Rationalizing Pre-Analysis Plans: Statistical Decisions Subject to Implementability

Maximilian Kasy Jann Spiess

June 21, 2022

Abstract

Pre-analysis plans (PAPs) are a potential remedy to the publication of spurious findings in empirical research, but they have been criticized for their costs and for preventing valid discoveries. In this article, we analyze the costs and benefits of pre-analysis plans by casting pre-commitment in empirical research as a mechanism-design problem. In our model, a decision-maker commits to a decision rule. Then an analyst chooses a PAP, observes data, and reports selected statistics to the decision-maker, who applies the decision rule. With conflicts of interest and private information, not all decision rules are implementable. We provide characterizations of implementable decision rules, where PAPs are optimal when there are many analyst degrees of freedom and high communication costs. These PAPs improve welfare by enlarging the space of implementable decision functions. This stands in contrast to single-agent statistical decision theory, where commitment devices are unnecessary if preferences are consistent across time.

Keywords: Pre-analysis plans, Statistical decisions, Implementability
JEL codes: C18, D8, I23

We thank Alex Frankel, Carlos Gonzalez Perez, Pascal Michaillat, Ted Miguel, Marco Ottaviani, Alex Teytelboym, and participants at the BITSS 2022 meeting, the 2022 AEA meetings, and the 2022 conference in Honor of Jim Powell for helpful discussions and suggestions.

1 Introduction

Background and motivation When writing their papers, researchers might cherry-pick the findings that they report. Cherry-picking can distort the inferences that we can draw from published findings, and has led some to argue that "most published findings are false" (Ioannidis, 2005). As a potential remedy, PAPs have become a precondition for the publication of experimental research in economics, for both field experiments and lab experiments.[1]

Pre-analysis plans in their ideal form enable valid inference by specifying a full mapping from the data to the set of statistics that are reported. By tying the analyst's hands, PAPs prevent the cherry-picking of results, and might provide a remedy for the distortions introduced by unacknowledged multiple hypothesis testing. This is the justification of PAPs that is most commonly invoked. As Gelman (2017) argues, pre-analysis plans play the same role for frequentist notions of bias and size control as randomized controlled trials play for causality – they are necessary for the very definition of these notions: The notion of the size of a statistical test depends on knowing the test decision for all counterfactual realizations of the data, not just for the observed realization. The same is true for the notion of the bias of an estimator. An ideal PAP specifies these counterfactual decisions and estimates.

The widespread adoption of pre-analysis plans has not gone uncontested, however, as evidenced by Coffman and Niederle (2015), Olken (2015), and Duflo et al. (2020), who discuss the costs and benefits of pre-analysis plans in experimental economics from a practitioners' perspective. PAPs have been criticized for putting a disproportionate burden on researchers and limiting their ability to learn interesting hypotheses from the data. For example, Duflo et al. (2020) argue that "an ex-post requirement of strict adherence to pre-specified plans, or the discounting of non-pre-specified work, may mean that some experiments do not take place, or that interesting observations and new theories are not explored and reported," and in their "call for moderation" suggest that exploratory analysis should be published alongside results based on a PAP. Drawing on a survey of researchers, Miguel (2021), on the other hand, argues that many researchers also appreciate the benefits of PAPs in protecting them from pressures by funders and implementation partners, and in enforcing greater clarity of research designs.

In this article, we aim to clarify the costs and benefits of pre-analysis plans by modeling statistical inference as a mechanism-design problem. To motivate this approach, note that in single-agent statistical decision theory rational decision-makers, if their preferences are consistent over time, have no need for the commitment device that is provided by a PAP. This holds in particular when a single decision-maker aims to construct tests that control size, or estimators that are unbiased; they have

[1]Just as in the case of randomized experiments, the adoption of PAPs in economics follows their prior adoption in clinical research; see for instance the guidelines of the FDA on PAPs, (Food and Drug Administration, 1998).

no reason to "cheat themselves." The situation is different, however, when there are multiple agents with conflicting interests, or when preferences are not consistent across time.

Setup In our model, we consider the interaction between a decision-maker and an analyst with private information and conflicting interests. One example of such a conflict of interest is between a researcher (analyst) who wants to reject a hypothesis, and a reader of their research (decision-maker) who has an interest in a valid statistical test of that same hypothesis; the relevant decision here is whether to reject the null hypothesis. Another example is the conflict of interest between a researcher (analyst) who wants to get published, and a journal (decision-maker) that only wants to publish studies on effects that are large enough to be interesting; the relevant decision here is whether to publish a study. A third example is the conflict of interest between a pharmaceutical company (analyst) who wants to sell drugs, and a medical regulatory agency (decision-maker) who wants to protect patient health; the relevant decision here is whether to approve a drug.

The mechanism-design approach that we propose takes the perspective of a decision-maker who wants to implement a statistical decision rule. Not all rules are implementable, however, when the analyst has divergent interests and private information. This mechanism-design perspective allows us to stay close to standard statistical theory, while taking into account the constraints that come from the social dimension of research. The analyst in our model observes a vector of binary data X_i, such as the outcomes of different hypothesis tests, with a distribution governed by the parameter θ. The analyst selectively reports the data X_i to the decision-maker. The decision-maker then makes a binary decision. While the analyst always wants a positive decision ("acceptance"), the decision-maker would like to only accept when the parameter of interest θ exceeds a given threshold. Additionally, the analyst incurs a communication cost that is increasing in the number of reported components i; this communication cost creates private information. In this baseline model with binary data, we characterize the set of implementable statistical decision rules, and analyze the problem of finding an optimal implementable decision rule.

Main results We first restrict attention to the set of symmetric decision rules. For such rules acceptance only depends on the number of successes ($X_i = 1$) and failures ($X_i = 0$). When communication costs and/or researcher degrees of freedom (number of components i available to the analyst) are low enough, the first best rule is directly implementable. This rule applies an optimal cutoff on the number of successes – if a sufficient number of components X_i equals 1, the decision-maker accepts. When communication costs and/or researcher degrees of freedom are very high, on the other hand, no symmetric rule without a PAP delivers positive utility to both analyst and decision-maker. For intermediate parameter values, a second-best rule is implementable without a PAP. This rule applies a cutoff on the number of successes, where the cutoff is distorted downwards relative to the first best. Intuitively, the decision-maker tolerates a certain degree of distorted inference in order to ensure analyst participation.

A PAP allows the decision-maker to break the symmetry between components. By ignoring all components i outside the PAP, she can maximize her expected utility by effectively reducing the number of components available to the analyst (the researcher degrees of freedom), down to the optimal value. At this optimal value, the maximum possible amount of information is communicated by the analyst, without the need to distort the cutoff applied in a way that reduces decision-maker utility.

After our analysis of the symmetric case, we characterize the full set of implementable decision rules in our model. We show that these rules have acceptance regions which are given by the union of acceptance regions for "simple" rules. Simple rules accept only outcome vectors which take a given set of values on a given subset of their components. In general, searching for an optimal decision rule by enumeration of all implementable rules (that is, of all unions of simple rules) is computationally infeasible, since the space of implementable decision rules is too large. We therefore consider a greedy algorithm to optimize over implementable decision rules. The greedy algorithm does not always attain the global optimum, but it achieves welfare which outperforms the symmetrical rules discussed above.

Implications In our model, there is a role for PAPs under some conditions. In particular, if the analyst has many choices (degrees of freedom) for her analysis, and if communication costs are high (there is a lot of private information), then PAPs can improve the welfare (statistical risk) of the decision-maker. If, on the other hand, the analyst faces a smaller number of choices and private information is limited, the decision-maker might be better off without requiring a PAP.

Our model clarifies a motive for pre-analysis plans beyond the goal of achieving valid inference. The pre-analysis plan overcomes a conflict of interest between decision-maker and analyst. If there was no conflict of interest, then a pre-analysis plan would not be necessary to ensure valid inference. If there is a conflict of interest, then the decision-maker will still be able to interpret the analyst's choices correctly in equilibrium, but the lack of a PAP will lead to inefficient outcomes, as the decision-maker is forced to (correctly) make conservative assumptions about analyst behavior.

Our interpretation of statistical inference as a mechanism-design problem connects the structure of pre-analysis plans to the set of implementable decision rules. In this interpretation, the availability of pre-analysis plans enlarges the space of implementable rules. The choice of an optimal pre-analysis plan then boils down to the optimization problem of finding an optimal decision rule among the implementable ones. We discuss fully optimal as well as computationally feasible approximately optimal choices.

Note also that in our model, commitment is somewhat subtle: There is no analyst commitment as such. However the decision-maker might commit to a response function, which might condition on a message (PAP) sent by the analyst before observing the data. This allows the decision-maker to *create* a commitment device for the analyst; whether or not doing so is optimal is a central question of our analysis.

Model variations In Sections 4.1 through 4.6 we consider a number of variations and extensions of our baseline model. These extension cover statistical hypothesis testing, the estimation and selection of multiple parameters, settings without commitment by the decision-maker, and settings with additional private or ex-ante unknown information, as well as different cost structures. The set of implementable decision rules takes a similar form in all of those cases. Let us sketch some of these extensions.

The extension to hypothesis testing in Section 4.1 most closely relates to common interpretations of pre-analysis plans. Here, the binary data in our model can be thought of as the outcomes of individual tests. The analyst's decision is then which of these outcomes to report, while the decision-maker decides when to reject the joint hypothesis. The decision-maker aims to maximize power subject to size control, while the analyst always wants a rejection of the null. In this setting, no non-trivial testing rule that controls size exists when communication costs and/or researcher degrees of freedom are very high. A PAP allows to reduce the researcher degrees of freedom down to the optimal point, with maximal power subject to size control.

In another extension in Section 4.4, we present a case where the analyst has additional private information about which components of the data are actually informative. When the decision-maker lacks this information, it becomes important that the analyst formulates the pre-analysis plan, instead of the decision-maker. This provides a justification for the common practice where an analyst files a pre-analysis plans, while a decision-maker checks the analyst's adherence to the plan.

In Section 4.5, we consider what happens when the number of components available to the analyst is unknown to the decision-maker. We show that such uncertainty over researcher degrees of freedom can have implications which are similar to the implications of communication costs: PAPs are optimal when the uncertainty is large enough.

We furthermore discuss settings with multiple parameters θ_i in Section 4.2, settings without decision-maker commitment in Section 4.3, and settings where the decision-maker rather than the analyst bears the cost of communication in Section 4.6.

Related literature Our article speaks, first, to the current debates around pre-registration – and other possible reforms – in empirical economics and other social- and life-sciences; cf. Coffman and Niederle (2015), Olken (2015), Duflo et al. (2020), and Miguel (2021).

Our article also contributes to a literature that spans statistics, econometrics and economic theory, and which models statistical inference in multi-agent settings. Drawing on classic references (Tullock, 1959; Sterling, 1959; Leamer, 1974), Glaeser (2006) considers the role of incentives in empirical research. A recent fast-growing strand of this literature explicitly models estimation and testing within multiple-agent models with conflicts of interest and private information. This includes Chassang et al. (2012); Tetenov (2016); Di Tillio et al. (2021, 2017); Spiess (2018); Henry and Ottaviani (2019); McCloskey and Michaillat (2020); Libgober (2020); Yoder (2020); Williams (2021); Abrams et al. (2021); Viviano et al. (2021). In this literature, Banerjee et al. (2020); Frankel and Kasy (2022); Andrews and Shapiro (2021); Gao (2022) consider the communication of scientific results to an audience with priors, information, or objectives that might differ from the sender's. Gao (2022), in particular, considers a sender who aims to convince a receiver of a good state in order to obtain a higher action by selectively disclosing data. Relative to these contributions, we focus on the role of implementability as a constraint on statistical decision theory that rationalizes pre-analysis plans.

