n,k,q) > \mathbb{1}_{\{q \geq t\}}$ holds for $q < t$ and a reverse inequality holds for $q > t$. An analogous argument to Step 1 in Theorem 4 now implies the result. □

The second step establishes that full revelation is optimal with respect to the linear prior when $t \leq 2/7$.

Step 2. For the linear prior and $t \leq 2/7$, $P_F > P(n,k)$.

Proof. Let $k[n] = \{\lceil nt \rceil \mid 0 < t \leq 2/7\}$ 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 maximal integer in $k[n]$. We wish to show that $P(n,k) = 1 - \frac{k(k+1)}{(n+1)(n+2)}$ is smaller than $P_F = 1 - t^2$ for $k \in k[n]$.

If $k = \kappa(n)$, the combinatorial identity (2) implies that it suffices to consider the largest sample size $n(k)$ such that $k/n(k) \geq 2/7$. This sample size is $\lfloor \frac{7k}{2} \rfloor$. Because the maximal t for which $\kappa(n)$ successes trigger the DM to take the action is given by $t = 2/7$, full revelation is optimal if $\frac{k(k+1)}{(n(k)+1)(n(k)+2)} > (2/7)^2$. This inequality holds for $k \geq 1$.

If $k \neq \kappa(n)$ then we must have $n \geq 4$ because for smaller sample sizes, $k[n] = \{\kappa(n)\}$ is a singleton. By definition, $(\kappa(n)-1)/n < 2/7$ and therefore $k/n < 2/7$. The maximal t for which k successes trigger the DM to take the action is given by $t = k/n$. Therefore, we need to prove that $1-(k/n)^2 > P(n,k)$ which holds if $k/n < n/(3n+2)$. The sequence $\{n/(3n+2)\}$ increases in n. Because $n/(3n+2) = 2/7$ for $n = 4$, the result follows. □

□

Proof of Theorem 5. Fix $t > 1/2$ and let $n' = \lceil (1-t)^{-1} \rceil - 2$. For any sample size n, let $k = \lceil t(n+2) \rceil - 1$ be the minimal number of successes that trigger the DM to take the action after obtaining the sample $(n,\, k)$. Since $n'/(n'+1) < t \leq (n'+1)/(n'+2)$, the integer n' is the smallest sample size such that only a unanimity of successes triggers the DM to take the action.

As argued in Step 2 in the proof of Theorem 3, the expected probability $P(n,k)$ for the uniform prior equals $1 - \frac{k}{n+1}$. For sample size n', the value of the objective function is $P(n',n') = 1 - n'/(n'+1)$. It is larger than the corresponding value $1-t$ for full revelation because $t > n'/(n'+1)$. All sample sizes $n < n'$ are dominated because the DM will not take the action after obtaining the sample (n,n).

For larger sample sizes, fix $l \geq 2$. The DM takes the action after obtaining the sample $(n, n-l+1)$ if $\frac{n'}{n'+1} < t \leq \frac{n-l+2}{n+2}$, i.e. if $n > l(n'+1)-2$. Because n is an integer, the smallest n that satisfies this condition is $n = l(n'+1)-1$. For this n, the value of the objective function is

$$1 - k/(n+1) = 1 - n'/(n'+1)$$

which is equal to the value of the objective function for sample size n'. For larger n's, this value is smaller than for sample size n' because $(n-l+1)/(n+1)$ is larger. Thus, sample size n' dominates all sample sizes with the exception of those given by $n = l(n'+1)-1$, which are also optimal for $t \leq (n-l+2)/(n+2)$, as in the statement of the Theorem. □

B Mathematical appendix

Lemma 1. *Let $n' \leq n$, $k' \leq k$, and $n' + k' < n + k$. Then, there exists $q^* > 0$ such that $P(n', k', q) > P(n, k, q)$ for $q \in (0, q^*)$. If $q^* < 1$ then $P(n, k, q) > P(n', k', q)$ for $q \in (q^*, 1)$.*

Proof. By definition, $P(n', k', 0) = P(n, k, 0) = 0$. The inequality $P'(n', k', q) > P'(n, k, q)$ holds for some ϵ and $q \in (0, \epsilon)$ if and only if

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

The function $r(q)$ is continuous in q and approaches 0 from above as q approaches 0 from above. The desired inequality follows.