Our work also relates to the literature on persuasion in economic theory. Glazer and Rubinstein (2004) consider a model similar to ours, but restrict verifiability to only one component and do not allow for communication before data are observed. The literature on Bayesian persuasion, as initiated by Kamenica and Gentzkow (2011) and reviewed in Kamenica (2019), similarly considers a sender with information unavailable to a receiver, where sender and receiver have divergent objectives. In contrast to our model, Bayesian persuasion considers sender commitment rather than receiver commitment, and focuses on the maximization of sender welfare. Our model also differs in that the signal space of the analyst is restricted to a subset of the components, implying that the concavification argument central to Bayesian persuasion does not apply. [2] Mathis (2008) studies models of persuasion with partial certifiability, and provides conditions for the existence of separating equilibria, i.e., full information revelation. These models require continuous types and actions, costless information transmission, and do not allow for receiver commitment. Separating equilibria in these models are analogous to the case in our model where first-best rules can be implemented.

Roadmap The rest of this article is structured as follows. Section 2 presents our formal baseline model. Section 3 characterizes implementable and optimal rules in this model. Section 4 discusses variations of our baseline model with different decision-maker objectives (frequentist hypothesis testing, multi-parameter settings) as well as alternative information and cost structures.

[2] A variation of the Bayesian persuasion model allows for signal-dependent costs for the sender (as in our model); cf Gentzkow and Kamenica (2014). Under posterior separability (a condition which is not satisfied in our model), the concavification approach generalizes to the costly persuasion model.

2 Setup

There are two agents, an analyst and a decision-maker. The analyst observes a vector $X = (X_1, \ldots, X_{\bar{n}})$, where $X_i \in \{0, 1\}$. The analyst can then report a subvector X_I to the decision-maker, where $I \subset \{1, \ldots, \bar{n}\}$. The decision-maker in turn has to make a binary decision $a \in \{0, 1\}$ based on this report. We will sometimes describe the decision $a = 1$ as "acceptance."

Distribution and prior The components X_i are i.i.d. draws from a Bernoulli distribution, $X_i \overset{\text{iid}}{\sim} \text{Ber}(\theta)$. Analyst and decision-maker share a common Beta prior over θ, $\theta \sim \text{Beta}(\alpha, \beta)$. The number of components $\bar{n}$ is known to both analyst and decision-maker.

Objectives The analyst's objective is given by

$$u^{\text{an}} = a - c \cdot |I|, \tag{1}$$

where $|I|$ is the size of the reported set, and c is the cost to the analyst of communicating one component j of X to decision-maker.

The decision-maker's objective is given by

$$u^{\text{d-m}} = a \cdot (\theta - \underline{\theta}). \tag{2}$$

$\underline{\theta}$ is a commonly known parameter determining the minimum value of θ beyond which the decision-maker would like to choose $a = 1$.

Pre-analysis plans and timeline The analyst might specify a PAP before observing the data X. For our purposes, a PAP consists of a list of indices $J \subseteq \{1, \ldots, \bar{n}\}$ which the analyst chooses and reports to the decision-maker.

The timeline of our model is as follows. The decision-maker first commits to a decision rule determining a as a function of analyst reports (J, I, X_I). Before observing any data, the analyst then reports a PAP, that is, a (possibly empty) subset $J \subseteq \{1, \ldots, \bar{n}\}$ of components. The analyst next observes X, chooses $I = I(X)$, and reports (I, X_I). Finally the decision rule $a = a(J, I, X_I)$ is applied and utilities are realized.

Implementability We next define implementability without PAPs; the role of PAPs in implementability is elaborated in Section 3 below. Consider a general reduced form mapping[3] $\bar{a}(x)$ from x to a, where x is a possible realization of X. We

[3] Here, we assume that all mappings are deterministic. We discuss extensions to randomized mappings, and define implementability of randomized rules, when we consider frequentist testing in Section 4.1.

say that the mapping $\bar{a}(x)$ is implementable if there exist mappings $I(x)$ and $a(I, x_I)$ such that for all x

$$\bar{a}(x) = a(I(x), x_{I(x)}) \qquad \text{and} \qquad I(x) \in \underset{I}{\operatorname{argmax}}\ a(I, x_I) - c \cdot |I|. \tag{3}$$

That is, it is optimal for the analyst to actually report $I(X)$ when observing $X = x$, for all x. This incentive-compatibility condition implies in particular that $I(x) \in \operatorname{argmin}_I\{|I| : a(I, x_I) = 1\}$ whenever $\bar{a}(x) = 1$, and $I(x) = \emptyset$ else. It also implies that $|I(x)| \leq 1/c$ for all x. Our goal will be to find implementable mappings $\bar{a}(x)$ that maximize the expected decision-maker utility $E[u^{\text{d-m}}]$.

Discussion of assumptions The model sketched here is chosen to describe the minimal interesting case of statistical inference as a mechanism design problem. This model leads to an extension of classical statistical decision theory, where implementability constrains the set of feasible decision rules.

Our model has two agents with conflicting interests. The decision-maker would like to only choose $a = 1$ if θ exceeds $\underline{\theta}$, while the analyst would always prefer $a = 1$. The model also has analyst private information; the analyst observes X while the decision-maker only observes the reported components X_I. It is the combination of these two model features that generates a potential role for a commitment device, that is, for a pre-analysis plan. The data available to the analyst take the form of i.i.d. binary indicators X_i. We can think of these as the outcomes of statistically independent hypothesis tests of the same hypothesis, an interpretation we expand on when we explicitly model joint frequentist testing in Section 4.1. The analyst decides which of these tests to report. We will drop the assumption of i.i.d. draws X_i in several model extensions considered later.

The PAP J constitutes "cheap talk" in our baseline model, and it might in principle be chosen by the decision-maker rather than the analyst. In an extension of the model where the analyst possesses prior private information, this ceases to be the case, however. The PAP in this baseline model serves as a symmetry-breaking device, differentiating components i which are a-priori exchangeable. In Section 4.4, we explicitly model additional private information of the researcher that makes it optimal for the researcher to choose J.

Interpretations We have described our model in terms of a generic analyst and decision-maker. There are several alternative interpretations of this model. We might think of the model as describing the conflict of interest between an analyst who always wants to reject some hypothesis, and a reader of their research who wants a valid statistical test of that same hypothesis. In this case, the decision a is whether to reject the hypothesis. We will consider frequentist hypothesis testing in greater detail in Section 4.1 below.

We might also think of the model as describing the conflict of interest between a researcher who wants to get published (in order to get tenure, for instance), and a journal that only wants to publish studies on effects that are large enough to be interesting; the relevant decision a here is whether to publish a study.

We might lastly think of the model as describing the conflict of interest between a pharmaceutical company who wants to sell drugs, and a medical regulatory agency (such as the Food and Drug Administration) who wants to protect patient health. In this case, the decision a is whether to approve a drug.

Notation We will use the following notation. The number of successes among the subset of components I is given by $s(X_I) = \sum_{i \in I} X_i$, and the number of successes among all components is $s(X) = \sum_i X_i$. We similarly write $t(X_I) = \sum_{i \in I}(1-X_i)$, and $t(X) = \sum_i (1 - X_i)$ for the number of failures. The maximal number of components the analyst is willing to report is denoted $\bar{n}^{max}$,

$$\bar{n}^{max} = \max \{n : \ 1 - cn \geq 0\} = \lfloor 1/c \rfloor, \tag{4}$$

where $\lfloor \cdot \rfloor$ denotes rounding down to the next integer. The number of successes (components such that $X_i = 1$) that the decision-maker needs to see among n components in order to be willing to accept, absent implementability constraints, is denoted $\underline{s}^{opt}(n)$. This is the optimal (first best) cutoff. The minimal cutoff for the number of successes below which the decision-maker prefers the outside option of never accepting is denoted $\underline{s}^{min}(n)$. Formally,

$$\begin{aligned} \underline{s}^{opt}(n) &= \min \left\{\underline{s} : \ E[\theta \big| s(X_{\{1,\dots,n\}}) = \underline{s}] \geq \underline{\theta}\right\}, \text{ and} \\ \underline{s}^{min}(n) &= \min \left\{\underline{s} : \ E[\theta \big| s(X_{\{1,\dots,n\}}) \geq \underline{s}] \geq \underline{\theta}\right\}. \end{aligned} \tag{5}$$

Since we have assumed a Beta prior, we have $E[\theta \big| s(X) = s] = \frac{\alpha+s}{\alpha+\beta+n}$, and therefore $\underline{s}^{opt}(n) = \lceil \underline{\theta} \cdot (\alpha + \beta + n) - \alpha \rceil$, where $\lceil \cdot \rceil$ denotes rounding up to the next integer.

3 Analysis

In this section, we characterize solutions to the model introduced in Section 2. Specifically, we characterize decision rules which are optimal for the decision-maker subject to the constraint of implementability. To build intuition before discussing the general case, we start by considering the special case $\bar{\mathbf{n}} = \mathbf{3}$. For this special case we show that for some parameter values a symmetric cutoff rule without pre-specification is optimal, while for other values a rule based on a pre-analysis plan dominates. In particular, unless the cost c is small, the unconstrained efficient mapping $\bar{a}(x)$ is not implementable.

We next consider general values of $\bar{n}$, but focus on decision rules which are **symmetric**, in the sense that they depend only on the reported number of components $|I|$ and the number of reported successes $s(X_I)$. Such rules can implement the unconstrained efficient solution when the reporting cost c is low enough or the number of analyst degrees of freedom $\bar{n}$ is small, but not otherwise. For larger values of $\bar{n}$ the decision-maker needs to distort the cutoff $\underline{s}$ down, relative to the first-best cutoff $\underline{s}^{opt}(\bar{n})$, in order to ensure analyst participation. This results in reduced decision-maker welfare. When $\bar{n}$ is too large, no symmetric decision rule can ensure positive welfare for the decision-maker.

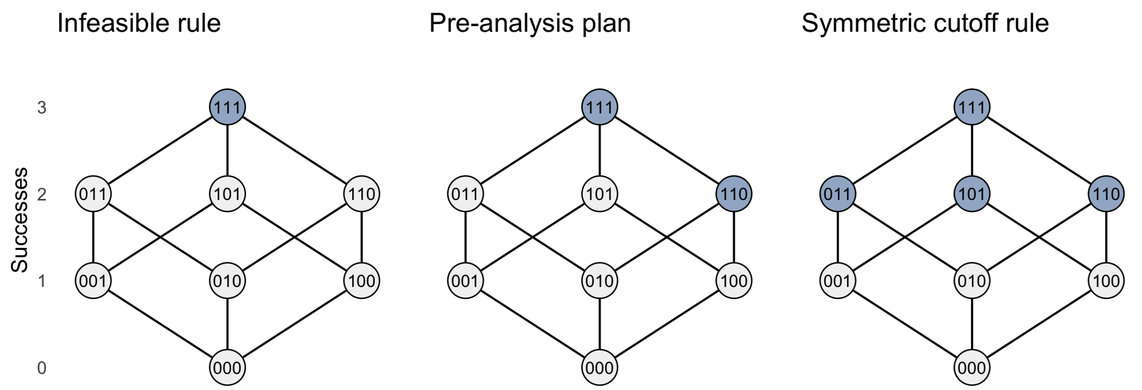
In this symmetric case, we next turn to decision rules based on **pre-analysis plans**. These rules are such that the decision-maker ignores all components which are not part of the pre-specified set J. We show that such decision rules based on pre-analysis plans dominate symmetric rules when the reporting cost c is large or the number of analyst degrees of freedom $\bar{n}$ is too big. Such pre-analysis plans allow the decision-maker to reduce the effective number of components $\bar{n}$ that the analyst can consider. The decision-maker can reduce $\bar{n}$ down to the value that results in maximal expected decision-maker welfare, and avoids distorted acceptance cutoffs.

Lastly, we consider the **general model** without symmetry restrictions. We prove that the implementable decision rules are exactly the rules which accept for values of X in sets which are given by unions of cylinder sets. These cylinder sets correspond to the acceptance regions for PAPs pre-specifying the maximal number of components, $\bar{n}^{max}$. We discuss feasible greedy choices from this general set of implementable rules in some examples.

3.1 Examples where $\bar{n} = 3$

In this section we show by example that in our model the form of the optimal decision rule depends on parameter values. For some parameter values, the decision rule is implementable using a PAP. For other parameter values, a symmetric decision rule is optimal and implements the first-best reduced-form decision rule. We illustrate these different scenarios in Figure 1. Throughout this section, we consider the special case of our model where $\bar{n} = 3$ and $c = 0.4$, so that $\bar{n}^{max} = 2$.

Figure 1: Solutions for $\bar{n} = 3$ and $\bar{n}^{max} = 2$



Notes: This panel shows different possible reduced form mappings from $X = (X_1, X_2, X_3)$ to $\bar{a}(X)$, where realizations of X such that $\bar{a}(X) = 1$ are marked by darker nodes. The vertical axis corresponds to the number of successes $s(X)$. The leftmost figure shows the infeasible rule which only accepts for $X = (1, 1, 1)$; this rule cannot be implemented when $\bar{n}^{max} < 3$. The second figure shows the decision rule implemented by the PAP registering components $J = \{1, 2\}$, and accepting for two reported successes among these. This is optimal when $\underline{s}^{opt}(3) = 3$ and $\underline{s}^{opt}(2) = 2$. The third figure shows the symmetric cutoff rule accepting for 2 successes or more. This is optimal when $\underline{s}^{opt}(3) = 2$.

Case I: Symmetric cutoff rule is optimal Suppose that $\underline{\theta}$ is such that

$$E[\theta] < \underline{\theta} < E[\theta|s(X) = 2].$$

For the uniform prior with $\alpha = \beta = 1$, for instance, this holds whenever $\underline{\theta} \in].5, .6[$. Then $\underline{s}^{opt}(3) = 2$. The unconstrained efficient solution is given by

$$\bar{a}(X) = \mathbf{1}(s(X) \geq 2).$$

This solution can be implemented by

$$a(J, I, X_I) = \mathbf{1}(s(X_I) \geq 2),$$

without requiring a PAP.

Case II: PAP is optimal Suppose now instead that $\underline{\theta}$ is such that

$$E[\theta|s(X) = 2] < E[\theta|s(X) \geq 2] < \underline{\theta} < E[\theta|s(X_{\{1,2\}}) = 2] < E[\theta|s(X) = 3].$$

For the uniform prior with $\alpha = \beta = 1$, for instance, this holds whenever $\underline{\theta} \in].7, .75[$. Then $\underline{s}^{opt}(3) = 3$ and $\underline{s}^{opt}(2) = 2$. The unconstrained efficient solution is given by

$$\bar{a}(X) = \mathbf{1}(s(X) = 3).$$

This rule is illustrated on the left of Figure 1. There is no incentive compatible implementation of this solution, however, since $\bar{n}^{max} = 2$. The PAP solution for $J = \{1, 2\}$,

$$a(J, I, X_I) = \mathbf{1}(I = J = \{1, 2\}, s(X_I) = 2),$$

yields positive expected welfare $E[u^{\text{d-m}}] > 0$, and is indeed the constrained optimal solution in this case. This solution is illustrated in the middle of Figure 1. No symmetric decision rule of the form $a(J, I, X_I) = \mathbf{1}(s(X_I) > \underline{s})$ which respects the analyst reporting constraint yields positive expected decision-maker welfare for this example, because $E[\theta|s(X) \geq 2] < \underline{\theta}$ by assumption.

3.2 Symmetric decision rules

We next return to the model with a general number of components $\bar{n}$, but we restrict our attention in this section to symmetric decision rules, which are invariant to permutations of the components $1, \ldots, \bar{n}$. For such decision rules, we can write (overloading notation)

$$a(J, I, X_I) = a(s(X_I), t(X_I)), \tag{6}$$

where $t(X_I) = |I| - s(X_I)$. In words, acceptance only depends on the number of successes among the reported components, and on the number of failures. Such decision rules ignore the pre-analysis plan, should the analyst report one. Below, we generalize this approach to rules based on pre-analysis plans which require symmetry only among the pre-specified components J, while ignoring all other components.

In the following we characterize optimal symmetric decision rules using the mechanism-design approach sketched in Section 2. We first characterize the set of all reduced-form decision rules $\bar{a}(\cdot)$ which are implementable by symmetric decision rules. We then describe the optimal symmetric decision rules subject to the constraint of implementability. Thereafter, we consider the special case of a uniform prior, which allows us to derive analytic solutions which can be plotted (cf. Figure 2). We then show that the qualitative comparative statics exhibited by the solution for the uniform case do, in fact, hold more generally.

Implementable rules As a first step, the following lemma characterizes the reduced-form decision rules $\bar{a}(X)$ which are implementable by symmetric decision rules of the form $a(s(X_I), t(X_I))$. It is worth emphasizing that the set of implementable decision rules is independent of decision-maker preferences, by construction. This will allow us to consider other decision-maker objectives in Section 4 below, for instance objectives corresponding to frequentist testing with size control, referring to the same implementability results derived in this section. The condition on c in the following lemma avoids indifference of the analyst.

Lemma 1 (Symmetrically implementable rules)**.** *Assume that* $(1/c) \notin \mathbb{N}$*, and consider decision rules of the form* $a(s(X_I), t(X_I))$*.* $\bar{a}(\cdot)$ *is a reduced-form decision rule implementable by such a rule* $a(\cdot)$ *if and only if it is of the form*

$$\bar{a}(X) = \mathbf{1}(s(X) \in \mathcal{S}), \tag{7}$$

where $\mathcal{S}$ *is a union of intervals of length at least* $\bar{n} - \bar{n}^{max}$*.*[4]

This Lemma implies in particular that if a decision rule is implementable and symmetric and accepts for $s(X) = \bar{n}$, then it also has to accept for any value of S in the interval $[\bar{n}^{max}, \bar{n}]$.

Optimal rules The following proposition derives the optimal implementable symmetric decision rule. This decision rule applies a cutoff which is illustrated by the bold blue line in the left panel of Figure 2. Note that the first-best reduced-form rule (which neglects the constraint of implementability) is given by $\bar{a}(X) = \mathbf{1}(s(X_I) \geq \underline{s}(\bar{n}))$. This first-best decision rule is implementable iff $\bar{n}^{max} \geq \underline{s}^{opt}(\bar{n})$.

Proposition 1 (The optimal symmetric rule)**.** *If* $\bar{n}^{max} \geq \underline{s}^{min}(\bar{n})$*, the optimal reduced-form decision rule that is symmetrically implementable takes the form*

$$\bar{a}(X) = \mathbf{1}(s(X) \geq \min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max})), \tag{8}$$

and can be implemented by

$$a(s(X_I), t(X_I)) = \mathbf{1}(s(X_I) \geq \min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max})). \tag{9}$$

If $\bar{n}^{max} < \underline{s}^{min}(\bar{n})$*, then the optimal symmetrically implementable decision rule is given by* $a(s(X_I), t(X_I)) \equiv 0$*.*

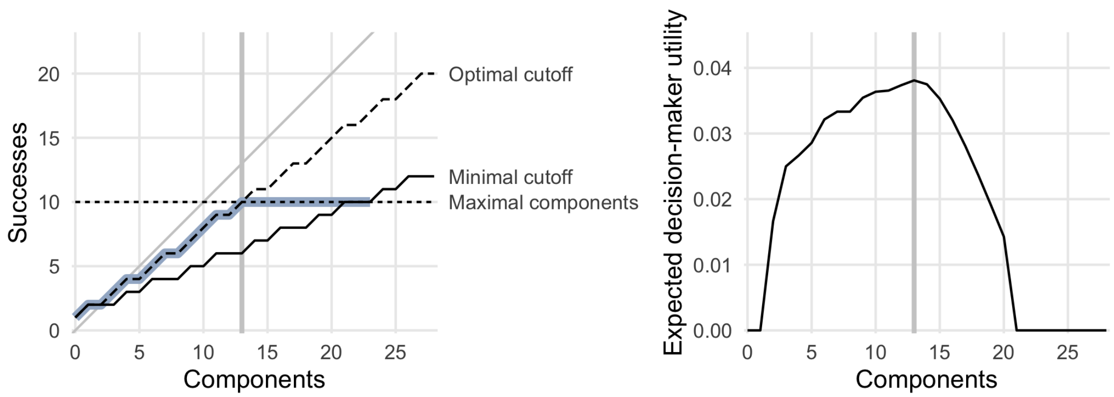
Uniform prior Assume now additionally that the prior for θ is uniform, so that the parameters of the prior Beta distribution are given by $\alpha = \beta = 1$. The number of successes S, which generally has a Beta-Binomial distribution in our model, has a discrete uniform distribution on $\{0, 1, \ldots, \bar{n}\}$ in this special case. The following identities are then immediate.

$$P(s(X) \geq \underline{s}) = \frac{\bar{n} + 1 - \underline{s}}{\bar{n} + 1}, \quad E[\theta | s(X) = s] = \frac{1 + s}{2 + \bar{n}}, \quad E[\theta | s(X) \geq \underline{s}] = \frac{1}{2 + \bar{n}} \cdot \frac{2 + \underline{s} + \bar{n}}{2}.$$

We can now derive the optimal symmetric decision rule in closed form, combining Proposition 1 with these identities.

[4] Note that intervals in the set of integers that are of length k are of cardinality $k + 1$.

Figure 2: Symmetric cutoff without PAP, uniform prior



Notes: This figure plots the solution described in Corollary 1, for the case where $\underline{\theta} = .7$ and $c = .1$. The left figure shows the optimal cutoff $\underline{s}^{opt}(\bar{n})$ and the minimal cutoff $\underline{s}^{min}(\bar{n})$ (the two upward sloping lines), the maximal number of components reported by the analyst $\bar{n}^{max}$ (the horizontal line), and then the actual cutoff (bold, in blue). The right figure plots the decision-maker's expected utility, as a function of $\bar{n}$. The vertical grey line indicates the number of components at which the analyst reporting constraint starts to bind; here, this is also the number of components which maximizes expected decision-maker utility.

Corollary 1 (Uniform distribution)**.** *Suppose that $\alpha = \beta = 1$. Then the optimal reduced-form symmetric decision rule, and the corresponding expected decision-maker welfare, are given by*

$$\bar{a}(X) = \mathbf{1}(s(X) \geq \min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max})) \cdot \mathbf{1}\left(\bar{n}^{max} \geq \underline{s}^{min}(\bar{n})\right),$$

$$E[u^{\text{d-m}}] = \left(\frac{2+\min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max}) + \bar{n}}{(2+\bar{n})2} - \underline{\theta}\right) \cdot \left(\frac{\bar{n}+1-\min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max})}{\bar{n}+1}\right) \cdot \mathbf{1}\left(\bar{n}^{max} \geq \underline{s}^{min}(\bar{n})\right),$$

where

$$\underline{s}^{opt}(\bar{n}) = \lceil \underline{\theta} \cdot (2+\bar{n}) - 1 \rceil, \quad \underline{s}^{min}(\bar{n}) = \lceil \underline{\theta} \cdot (4+2\bar{n}) - 2 - \bar{n} \rceil, \quad \bar{n}^{max} = \lfloor 1/c \rfloor.$$

Three regimes Figure 2 plots the solution described in Corollary 1, for the case where $\underline{\theta} = .7$ and $c = .1$. This figure shows that the optimal symmetric cutoff rule is determined by one of three regimes, depending on the value of $\bar{n}$ relative to c. For $\bar{n}$ **small** enough (given c), $\underline{s}^{opt}(\bar{n})$ is less than $\bar{n}^{max}$, and the first best decision can be implemented by using the cutoff $\underline{s}^{opt}(\bar{n})$. For $\bar{n}$ **large**, we have that $\underline{s}^{min}(\bar{n})$ is greater than $\bar{n}^{max}$, and there is no cutoff that would guarantee both analyst and decision-maker participation; acceptance never occurs. For **intermediate** values of $\bar{n}$, we have that $\bar{n}^{max}$ lies between $\underline{s}^{min}(\bar{n})$ and $\underline{s}^{opt}(\bar{n})$. The decision-maker then uses the cutoff $\bar{n}^{max}$, which is just small enough to ensure analyst participation, while still guaranteeing positive expected welfare to the decision-maker.