By definition, $P(n', k', 1) = P(n, k, 1) = 1$. Thus, to conclude the proof, it suffices to show that there is at most one interior point of intersection of the two functions. Assume to the contrary that there are two interior points q_1 and $q_2 > q_1$ that satisfy $P(n', k', q_i) = P(n, k, q_i)$.[9] Because $P(n, k, q)$ is below $P(n', k', q)$ in a neighborhood of 0 (see above), we have that $P'(n, k, q_1) > P'(n', k', q_1)$ implying that $r(q_1) > 1$. Analogously, $P'(n, k, q_2) < P'(n', k', q_2)$ implying that $r(q_2) < 1$, and $P'(n, k, 1) > P'(n', k', 1)$ (because $P(n', k', 1) = P(n, k, 1) = 1$) implying that $r(1) > 1$. Thus, $r(q)$ has at least two interior extremums in contradiction to the fact that by definition $r(q)$ has at most one interior extremum. □

Lemma 2. *For $k \geq 2$, $P(n, 1, \frac{1}{n+1}) \geq P(k(n+1) - 1, k, \frac{1}{n+1})$.*

Proof. Let $q = \frac{1}{n+1}$. We prove that $\overline{P}(k(n+1) - 1, k - 1) \geq \overline{P}(n, 0)$ where $\overline{P}(n, l) = 1 - P(n, l + 1, q)$. Suppose $k = 2$. Then,

$$\overline{P}(2(n+1)-1,1) - \overline{P}(n,0) = (1-q)^n \left(\frac{3n+1}{n+1}(1-q)^n - 1\right)$$

is non-negative if and only if $\frac{3n+1}{n+1} \geq \left(1 + \frac{1}{n}\right)^n$. This inequality can be verified numerically for $n \leq 6$. And for $n \geq 7$, we observe that $\frac{3n+1}{n+1} > e$ together with the combinatorial inequality $e \geq \left(1 + \frac{1}{n}\right)^n$ imply the desired inequality.

Suppose $k \geq 3$. By adding and subtracting terms of the form $\overline{P}((n+1)i + n, i)$ for $i \in \{1, 2, \cdots, k-1\}$, we obtain:

$$\begin{aligned}\overline{P}(k(n+1)-1,k-1) - \overline{P}(n,0) &= \textstyle\sum_{i=1}^{k-1} \overline{P}((n+1)i+n,i) - \overline{P}((n+1)(i-1)+n,i-1) \\ &= \textstyle\sum_{i=1}^{k-1} \Delta_i.\end{aligned}$$

[9] The proof for three or more interior points of intersection is analogous.

We can further expand each Δ_i by adding and subtracting n expressions of the form $\overline{P}((n+1)i+n-j-1,i)$ as follows:

$$\begin{aligned}
\Delta_i &= \textstyle\sum_{j=0}^{n} \overbrace{\left(\overline{P}((n+1)i+n-j,i) - \overline{P}((n+1)i+n-j-1,i)\right)}^{\delta_j} \\
&\quad + \underbrace{\overline{P}((n+1)i-1,i) - \overline{P}((n+1)(i-1)+n,i-1)}_{\alpha} \\
&= \sum_{j=0}^{n} \delta_j + \alpha \qquad (2)
\end{aligned}$$

To complete the proof, it suffices to show that $\frac{\alpha}{n+1} \geq -\delta_j$ for every j. Standard combinatorial identities imply that $-\delta_j = q\binom{(n+1)i+n-j-1}{i} q^i (1-q)^{(ni+n-j-1)}$, and $\alpha = \binom{(n+1)i-1}{i} q^i (1-q)^{(ni-1)}$ by definition. We thus need to show that

$$\binom{(n+1)i-1}{i} \geq \binom{(n+1)i+n-j-1}{i} \left(\frac{n}{1+n}\right)^{(n-j)},$$

which holds because the LHS is a constant whereas the RHS increases in j and is equal to the LHS for $j = n$. □

Lemma 3. *Fix an integer $m \geq 1$. If $n \in \mathcal{D}$ then $n = n(l) = (2m+1)l - 2$ for some $l \geq 1$.*

Proof. Fix $l \geq 1$ 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 $0 \leq j \leq m$ 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 4. *Fix an integer $m \geq 1$. For any $n > 2m-1$ and $k \in k[n]$ there exists $n_1 \in \mathcal{D}$ such that (i) $n_1 \leq n$, (ii) $\frac{\kappa(n_1)}{n_1} \leq \frac{k}{n}$, and (iii) $n_1 - \kappa(n_1) \leq n - k$.*

Proof. Fix $n > 2m-1$ and $k \in k[n]$. Let $\underline{n} \leq n$ be a weakly smaller sample size with $n-k+1$ terms in $P(\underline{n}, \kappa(\underline{n}), q)$. Such a sample size exists because the number of terms in $P(n, \kappa(n), q)$ is greater or equal to $n-k+1$, and the number of terms in $P(n', \kappa(n'), q)$ is either the same or larger by 1 than the number of terms in $P(n'-1, \kappa(n'-1), q)$. By the proof of Lemma 3, the maximal $n \in \mathcal{D}$ that is smaller than $\underline{n}$ satisfies the desired properties. □

Lemma 5. *For any integer $m \geq 1$, $\Delta(2) > 0$.*

Proof. Fix integer $m \geq 1$. Then $\Delta(2)$ is equal to:

$$P\left(2m-1, m, \frac{2m+1}{4m}\right) - P\left(4m, 2m+1, \frac{2m+1}{4m}\right).$$

We first observe that

$$\begin{aligned} P(2m-1, m, \tfrac{2m+1}{4m}) &= \int_0^{\frac{2m+1}{4m}} P'(2m-1, m, q)dq \\ &= \tfrac{1}{2} + \int_{\frac{1}{2}}^{\frac{2m+1}{4m}} m\tbinom{2m-1}{m} q^{m-1}(1-q)^{m-1}dq \\ &\geq_{(1)} \tfrac{1}{2} + \tfrac{1}{4}\tbinom{2m-1}{m}\left(\tfrac{2m+1}{4m}\right)^{m-1}\left(\tfrac{2m-1}{4m}\right)^{m-1} \\ &=_{(2)} \tfrac{1}{2} + \tfrac{1}{2^{4m-1}}\tbinom{2m}{m}\left(4 - \tfrac{1}{m^2}\right)^{m-1} \end{aligned}$$

where inequality (1) holds because $P'(2m-1, m, q)$ decreases in q over, and we use the identity $\binom{2m-1}{m} = \frac{1}{2}\binom{2m}{m}$ to obtain equality (2).

We also observe that

$$P(4m, 2m+1, \frac{2m+1}{4m}) <_{(1)} P(4m-2, 2m-1, \frac{1}{2}) = \frac{1}{2} + \frac{1}{2^{4m-1}}\binom{4m-2}{2m-1}$$

where inequality (1) is obtained by applying twice the inequality $P\left(n, k, \frac{k}{n}\right) < P\left(n-1, k-1, \frac{k-1}{n-1}\right)$ (Anderson and Samuels (1967), Theorem 2.2).

To prove the result, it thus suffices to show that

$$\tbinom{2m}{m}\left(4 - \tfrac{1}{m^2}\right)^{m-1} \geq \tbinom{4m-2}{2m-1}.$$

For $m = 1$, the inequality holds with equality. For $m \geq 2$, we use the identity $\binom{2(n+1)}{n+1} = \frac{4n+2}{n+1}\binom{2n}{n}$ to rewrite the RHS as

$$\binom{4m-2}{2m-1} = \binom{2m}{m}\left(4 - \frac{2}{2m-1}\right) \cdot \left(4 - \frac{2}{2m-2}\right) \cdot \ldots \cdot \left(4 - \frac{2}{m+1}\right)$$

and observe that $(4 - \frac{1}{m^2})$ is larger than any of the $m-1$ terms on the RHS of the last identity. $\square$

Lemma 6. *For any integer $m \geq 1$, $\Delta(l) > 0$ for $l \geq 3$.*

Proof. Fix integers $m, l \geq 1$. Then $\Delta(l)$ is equal to:

$$P\left(2m-1, m, \tfrac{(m+1)l-1}{(2m+1)l-2}\right) - P\left((2m+1)l-2, (m+1)l-1, \tfrac{(m+1)l-1}{(2m+1)l-2}\right).$$