The decision-maker's expected welfare behaves correspondingly. For small $\bar{n}$, welfare is increasing in $\bar{n}$, since the Blackwell-informativeness of $X = (X_1, \ldots, X_{\bar{n}})$ is increasing in $\bar{n}$. For large $\bar{n}$, expected welfare equals 0, since no symmetric rule exists that can give positive expected welfare to both the analyst and the decision-maker. For intermediate $\bar{n}$, welfare is decreasing in $\bar{n}$. The reason is that the cutoff is increasingly distorted downward, relative to the decision-maker optimum, in order to ensure analyst participation.

These considerations generalize; they do not depend on the uniform prior for θ, as shown by the following corollary.

Corollary 2 (Comparative statics of decision-maker welfare for symmetric rules)**.** *Consider expected welfare $E[u^{\text{d-m}}]$ of the decision-maker for the optimal symmetric decision rule, as a function of $\bar{n}$, holding all other parameters of the model constant. Then the following holds true:*

1. *For $\bar{n} = 0$,*
   $E[u^{\text{d-m}}] = 0$.

2. *For $\bar{n}$ small enough*
   so that $\underline{s}^{opt}(\bar{n}) = \lceil \underline{\theta} \cdot (\alpha + \beta + \bar{n}) - \alpha \rceil \leq \bar{n}^{max} = \lfloor 1/c \rfloor$,
   $E[u^{\text{d-m}}]$ is increasing in $\bar{n}$.

3. *For intermediate values of $\bar{n}$,*
   if there exists a value of $\bar{n}$ such that $\underline{s}^{min}(\bar{n}) \leq \min(\bar{n}, \bar{n}^{max})$,
   $E[u^{\text{d-m}}] > 0$.

4. *For $\bar{n}$ large enough*
   so that $\underline{s}^{min}(\bar{n}) > \bar{n}^{max} = \lfloor 1/c \rfloor$,
   $E[u^{\text{d-m}}] = 0$.

Pre-analysis plans The preceding analysis suggests that, when $\bar{n}$ is large, it might beneficial for the decision-maker to somehow reduce the number of components $\bar{n}$, that is, to reduce the analyst degrees of freedom. This is exactly what a pre-analysis plan allows them to achieve.

Generalizing our analysis for symmetric rules, consider now rules which are symmetric among the pre-registered and reported components $i \in I \cap J$, while ignoring all other components, so that

$$a(J, I, X_I) = a(|J|, s(X_{I \cap J}), t(X_{I \cap J})). \tag{10}$$

Proposition 2 (Optimal PAP when full pre-specification is required)**.** *Suppose that $\underline{s}^{opt}(\bar{n}) \geq \bar{n}^{max}$. Consider decision rules of the form $a(|J|, s(X_{I \cap J}), t(X_{I \cap J}))$. An optimal decision rule of this form is given by*

$$a(|J|, s(X_{I \cap J}), t(X_{I \cap J})) = \mathbf{1}\left(|J| = \bar{n}^* \text{ and } s(X_{I \cap J}) \geq \bar{n}^{max}\right), \tag{11}$$

where

$$\bar{n}^* = \underset{n}{\text{argmax}}\ E[\mathbf{1}(s(X_{\{1,\dots,n\}}) \geq \bar{n}^{max}) \cdot (\theta - \underline{\theta})]. \tag{12}$$

An immediate consequence of this result is the following case distinction. If $\underline{s}(\bar{n}) \leq \bar{n}^{max}$, then the **first-best** reduced-form decision rule can be implemented, and is given by $\bar{a}(X) = \mathbf{1}(s(X) \geq \underline{s}(\bar{n}))$. If $\underline{s}(\bar{n}) > \bar{n}^{max}$, then the first-best is not implementable, and symmetric decision rules are dominated by decision rules requiring a **pre-analysis plan**. The optimal pre-analysis plans of Proposition 2 reduce the effective number of components available to the analyst down to the optimal number, corresponding to the peak of the expected decision-maker utility as shown in Figure 2. The pre-analysis plan effectively acts as a symmetry-breaking device that expands the set of implementable symmetric decision rules (where symmetry applies only to the pre-specified components).

3.3 General implementable decision rules

In the previous section we considered symmetric decision rules, as well as symmetric rules with a PAP. We now return to fully general rules. The following lemma characterizes the set of implementable reduced-form decision rules $\bar{a}(x)$. This lemma generalizes Lemma 1 discussed above, where we considered reduced-form rules implementable by symmetric rules.

Lemma 2. *The implementable reduced-form decision rules $\bar{a}(x)$ are exactly those that are of the form*

$$\bar{a}(x) = \mathbf{1}(x \in \cup_k C_{I_k, w_k}), \tag{13}$$

for some set of $\{(I_k, w_k)\}$, where $C_{I,w}$ are the cylinder sets

$$C_{I,w} = \{x :\ x_I = w\}, \tag{14}$$

and $|I_k| = \bar{n}^{max}$ for all k.

This lemma shows that the set of values of x for which we get $\bar{a}(x) = 1$ is necessarily given by a union of sets of the form $C_{I,w} = \{x :\ x_I = w\}$, if $\bar{a}(\cdot)$ is implementable. These are cylinder sets; they fix the value of x on a subset of components I, to the values specified by w. Furthermore, the size of I is (at most) $\bar{n}^{max}$ – the cylinder sets pin down (at most) $\bar{n}^{max}$ components.

To gain some intuition, consider again the example where $\bar{n} = 3$ but $\bar{n}^{max} = 2$, as discussed in Section 3.1 and depicted in Figure 1. In this example, the cylinder sets of Lemma 2 correspond to edges of the cube shown in Figure 1. The set $\{(1, 1, 1)\}$ is not implementable (infeasible), since it is not a union of such edges. The sets $\{x : x_1 = x_2 = 1\}$ (corresponding to a pre-analysis plan), and $\{x : s(x) \geq 2\}$ (corresponding to the symmetric cutoff rule) however, are implementable, as they can be written as a union of edges.

Brute-force optimization The optimal decision rule (from the decision-maker's point of view) maximizes the expected utility $E[u^{\text{d-m}}]$ subject to the constraint of implementability. In light of Lemma 2, we can write this problem as a problem of maximizing over implementable sets $\boldsymbol{A} = \{x : \ a(x) = 1\}$,

$$\max_{\boldsymbol{A}\in\mathscr{A}} \sum_{x\in\boldsymbol{A}} \omega_x, \text{ where} \qquad \mathscr{A} = \{\boldsymbol{A} : \ \boldsymbol{A} = \cup_k C_{I_k,w_k}\}, \tag{15}$$
$$C_{I,w} = \{x : \ x_I = w\}, \text{ and} \qquad \omega_x = P(x) \cdot (E[\theta|s(X) = s(x)] - \underline{\theta})\,.$$

This problem can, in, principle, be solved by brute-force enumeration of all possible sets $\boldsymbol{A}$ (there are a finite number), maximizing the corresponding expected decision-maker utility.

In practice, however, enumeration quickly becomes computationally infeasible, due to the large number of possible sets $\boldsymbol{A}$. There are $\binom{\bar{n}}{\bar{n}^{max}}$ possible subsets I, and for each of those $2^{\bar{n}^{max}}$ possible values of w. The number of possible cylinder sets $C_{I,w}$ (not taking advantage of symmetries in the problem), thus is given by $\binom{\bar{n}}{\bar{n}^{max}} \cdot 2^{\bar{n}^{max}}$. The first term grows as $\bar{n}^{\bar{n}^{max}}$, for fixed $\bar{n}^{max}$; the second term grows exponentially in $\bar{n}^{max}$. The size of $\mathscr{A}$, finally, grows as 2 to the power of the number of possible cylinder sets $C_{I,w}$.

Greedy optimization This observation suggests other approaches to optimization. We consider the following greedy algorithm: Start with the empty set $\boldsymbol{A} = \emptyset$. At each step of the algorithm, augment $\boldsymbol{A}$ by a set $C_{I,w}$, where w is a vector of ones, and I is chosen among all possible subsets of $1, \ldots, \bar{n}$, in order to yield the largest increase of $E[u^{\text{d-m}}]$. This step is further simplified by using the fact that all components not considered yet by $\boldsymbol{A}$ are symmetric, so we do not need to consider all of them. The algorithm stops when no more set I can be found which leads to an increase of $E[u^{\text{d-m}}]$.

This greedy algorithm leads to reasonable results in the numerical examples discussed in Appendix A.1. It is not guaranteed to find the global optimum, however. The underlying reason is a lack of supermodularity in the combinatorial optimization problem. Convergence of the greedy algorithm to the global optimum requires the absence of complementarities: Augmenting $\boldsymbol{A}$ by two sets $C_{I,w}$ and $C_{I',w'}$ should lead to a lower increase of $E[u^{\text{d-m}}]$ then the sum of increases from augmenting $\boldsymbol{A}$ by either of the two sets separately. This does not always hold, as shown in an example in Appendix A.1, where the greedy algorithm does not find the global optimum. That said, in the examples we considered, it appears that the solution found by the greedy algorithm still performs quite well, and better than simpler alternatives.

4 Model variations and extensions

In this section, we discuss several variations and extensions of our baseline model. We first consider alternative decision-maker objectives, while maintaining the same analyst objective and information structure as before. This immediately implies that the same reduced form decision functions $\bar{a}(\cdot)$ are implementable as in the baseline model; in particular Lemma 1 and Lemma 2 above continue to describe the set of symmetrically implementable decision functions, and the set of implementable decision functions, respectively. We then consider alternative information and cost structures. For the sake of exposition, we restrict our attention throughout to symmetric cutoff rules, as in Proposition 1, where symmetry is possibly restricted to the pre-registered components.

Before discussing the specific model variations in greater detail, let us briefly preview and discuss the purpose of introducing each of these variations and extensions.

In Section 4.1 we replace the decision-maker objective with a frequentist testing problem, where the goal is to maximize power subject to a constraint on the size of a test. This extension shows that our qualitative conclusions hold up when we replace decision-maker utility maximization by frequentist testing.

In Section 4.2, we replace the decision-maker objective with the objective $u^{\text{d-m}}(a) = a \cdot \sum_{i \in I} (\theta_i - \underline{\theta})$, where now there is a different parameter θ_i corresponding to each component X_i of the data, but these parameters are ex-ante correlated. This extension shows that our qualitative conclusions again hold up when we allow for multiple parameters, as long as they are not statistically independent in the prior.

In Section 4.3, we drop the assumption of decision-maker commitment. We show that the optimal symmetrically implementable rules derived in Proposition 1 are sustained by a perfect Bayesian equilibrium in this model without commitment. Moreover, these rules are optimal from the decision-maker's point of view, among the symmetrically implementable perfect Bayesian equilibria.

In Section 4.4, we allow for the possibility of additinoal asymmetric information between the decision-maker and the analyst, in terms of the available set of hypotheses. This extension implies that it is better for the analyst (rather than the decision-maker) to choose the PAP, while all other conclusions from the baseline model continue to hold.

In Section 4.5, we allow for the possibility that the number of components $\bar{n}$ (degrees of freedom) is unknown to both the analyst and the decision-maker ex-ante. This extension shows that uncertainty over $\bar{n}$ provides an additional rationale for PAPs that is separate from costs of communication.

In Section 4.6, we assume that the decision-maker, rather than the analyst, bears the cost $c \cdot |I|$ of communication. For this model variation, implementability is not a binding constraint. Our qualitative conclusions again continue to hold, nonetheless, for the implied decision problem.