We observe that

$$\begin{aligned}
P\left((2m+1)l-2,(m+1)l-1,\tfrac{(m+1)l-1}{(2m+1)l-2}\right) &<_{(1)} P\left(2ml-2,ml-1,\tfrac{ml-1}{2ml-2}\right)\\
&= \tfrac{1}{2}+\tfrac{1}{2^{2(ml-1)+1}}\tbinom{2ml-2}{ml-1}\\
&\leq_{(2)} \tfrac{1}{2}+\tfrac{1}{2}\tfrac{1}{\sqrt{\pi(ml-1+\frac{1}{4})}}
\end{aligned}$$

where inequality (1) is obtained by applying the identity $P\left(n,k,\frac{k}{n}\right)<P\left(n-1,k-1,\frac{k-1}{n-1}\right)$ for l times, and inequality (2) holds because $\binom{2n}{n}\leq\frac{4^n}{\sqrt{\pi\left(n+\frac{1}{4}\right)}}$. Thus,

$$\Delta(l)>P\left(2m-1,m,\tfrac{(m+1)l-1}{(2m+1)l-2}\right)-\tfrac{1}{2}-\tfrac{1}{2}\tfrac{1}{\sqrt{\pi(ml-1+\frac{1}{4})}}\equiv d(l).$$

To complete the proof we treat l as a continuous variable on $[3,\infty)$ and show that $d'(l)\geq 0$ and $d(3)>0$.

The derivative $d'(l)$ is positive if and only if

$$\frac{((2m+1)l-2)^2}{(4ml-3)^{\frac{3}{2}}\sqrt{\pi}}\geq\frac{\binom{2m-1}{m}}{2}\left(\frac{((m+1)l-1)(ml-1)}{((2m+1)l-2)^2}\right)^{m-1}. \tag{3}$$

The RHS is bounded above by $\frac{2}{\sqrt{\pi(4m+1)}}$. This is because $\binom{2m-1}{m}=\frac{1}{2}\binom{2m}{m}$, $\binom{2m}{m}\leq\frac{4^m}{\sqrt{\pi\left(m+\frac{1}{4}\right)}}$ and $\frac{ab}{(a+b)^2}\leq\frac{1}{4}$ where $a=(m+1)l-1$ and $b=ml-1$. The LHS of inequality (3) increases in l. Thus, to verify that $d'(l)$ is positive, it suffices to verify that $d'(3)$ is positive, which holds because the inequality $\frac{((2m+1)3-2)^2}{(12m-3)^{\frac{3}{2}}}\geq\frac{2}{\sqrt{4m+1}}$ holds for $m\geq 1$.

It remains to show that $d(3)>0$ for $m\geq 1$. For $m=1$ we directly verify the inequality. For $m\geq 2$, we start by obtaining a lower bound on

$$\begin{aligned}
P(2m-1,m,\tfrac{3m+2}{6m+1}) &= \int_0^{\frac{3m+2}{6m+1}}P'(2m-1,m,q)dq\\
&= \tfrac{1}{2}+\int_{\frac{1}{2}}^{\frac{3m+2}{6m+1}}m\tbinom{2m-1}{m}q^{m-1}(1-q)^{m-1}dq\\
&\geq_{(1)} \tfrac{1}{2}+\tfrac{3m}{2(6m+1)}\tbinom{2m-1}{m}\left(\tfrac{3m+2}{6m+1}\right)^{m-1}\left(\tfrac{3m-1}{6m+1}\right)^{m-1}\\
&\geq_{(2)} \tfrac{1}{2}+\sqrt{\tfrac{3}{\pi(3m+1)}}\tfrac{3m}{(6m+1)}\left(\tfrac{(6m+4)(6m-2)}{(6m+1)^2}\right)^{m-1}.
\end{aligned}$$

Inequality (1) holds because $P'(2m-1,m,q)$ decreases on $\left(\frac{1}{2},\frac{3m+2}{6m+1}\right)$. We use the identity $\binom{2m-1}{m}=\frac{1}{2}\binom{2m}{m}$ and the inequality $\binom{2m}{m}\geq\frac{4^m}{\sqrt{\pi\left(m+\frac{1}{3}\right)}}$ to obtain (2).

Thus, it suffices to prove that

$$\left(\frac{(6m+4)(6m-2)}{(6m+1)^2}\right)^{m-1}>\frac{6m+1}{3m}\sqrt{\frac{(3m+1)}{3(8m-3)}}.$$

Because the LHS of the above inequality attains its minimum at $m=2$ and the RHS decreases in $m\geq 2$, we complete the proof by verifying it for $m=2$. □