4.1 Frequentist testing

In the baseline model, we interpreted the function a as the decision rule of a decision-maker who aims to maximize the expectation of their utility $u^{\text{d-m}}$. Alternatively, we could interpret a as the testing decision of a reader based on evidence X_I presented by an analyst, where the components X_i are themselves the outcomes of statistically independent hypothesis tests. This is the perspective we take in this section.

Assumptions Consider the same model as outlined in Section 2, except that now the decision-maker wants to test the null hypothesis $\theta \leq \underline{\theta}$, for some significance level $\underline{\theta}$, e.g. $\underline{\theta} = .05$. Since, by assumption, $X_i \sim \text{Ber}(\theta)$, this implies that X_i itself is a valid test that controls size for the hypothesis $\theta \leq \underline{\theta}$. Any statistical test X_i can be written in this form after suitable re-parametrization: We can replace the original underlying parameter by θ, which is *defined* as the rejection probability of the test.

Tests with higher power can be obtained by combining the evidence from multiple components X_i. We next discuss the construction of first-best and implementable testing rules. We require that these rules satisfy the size constraint, and we aim to maximize power. Note that the analyst's incentives are the same as before. It immediately follows that the (symmetrically) implementable non-randomized tests $\bar{a}(X)$ are exactly those described by Lemma 1 and Lemma 2.

The first-best rule We start by characterizing the first-best testing rule, sidestepping the question of implementability. The number of successes $S = s(X)$ is a sufficient statistic for θ. Furthermore, $s(X)$ satisfies the monotone likelihood ratio property. We can therefore apply Theorem 3.4.1 in Lehmann and Romano (2006), which shows that

$$\bar{a}(X) = \mathbf{1}(s(X) + U \geq \underline{s}^{test}(\bar{n})), \tag{16}$$

where U is distributed Uniform$([0, 1])$, is a uniformly most powerful test for the null $\theta \leq \underline{\theta}$. Here U is a tie-braking device to ensure exact size control, despite the discreteness of the test-statistic $s(X)$.[5]

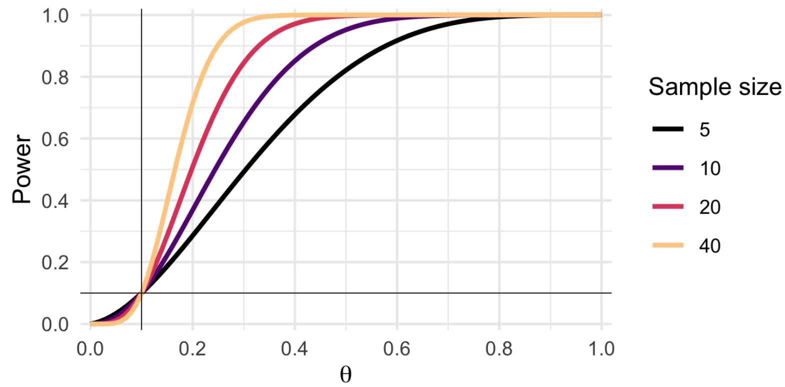
The ex-ante probability conditional on θ that such a test, using a cutoff $\underline{s}$, rejects the null is given by

$$\begin{aligned} p(\theta; \bar{n}, \underline{s}) &= E[\bar{a}(X)|\theta] = P(s(X) + U \geq \underline{s}|\theta) \\ &= \sum_{j=\lceil \underline{s} \rceil}^{\bar{n}} \binom{\bar{n}}{j} \theta^j (1-\theta)^{\bar{n}-j} + (\lceil \underline{s} \rceil - \underline{s}) \cdot \binom{\bar{n}}{j} \theta^{(\lceil \underline{s} \rceil - 1)} (1-\theta)^{\bar{n}+1-\lceil \underline{s} \rceil}. \end{aligned}$$

The last term reflects the probability of rejection due to the tie-breaking device. For any θ under the null, that is for $\theta \leq \underline{\theta}$, this probability is bounded above by $p(\underline{\theta}; \bar{n}, \underline{s})$.

[5] Allowing for more general forms of randomization can increase the set of implementable testing rules. Such more general randomized tests are considered in Proposition 3 below, and in the context of a numerical example in Appendix A.2.

Figure 3: Power curves for different sample sizes



Notes: This figure illustrates the value of increased sample size for the first-best rule by showing the power curves for different values of n, for $\underline{\theta} = .1$ and the test of Equation (16).

The critical value $\underline{s}^{test}(\bar{n})$ is given by the unique value which yields size control,[6]

$$p(\underline{\theta}; \bar{n}, \underline{s}^{test}(\bar{n})) = \underline{\theta}. \tag{17}$$

This equation assumes that the cutoff defining the null hypothesis ($\underline{\theta}$ on the left hand side), and thus the size of each of the X_i as tests of this null, is the same as the size of the joint test ($\underline{\theta}$ on the right hand side). In principle these two cutoffs could be different. The following immediately generalizes to the case with different cutoffs.

Blackwell dominance implies that a larger value of $\bar{n}$, when using the cutoff $\underline{s}^{test}(\bar{n})$, leads to a more powerful test. Specifically, $p(\theta; \bar{n}, \underline{s}^{test}(\bar{n}))$ is monotonically increasing in $\bar{n}$ for $\theta > \underline{\theta}$, and is decreasing in $\bar{n}$ for $\theta < \underline{\theta}$. Figure 3 illustrates this point for $\underline{\theta} = .1$ and different values of $\bar{n}$.

Non-randomized symmetrically implementable tests and PAPs We next consider optimal symmetric testing rules subject to implementability. If $\underline{s}^{test}(\bar{n}) \leq \bar{n}^{max}$, the first-best testing rule of Equation 16 can be directly implemented, without a PAP. If $\underline{s}^{test}(\bar{n}) > \bar{n}^{max}$, this is not the case. Furthermore, if we restrict our attention to non-randomized and symmetrically implementable rules, as in Section 3.2, we obtain a stark negative result: No non-trivial symmetric non-randomized tests that control size exist, in this case. This illustrates the common intuition that PAPs are necessary for size control in statistical testing.

Our proof of this claim builds directly on the implementability results from Section 3.2. In order to define implementability for randomized rules, we assume that

[6]Using Chernoff's inequality, $\underline{s}^{test}(\bar{n})$ can be bounded as follows, resulting in alternative valid but conservative critical values: $\underline{s}^{test}(\bar{n}) \leq \underline{\theta}\bar{n} + \sqrt{-3\log(\underline{\theta})}\sqrt{\underline{\theta}\bar{n}}$.

any randomization device U is drawn by the decision-maker and known to the analyst before they choose I. Conditional on U, all previous definitions and results apply.

Proposition 3.

1. *If* $\underline{s}^{test}(\bar{n}) \leq \bar{n}^{max}$ *then the symmetric cutoff rule* $a(I, X_I) = \mathbf{1}(s(X_I) + U \geq \underline{s}^{test}(\bar{n}))$ *implements a uniformly most powerful test based on* X*, subject to size control.*

2. *If* $\lfloor \underline{s}^{test}(\bar{n}) \rfloor > \bar{n}^{max}$ *then no uniformly most powerful test is implementable.*

3. *Furthermore, if* $\underline{s}^{test}(\bar{n}) > \bar{n}^{max}$ *and* $\underline{\theta} < .5$*, then no non-randomized symmetric decision rule of the form* $a(s(X_I), t(X_I))$ *exists which implements a test that controls size under the null and has positive power under any alternative.*

It immediately follows, in analogy to Proposition 2, that a PAP provides value by ensuring size control for a non-trivial non-randomized and symmetric test whenever $\underline{s}^{test}(\bar{n}) > \bar{n}^{max}$. Furthermore, in light of the monotonicity of power noted above, the optimal choice for the size $|J|$ of the PAP is given by

$$\max\{\bar{n} : \underline{s}^{test}(\bar{n}) \leq \bar{n}^{max}\}. \tag{18}$$

Summary and implications The application of our model to hypothesis testing shows that the take-aways from the previous section extend: When analyst and decision-maker have misaligned preferences, and analyst degrees of freedom $\bar{n}$ and/or the cost of communication c are sufficiently high, then a pre-analysis plan improves "welfare" – in this case, size and power. The optimal PAP delivers both size control and non-trivial power.

4.2 Multiple parameters

We next consider a variation of our baseline model where instead of one parameter θ governing the distribution of all the X_i, we have separate parameters θ_i for each outcome. A possible interpretation of this model is the drug-approval process, where the i correspond to different sub-populations, and θ_i describes the effectiveness of the drug for sub-population i. The medical authority's objective is to approve the drug for subpopulations where it is effective, but not for other sub-populations. A key feature of the following model is that the parameters θ_i are a-priori dependent; this dependency is captured by a hierarchical Bayesian model. This dependency implies that selective reporting distorts the posterior for the reported components; absent such dependency the model would be separable across components.

Assumptions Consider a variation of our baseline model where there are parameters θ_i for every i. Suppose that the joint distribution of data and parameters is given by the hierarchical model:

$$\begin{aligned} X_i|\theta_1, \ldots \theta_{\bar{n}}, \bar{\theta} &\sim \text{Ber}(\theta_i) \\ \theta_i|\bar{\theta} &\sim \text{Beta}(m\bar{\theta}, m(1-\bar{\theta})) \\ \bar{\theta} &\sim \text{Beta}(\alpha, \beta). \end{aligned} \tag{19}$$

Assume that the X_i and the θ_i are conditionally independent across i, and m is common knowledge. Suppose that the decision-maker's objective is given by

$$u^{\text{d-m}}(a) = a \cdot \sum_{i \in I} (\theta_i - \underline{\theta}), \tag{20}$$

while the analyst's utility is as before. The latter implies again that the (symmetrically) implementable decision rules $\bar{a}(X)$ are exactly those described by Lemma 1 and Lemma 2.

The data generating process in this model reduces to our baseline model when we take the limit $m \to \infty$; in this limit, the θ_i are perfectly correlated. Decision-maker utility in this limit, however, is multiplied by $|I|$, relative to the baseline model.

The first-best rule By integrating out the θ_i, we immediately get $X_i|\bar{\theta} \overset{iid}{\sim} \text{Ber}(\bar{\theta})$ and thus $s(X)|\bar{\theta} \sim \text{Bin}(\bar{n}, \bar{\theta})$, so that

$$E[\bar{\theta}|s(X)] = \frac{\alpha + s(X)}{\alpha + \beta + \bar{n}}.$$

By the law of iterated expectations, and by conditional independence of θ_i and $s(X)$ given $\bar{\theta}$,

$$\begin{aligned} E[\theta_i|X] &= E[\theta_i|X_i, s(X)] = E\left[E[\theta_i|X_i, \bar{\theta}]|X_i, s(X)\right] \\ &= E\left[\frac{m\bar{\theta} + X_i}{m+1}\Big|X_i, s(X)\right] = \frac{1}{m+1}\left[m \cdot \frac{\alpha + s(X)}{\alpha + \beta + \bar{n}} + X_i\right]. \end{aligned}$$

The posterior mean of θ_i given the full data is thus linear in the total number of successes $s(X)$ and in the component-specific value X_i. From this expression for the posterior mean, it is immediate that the first-best decision rule, for the decision-maker, is given by

$$\begin{aligned} I(X) &= \{i : \; X_i = 1\} \\ \bar{a}(X) &= \mathbf{1}(s(X) \geq \underline{s}^{mult}(\bar{n})) \\ \underline{s}^{mult}(\bar{n}) &= \min\{s : \; E[\theta_i|X_i = 1, s(X) = s] \geq \underline{\theta}\} \\ &= \left\lceil \tfrac{(m+1)\underline{\theta} - 1}{m} \cdot (\alpha + \beta + \bar{n}) - \alpha \right\rceil. \end{aligned} \tag{21}$$

Notice that now, in contrast to the baseline model, the reporting rule $I(X)$ directly enters decision-maker utility; the decision-maker wants only successes to be reported. This first-best decision rule is implementable iff $\underline{s}^{mult}(\bar{n}) \leq \bar{n}^{max}$. $\underline{s}^{mult}(\bar{n})$ is increasing in $\bar{n}$ as long as $(m+1)\underline{\theta} \geq 1$.

Symmetric cutoff rules and PAPs Consider now a general symmetric cutoff rule of the form $a = \mathbf{1}(s(X_I) \geq \underline{s})$. As before, the analyst's best response is to only report successes, where

$$|I(X)| = s(X_{I(X)}) = \underline{s} \cdot \mathbf{1}(s(X) \geq \underline{s}, \bar{n}^{max} \geq \underline{s}).$$

It immediately follows that $E[u^{\text{d-m}}] = 0$ when $\bar{n}^{max} < \underline{s}$. When $\bar{n}^{max} \geq \underline{s}$, then

$$\begin{aligned}
E[u^{\text{d-m}}] &= \underline{s} \cdot \left(E[E[\theta_i | X_i = 1, \bar{\theta}] | s(X) \geq \underline{s}] - \underline{\theta}\right) \cdot P(s(X) \geq \underline{s}) \\
&= \underline{s} \cdot \left(\frac{1}{m+1} \left[m \cdot \frac{\alpha + E[s(X) | s(X) \geq \underline{s}]}{\alpha + \beta + \bar{n}} + 1 \right] - \underline{\theta} \right) \cdot P(s(X) \geq \underline{s}).
\end{aligned}$$

By the same argument as in the proof of Proposition 1, it is optimal for the decision-maker to choose a cutoff $\underline{s} = \min(\underline{s}^{mult}(\bar{n}), \bar{n}^{max})$, as long as expected decision-maker utility for this cutoff is positive, and to never accept for larger $\bar{n}$. Furthermore, since $E[u^{\text{d-m}}]$ is decreasing in $\bar{n}$ beyond some point, it is optimal to use a PAP to reduce $\bar{n}$ to the utility maximizing value, as in Proposition 2.

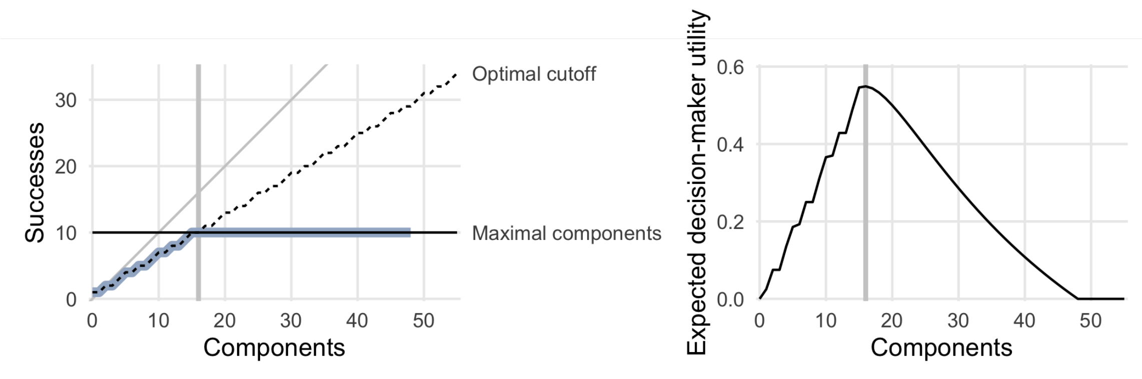
Uniform prior Assume now that the prior for $\bar{\theta}$ is uniform, so that $\alpha = \beta = 1$. As before, it follows that $s(X)$ is uniform on $\{0, \ldots, \bar{n}\}$, and we can calculate

$$E[u^{\text{d-m}}] = \underline{s} \cdot \left(\frac{1}{m+1} \left[m \cdot \frac{\alpha + (\underline{s} + \bar{n})/2}{\alpha + \beta + \bar{n}} + 1 \right] - \underline{\theta} \right) \cdot \frac{\bar{n} + 1 - \underline{s}}{\bar{n} + 1}. \tag{22}$$

Figure 4 plots an example of this solution. Note that relative to our baseline model, the effect of selective reporting on decision-maker utility is attenuated in this model. In the plot, this is reflected in decision-maker utility going to 0 more slowly, and remaining positive even for relatively large $\bar{n}$, despite considerable distortion of $\underline{s}$ relative to the optimum.

Summary and implications When the analyst observes measurements from *different* parameters that come from a joint prior, then the decision-maker is still more likely to approve the analyst's submission when more successes are submitted, and the conclusions from the main model extend. Specifically, when analyst degrees of freedom $\bar{n}$ and/or the cost of communication c are sufficiently high, then a pre-analysis plan improves welfare. However, the effect of misalignment and communication cost is now attenuated since only the selected components enter the decision-maker's payoff, rather than the overall parameter.

Figure 4: Symmetric cutoff, multiple parameters, uniform prior



Notes: This figure plots the solution described in Equation 22, for the case where $\underline{\theta} = .7$, $c = .1$, and $m = 3$. The figure on the left shows the optimal cutoff $\underline{s}^{mult}(\bar{n})$ (the upward sloping line), the maximal number of components $\bar{n}^{max}$ reported by the analyst (the horizontal line), and then the actual cutoff (bold, in blue). The figure on the right plots the decision-maker's expected utility, as a function of $\bar{n}$. The vertical grey line indicates the number of components at which the analyst reporting constraint starts to bind; this is also the number of components which maximizes expected decision-maker utility.

4.3 No commitment

Throughout, we have assumed that the decision-maker is able to commit to a decision rule $a(\cdot)$. As it turns out, the optimal cutoff rule derived in Proposition 1, and the corresponding rule with pre-registration characterized in Proposition 2, can be maintained without commitment, as well. Put differently, these rules are supported by a perfect Bayes Nash equilibrium. The same is true for a range of other cutoff rules.

Assumptions Consider the same model as outlined in Section 2, except that now the timeline is given as follows. Before observing any data, the analyst reports a PAP, that is, a subset $J \subseteq \{1, \ldots, \bar{n}\}$ of components. The analyst then observes X, chooses $I = I(X)$, and reports (I, X_I). Then the decision-maker observes (J, I, X_I) and updates their beliefs. Based on these updated beliefs, the decision-maker chooses a to maximize the expectation of $u^{\text{d-m}}$, and then utilities are realized.

Proposition 4. *Consider the decision rule $a(s(X_I), t(X_I)) = \mathbf{1}(s(X_I) \geq \underline{s})$, and an analyst response $I(X)$ such that $s(X_{I(X)}) = |I(X)| = \underline{s} \cdot \mathbf{1}(s(X) \geq \underline{s})$.*

- *This decision rule $a(\cdot)$, and any analyst response of this form, are sustained by a perfect Bayesian equilibrium if and only if $\underline{s} \in [\underline{s}^{min}(\bar{n}), \bar{n}^{max}]$.*

- *Among these equilibria, the highest decision-maker expected utility is achieved* *when* $\underline{s} = \min(\underline{s}^{opt}(\bar{n}), \bar{n}^{max})$*, and the highest analyst expected utility is achieved when* $\underline{s} = \underline{s}^{min}(\bar{n})$*.*

The proof of Proposition 4 immediately extends to the case of pre-registration, as shown by the following corollary:

Corollary 3. *Consider the decision rule* $a(J, I, X_I) = \mathbf{1}(|J| = n \text{ and } s(X_{I\cap J}) \geq \underline{s})$*, for some* $n \leq \bar{n}$*, and an analyst response* $J, I(X)$ *such that* $|J| = n$*,* $I(X) \subseteq J$*,* $s(X_{I(X)}) = |I(X)| = \underline{s} \cdot \mathbf{1}(s(X_J) \geq \underline{s})$*. This decision rule* $a(\cdot)$*, and any analyst response of this form, are sustained by a perfect Bayesian equilibrium if and only if* $\underline{s} \in [\underline{s}^{min}(n), \bar{n}^{max}]$*.*

One point worth emphasizing about both these results is that perfect Bayesian equilibrium does not constrain decision-maker beliefs conditional on off-equilibrium (zero probability) reports by the analyst. Pessimistic off-equilibrium beliefs, leading to $a = 0$, can therefore sustain many different decision-maker response functions, and the corresponding equlibria. A natural refinement of perfect Bayesian equilibrium in our setting would require consistency of beliefs with verifiable knowledge, based on the reported X_I. The most pessimistic off-equlibrium belief consistent with this refinement implies that $s(X) = s(X_I)$, so that $a = 1$ whenever $s(X_I) \geq \underline{s}^{opt}(\bar{n})$. This refinement does not substantively affect either result as long as $\underline{s} \leq \underline{s}^{opt}(\bar{n})$ (for Proposition 4) or $\bar{n}^{max} < \underline{s}^{opt}(\bar{n})$ (for Corollary 3): Under these conditions, the additional constraints imposed by this refinement already hold for all the proposed equilibria.

Summary and implications Our baseline model assumes commitment of the decision-maker to a decision rule $a(\cdot)$. We believe that this is a reasonable assumption for statistical methodology in academic research, where strong norms constrain acceptable reporting behavior. We also believe that this is an accurate description of settings such as the drug approval process, where legal rules constrain the medical authority's approval decisions.

Proposition 4 shows that our conclusions are, however, also robust to settings where the decision-maker cannot credibly commit. The optimal symmetric decision rules derived in Proposition 1 are sustained by a perfect Bayesian equilibrium absent commitment. Furthermore, they are optimal for the decision-maker among all symmetric rules which can arise in equilibrium. Analogous statements hold for the case with pre-registration, as in Corollary 3.

4.4 Unknown set of components and analyst private information

In the model considered thus far, there was no reason for the analyst, rather than the decision-maker, to be the party choosing the PAP J. Furthermore, it was assumed that the set of available components i is known to both analyst and decision-maker ex-ante.

We next discuss two extensions of the baseline model where there is additional hidden information. In the first extension, the number of available components is common knowledge, but the analyst has private information about the *identity* of the available components at the time of choosing a PAP. This version of the model rationalizes the role of the analyst in determining the PAP J, while leaving all other conclusions unchanged. In the second extension, considered in Section 4.5, neither analyst nor decision-maker know the available *number* of components $\bar{n}$ at the time of committing to a PAP, and the decision-maker does not know $\bar{n}$ at the time of choosing a.

Assumptions Suppose that the assumptions of the baseline model hold, except that availability of components i is determined by a vector $W \in \{0,1\}^{\bar{n}}$. After choosing J, the analyst observes the vector $X' = (W_1 X_1, \ldots, W_{\bar{n}} X_{\bar{n}})$, and chooses a set I of coordinates of X' to report. The decision-maker's decision rule is of the form $a(J, I, X'_I)$.

The number of available components, $\bar{n}' = |W|$, is common knowledge of the decision-maker and the analyst. The analyst additionally knows W ex-ante, while the decision-maker does not. The decision-maker's prior over W given $\bar{n}'$ is uniform over all permutations of the components i, i.e., $P(W = w|\theta, X) = \mathbf{1}(|w| = \bar{n}')/\binom{\bar{n}}{\bar{n}'}$.

The first-best rule Denote $s(X') = \sum_i W_i X_i$. The decision-maker's first-best decision rule, if they could observe the full vector X', is given by

$$\bar{a}(X') = \mathbf{1}\left(\frac{\alpha + s(X')}{\alpha + \beta + \bar{n}'} \geq \underline{\theta}\right). \tag{23}$$

As in the baseline model, this rule is only implementable if $\underline{s}^{opt}(\bar{n}') \leq \bar{n}^{max}$, where $\underline{s}^{opt}(\cdot)$ is defined as before.

Symmetric cutoff rules and PAPs Suppose that the decision-maker applies a symmetric cutoff rule of the form

$$a(J, I, X'_I) = \mathbf{1}(s(X'_I) \geq \underline{s}).$$

From the decision-maker's perspective, this setting looks just like the baseline model for $\bar{n}'$ components (rather than $\bar{n}$). As in the baseline model, the analyst will report a subset of components I such that $|I| = s(X'_I) = \underline{s}$ iff $s(X') \geq \underline{s}$ and $\bar{n}^{max} \geq \underline{s}$; otherwise the analyst will report $I = \emptyset$.

Symmetric reduced form cutoff rules of the form $\bar{a}(X') = \mathbf{1}(s(X') \geq \underline{s})$ are therefore implementable for $\bar{n}^{max} \geq \underline{s}$. Decision-maker expected welfare for such implementable rules is given by

$$E[(\theta - \underline{\theta}) \cdot \mathbf{1}(s(X') \geq \underline{s})].$$

This has exactly the same form as in our baseline model, with $s(X')$ taking the role of $s(X)$, and $\bar{n}'$ taking the role of $\bar{n}$. All results from the baseline model thus carry over. In particular, Proposition 1 and Proposition 2, where we derived the optimal symmetric rule and the optimal PAP when full pre-specification is required, apply verbatim, after replacing $\bar{n}$ by $\bar{n}'$. The optimal PAP, to be chosen by the analyst, has size $|J|$ such that decision-maker utility is maximized, which might be less than $\bar{n}'$ whenever the latter is large, or when the cost c is large.

Summary and implications The structure of pre-analysis plans from the baseline model remains intact when the analyst has private information. However, while in the baseline model the decision-maker could have chosen the pre-analysis plan, now it needs to be the analyst who has to pick the PAP J. Expected welfare would be strictly lower if the decision-maker were to select J, since the PAP reflects which components are (more) informative about the hypothesis of interest. The variables W_i might be interpreted as a stylized description of private information available to the analyst, regarding either the validity of alternative identification approaches i, or the expected standard errors of alternative estimators i.

4.5 Unknown number of components

In the previous subsection we considered a variant of our model where the analyst has private information about the *identity* of informative components. Now we consider a variant where they have private information about the *number* of informative components, with very different implications. In particular we find that such private information about $\bar{n}$ can rationalize PAPs even in the absence of a communication cost c.

Assumptions Consider the same model as outlined in Section 2, except that now the timeline is as follows. Neither decision-maker nor analyst know the available number of components $\bar{n}$ before the analyst commits to a PAP. Before observing any data or $\bar{n}$, the analyst reports a PAP, that is, a set $J \subseteq \{1, 2, \ldots\}$ of components. The analyst then observes $\bar{n}$ and X, chooses $I = I(X)$, and reports (I, X_I). Then the decision-maker observes (J, I, X_I), but not $\bar{n}$, and updates their beliefs, about both $\bar{n}$ and θ. Based on these updated beliefs, the decision-maker chooses a to maximize the expectation of $u^{\text{d-m}}$, and then utilities are realized.

Assume that $\theta \sim \text{Beta}(\alpha, \beta)$, $X_i|\theta \overset{iid}{\sim} \text{Ber}(\theta)$ and that $\bar{n} \perp \theta, X_1, X_2, \ldots$. Then

$$
\begin{aligned}
P(s(X) = s|\bar{n}) &= \binom{\bar{n}}{s} \frac{B(\alpha + s, \beta + \bar{n} - s)}{B(\alpha, \beta)}, \\
P(\bar{n}|s(X)) &= P(s(X)|\bar{n}) \cdot P(\bar{n})/P(S), \text{ and} \\
E[\theta|s(X)] &= E[E[\theta|s(X), \bar{n}]|s(X)] = E\left[\frac{\alpha + s(X)}{\alpha + \beta + \bar{n}}\middle| s(X)\right] \\
&= (\alpha + s(X)) \cdot E\left[\frac{1}{\alpha + \beta + \bar{n}}\middle| s(X)\right]. \qquad (24)
\end{aligned}
$$

The last expectation averages over the distribution $P(\bar{n}|s(X))$. This expectation is in general decreasing in $s(X)$. Because of this effect, reducing the prior dispersion of the number of components available to the analyst, by means of a PAP, can increase decision-maker welfare, by making $s(X)$ a more informative signals about θ, relative to the unrestricted case.

Symmetric cutoff rules and PAPs The following Proposition 5 provides an example with private information about $\bar{n}$, where a PAP increases decision-maker welfare, even when the cost of communication c equals 0. This example requires that there is sufficient prior dispersion of $\bar{n}$.

Proposition 5. *Suppose that $\bar{n}$ has two points of support $\bar{n}_1 < \bar{n}_2$ which it takes with equal probability, and that $c = 0$. Suppose that the decision-maker uses a cutoff rule of the form $a(I, X_I) = \mathbf{1}(s(X_I) \geq \underline{s})$. Then there exist values for $\bar{n}_1, \bar{n}_2$ such that $E[u^{\text{d-m}}]$ is strictly larger with a PAP restricting the analyst to only the first $\bar{n}_1$ components than with no PAP.*

Summary and implications When the number of components is unknown and we restrict decision rules to simple cutoff rules, then there is an additional role for pre-analysis plans, which is not driven by the cost of communication c. Specifically, a PAP can increase welfare by restricting the analyst to only reporting components that are available with high probability. If there is no such PAP, then the decision-maker would have to discount the report by the analyst to account for the possibility of a higher number of components being available, even in the state of the world where this possibility is not realized.

4.6 Decision-maker bears the cost of communication

In this section, we consider a variation of our baseline model where the decision-maker, rather than the analyst, bears the cost of communication, $c \cdot |I|$. Implementability becomes trivial in this model variation, so that we can focus on the decision-maker's decision problem. Nevertheless, the qualitative conclusions from the baseline model again carry over.

Assumptions Consider the model introduced in Section 2, except that the analyst and decision-maker's objectives are given by

$$\begin{aligned} u^{\text{an}} &= a, \\ u^{\text{d-m}} &= a \cdot (\theta - \underline{\theta}) - c \cdot |I|. \end{aligned} \tag{25}$$

Relative to the model of Section 2, this shifts the cost $c \cdot |I|$ from the analyst to the decision-maker.

Symmetric cutoff rules and PAPs Consider a symmetric cutoff rule of the form $a = \mathbf{1}(s(X_I) \geq \underline{s})$. We assume (as a matter of tie-breaking) that the analyst will report $s(X_I) = \underline{s}$ successes if and only if $s(X) \geq \underline{s}$.[7] That is, we assume that they report no more components than necessary for acceptance. Using the law of iterated expectations, conditioning on $s(X)$, the decision-maker's expected utility is then given by

$$E[u^{\text{d-m}}] = P(s(X) \geq \underline{s}) \cdot \left(E[\theta | s(X) \geq \underline{s}] - \underline{\theta} - c \cdot \underline{s}\right).$$

The optimal cutoff $\underline{s} = \underline{s}^{dc}(\bar{n})$ maximizes this expected utility. Using this optimal cutoff, expected decision-maker utility is, again, not monotonic in $\bar{n}$; cf. Figure 5 below for the uniform prior case. As a consequence, PAPs can again improve welfare by allowing the decision-maker to reduce $\bar{n}$ down to an interior optimum.

Uniform prior Assume now that the prior for θ is uniform, so that $\alpha = \beta = 1$. Using the expressions derived in Section 3 for this case, we can write $E[u^{\text{d-m}}]$ (for general $\underline{s}$) more explicitly as

$$E[u^{\text{d-m}}] = \frac{\bar{n} + 1 - \underline{s}}{\bar{n} + 1} \cdot \left(\frac{1}{2 + \bar{n}} \cdot \frac{2 + \underline{s} + \bar{n}}{2} - \underline{\theta} - c \cdot \underline{s} \right),$$

and the marginal return to decreasing the cutoff from $\underline{s} + 1$ to $\underline{s}$ is given by

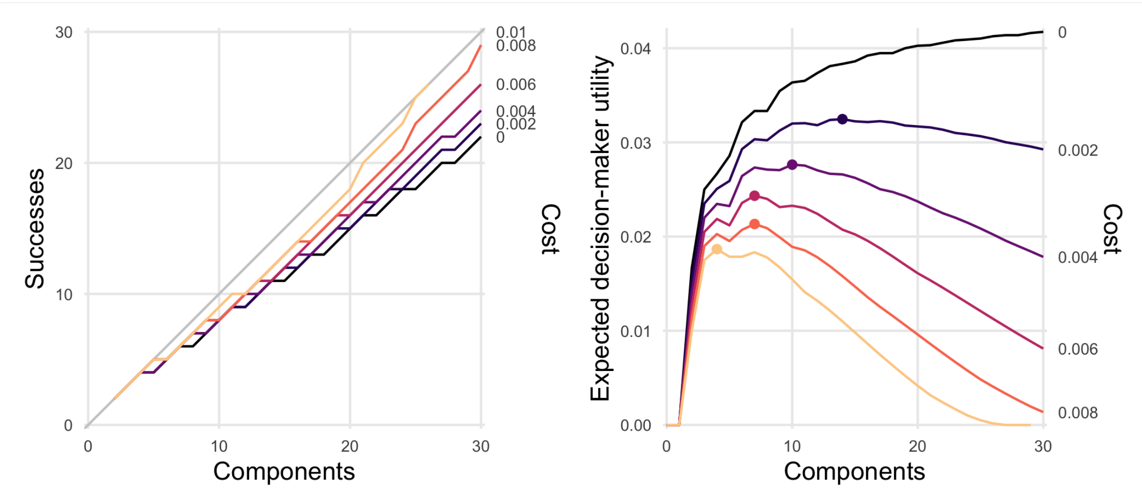
$$\begin{aligned} &P(s(X) = \underline{s}) \cdot (E[\theta | s(X) = \underline{s}] - \underline{\theta} - c \cdot (\underline{s} + 1)) + P(s(X) \geq \underline{s}) \cdot c \\ &= \frac{1}{\bar{n} + 1} \cdot \left(\frac{1 + \underline{s}}{2 + \bar{n}} - \underline{\theta} - c \cdot (\underline{s} + 1) \right) + \frac{\bar{n} + 1 - \underline{s}}{\bar{n} + 1} \cdot c. \end{aligned}$$

The optimal cutoff $\underline{s} = \underline{s}^{dc}(\bar{n})$ is given by

$$\begin{aligned} \underline{s}^{dc}(\bar{n}) &= \min \left\{ \underline{s} : \ E[\theta \middle| s(X) = \underline{s}] \geq \underline{\theta} + c \left(\underline{s} + 1 - \frac{P(s(X) \geq \underline{s})}{P(s(X) = \underline{s})} \right) \right\} \\ &= \left\lceil \frac{(\underline{\theta} - c \cdot \bar{n}) \cdot (2 + \bar{n}) - 1}{1 - 2 \cdot c \cdot (2 + \bar{n})} \right\rceil. \end{aligned} \tag{26}$$

[7] A model variation where the analyst pays an arbitrarily small fraction of the cost $c \cdot |I|$ would ensure this.

Figure 5: Decision-maker bears the cost, symmetric cutoff rules and uniform prior



Notes: This figure plots the solution described in Equation 26, for the case where $\underline{\theta} = .7$, and different values of c. The figure on the left shows the optimal threshold $\underline{s}^{dc}(\bar{n})$ of the decision-maker, taking into account the cost c, for different values of c. For the case $c = 0$, we obtain the first-best optimal cutoff $\underline{s}^{opt}(\bar{n})$ from the baseline model.
The figure on the right plots the decision-maker's expected utility, as a function of $\bar{n}$, for the same values of c. Dots mark out the respective maximizers and maxima of $u^{\text{d-m}}$. The optimal PAP reduces $\bar{n}$ to these maximizers.

Relative to $\underline{s}^{opt}(\bar{n})$ in our baseline model, this definition of $\underline{s}^{dc}(\bar{n})$ replaces $\underline{\theta}$ by $\underline{\theta} + c\left(\underline{s} + 1 - \frac{P(s(X) \geq \underline{s})}{P(s(X) = \underline{s})}\right)$. To see that the condition Equation 26 based on the marginal return is sufficient for an optimum, note that $E[\theta | s(X) = \underline{s}]$ is increasing in $\underline{s}$, while $\underline{s} + 1 - \frac{P(s(X) \geq \underline{s})}{P(s(X) = \underline{s})}$ is decreasing in this specific case. For $c = 0$, we recover $\underline{s}^{opt}(\bar{n})$ from the baseline model for the case of a uniform prior.

Summary and implications When the decision-maker bears the cost of communication, the structure of the game and its solution remains similar. The decision-maker will still ask the analyst to submit only successes, but not more than required for acceptance. For large values of $\bar{n}$, or large cost c, a PAP can again improve decision-maker utility.

References

Abrams, E., Libgober, J., and List, J. A. (2021). Research registries and the credibility crisis: An empirical and theoretical investigation.

Andrews, I. and Shapiro, J. M. (2021). A model of scientific communication. *Econometrica*, 89(5):2117–2142.

Banerjee, A. V., Chassang, S., Montero, S., and Snowberg, E. (2020). A theory of experimenters: Robustness, randomization, and balance. *American Economic Review*, 110(4):1206–1230.

Chassang, S., Padró I Miquel, G., and Snowberg, E. (2012). Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments. *The American Economic Review*, 102(4):1279–1309.

Coffman, L. C. and Niederle, M. (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives*, 29(3):81–98.

Di Tillio, A., Ottaviani, M., and Sørensen, P. N. (2017). Persuasion bias in science: Can economics help? *Economic Journal*, 127(605):266–304.

Di Tillio, A., Ottaviani, M., and Sørensen, P. N. (2021). Strategic sample selection. *Econometrica*, 89(2):911–953.

Duflo, E., Banerjee, A. V., Finkelstein, A., Katz, L. F., Olken, B., and Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics. *NBER Working Paper*, (w26993).

Food and Drug Administration (1998). Guidance for industry: Statistical principles for clinical trials. *US Department of Health and Human Services.*

Frankel, A. and Kasy, M. (2022). Which findings should be published? *American Economic Journal: Microeconomics*, 14(1):1–38.

Gao, Y. (2022). Inference from selectively disclosed data. *arXiv preprint arXiv:2204.07191.*

Gelman, A. (2017). How is preregistration like random sampling and controlled experimentation. `https://statmodeling.stat.columbia.edu/2017/03/09/preregistration-like-random-sampling-controlled-experimentation/`.

Gentzkow, M. and Kamenica, E. (2014). Costly persuasion. *American Economic Review*, 104(5):457–62.

Glaeser, E. L. (2006). Researcher Incentives and Empirical Methods. Technical Report t0329, National Bureau of Economic Research.

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

Henry, E. and Ottaviani, M. (2019). Research and the Approval Process: The Organization of Persuasion. *American Economic Review*, 109(3):911–955.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med*, 2(8).

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Leamer, E. E. (1974). False Models and Post-Data Model Construction. *Journal of the American Statistical Association*, 69(345):122–131.

Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

Libgober, J. (2020). False Positives and Transparency.

Mathis, J. (2008). Full revelation of information in sender–receiver games of persuasion. *Journal of Economic Theory*, 143(1):571–584.

McCloskey, A. and Michaillat, P. (2020). Incentive-Compatible Critical Values. Technical Report 2005.04141.

Miguel, E. (2021). Evidence on research transparency in economics. *Journal of Economic Perspectives*, 35(3):193–214.

Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3):61–80.

Spiess, J. (2018). Optimal estimation when researcher and social preferences are misaligned.

Sterling, T. D. (1959). Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance–Or Vice Versa. *Journal of the American Statistical Association*, 54(285):30–34.

Tetenov, A. (2016). An economic theory of statistical testing. Technical Report CWP50/16, cemmap working paper.

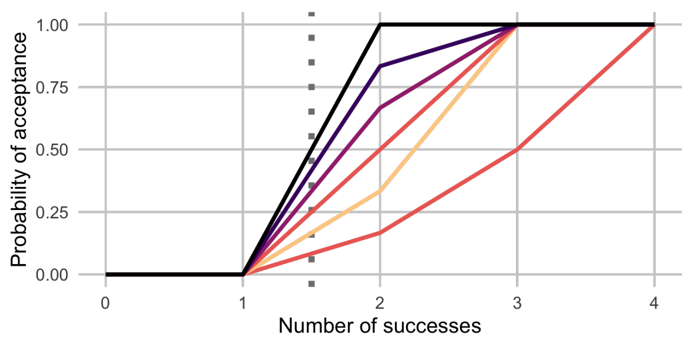
Tullock, G. (1959). Publication Decisions and Tests of Significance—A Comment. *Journal of the American Statistical Association*, 54(287):593–593.

Viviano, D., Wuthrich, K., and Niehaus, P. (2021). (When) should you adjust inferences for multiple hypothesis testing?

Williams, C. (2021). Preregistration and Incentives.

Yoder, N. (2020). Designing Incentives for Heterogeneous Researchers.

Figure 6: Greedy solution for $\bar{n} = 4$



Notes: The lines plot the conditional acceptance probability $P(a(X) = 1|s(X) = s)$ as a function of s, for each of the steps of the greedy algorithm, starting from the bottom. Brighter hues correspond to higher decision-maker utility; the best decision rule corresponds to the second step of the algorithm.

A Further numerical examples

In this appendix, we illustrate some further subtleties of our analysis in the context of several numerical examples.

A.1 The greedy algorithm

To illustrate the structure of optimal decision rules in the general case of our baseline model, as discussed in Section 3.3, let us consider some numerical examples. Suppose that $\bar{n} = 4$, $\bar{n}^{max} = 2$, and $\underline{\theta} = 0.6$. The case $\bar{n} = 4$ is the minimal one where the complications discussed below arise. This example illustrates that the optimal solution is not always given by a symmetric cutoff rule for a subset of components, as in Proposition 2.

For any reduced-form decision rule $\bar{a}(x)$, we can plot $P(a(X) = 1|s(X) = s)$; this function summarizes all the information relevant for decision-maker expected welfare. Figure 6 does so for a sequence of implementable decision rules $\bar{a}(x)$. The bottom line corresponds to the minimal implementable rule, $\bar{a}(X) = \mathbf{1}(X_1 = X_2 = 1)$. The top line corresponds to the cutoff rule $\bar{a}(X) = \mathbf{1}(s(X) \geq 2)$. The intermediate lines correspond to steps of a greedy algorithm for finding the utility-maximizing implementable rule.

As can be seen from Figure 6, in this example the best step of the greedy algorithm is the second step, which corresponds to the decision rule

$$\bar{a}(X) = \mathbf{1}(X_1 = X_2 = 1 \text{ or } X_3 = X_4 = 1).$$

The expected decision-maker utility for the minimal implementable rule, $\bar{a}(X) = \mathbf{1}(X_1 = X_2 = 1)$ (the first step of the greedy algorithm, which can be implemented using a PAP), is equal to 0.05. The expected decision-maker utility for the minimal cutoff rule, $\bar{a}(X) = \mathbf{1}(s(X_I) \geq 2)$ (the last step of the greedy algorithm), is equal to 0.04. The expected decision-maker utility for the second step of the greedy algorithm, $\bar{a}(X) = \mathbf{1}(X_1 = X_2 = 1 \text{ or } X_3 = X_4 = 1)$, is equal to $0.05\dot{3}$.

The first-best rule is $a(X) = \mathbf{1}(s(X_I) \geq 3)$ in this example, since $\underline{s}^{opt} = 3$. However, this first-best rule is not feasible with $\bar{n}^{max} = 2$; we cannot avoid including realizations with only two successes in any non-empty acceptance set. The optimal solution therefore trades off adding additional (utility-increasing) realizations with three successes against adding additional (utility-decreasing) realizations with only two. In this case, the optimal solution is reached by the greedy algorithm after two steps, at which point all realizations with positive utility are accepted. Adding more elements to the acceptance-set cannot be optimal since it only adds realizations with negative utility.

Suppose now instead that $\bar{n} = 4$, $\bar{n}^{max} = 1$, and $\underline{\theta} = 0.2$. In this case, $\underline{s}^{opt}(\bar{n}) = 1$, and the first-best cut-off rule $\bar{a}(X) = \mathbf{1}(s(X) \geq 1)$ is actually implementable. The greedy algorithm, however, does not find this rule. Instead, it chooses the rule that always accepts $\bar{a}(X) \equiv 1$ as its solution. This example illustrates that the greedy algorithm is not guaranteed to find the global optimum.

A.2 PAPs with randomized tests

Allowing for randomized rules may increase the space of implementable decision rules, going beyond our characterization in Section 3. In particular, in the testing example of Section 4.1, non-trivial symmetrical randomized cutoff rules may be available even for $\underline{s}^{test}(\bar{n}) > \bar{n}^{max}$. However, pre-analysis plans with symmetrical cutoff rules can still increase power even when decision rules can include randomization.

To illustrate this point in a simple example, consider the case $\bar{n} = 2$, and assume that $\underline{s}^{test}(\bar{n}) = 1.5$, $\bar{n}^{max} = 1$, and $\underline{\theta} < .5$. In this case, there is no non-trivial non-randomized symmetrical decision rule on the full sample that ensures size control. However, provided that $1/2 < c < \frac{1}{2-\underline{\theta}}$, the symmetric randomized rule

$$a(I, X_I) \sim \text{Ber}\left(\frac{1}{2-\underline{\theta}} \cdot \mathbf{1}(s(X_I) \geq 1)\right)$$

is implementable and has (non-trivial) power $\theta \cdot \frac{2-\theta}{2-\underline{\theta}}$, which exceeds $\underline{\theta}$ iff $\theta > \underline{\theta}$. At the same time, the PAP that restricts the analyst to reporting the first component only, has first-best power θ, and improves over any implementable symmetric rule on the full sample.

B Proofs

Section 3

Proof of Lemma 1. Consider a decision rule of the form $a(s(X_I), t(X_I))$. We show that the corresponding reduced form rule $\bar{a}(\cdot)$ takes the claimed form. Let $I(\cdot)$ be a best response of the researcher to the decision rule $a(\cdot)$, and let $\bar{a}(\cdot)$ be the corresponding reduced-form rule. Let x be such that $\bar{a}(x) = 1$. Then $|I(x)| \leq \bar{n}^{max}$ – otherwise the researcher would prefer reporting $I = \emptyset$. Denote $s = s(x_{I(x)})$ and $t = t(x_{I(x)})$. We get $s + t \leq \bar{n}^{max}$ and $a(s,t) = 1$. It follows that $\bar{a}(x') = 1$ for all x' such that $s(x') \in [s, \bar{n} - t]$. This holds because for any such x', there exists a report I with $|I| \leq \bar{n}^{max}$ for which $s(x'_I) = s$ and $t(x'_I) = t$, and thus $a = 1$. Since $c^{-1} \notin \mathbb{N}$, the researcher strictly prefers acceptance to non-acceptance in this case. Since $s + t \leq \bar{n}^{max}$ the interval $[s, \bar{n} - t]$ has length at least $\bar{n} - \bar{n}^{max}$, so we can take the union over such intervals to obtain the claim.

Consider now reversely a reduced form decision rule $\bar{a}(x)$ of the form described in the statement of the lemma. It remains to show that there is some decision rule $a(s,t)$ which implements $\bar{a}(x)$. It is without loss of generality to assume that $\mathcal{S}$ is a union of intervals of length exactly equal to $\bar{n} - \bar{n}^{max}$. Let s and $\bar{n} - t$ be the upper and lower bounds of such an interval, so that $s + t = \bar{n}^{max}$. Choose $a(s,t) = 1$ for all points (s,t) corresponding to one of the intervals constituting $\mathcal{S}$, and $a(s,t) = 0$ for all other points. It is easily verified that $a(\cdot)$ implements $\bar{a}(\cdot)$. □

Proof of Proposition 1. By Lemma 1, we can exactly implement the reduced-form decision rules

$$\bar{a}(X) = \mathbf{1}(S \in \mathcal{S})$$

with each interval in $\mathcal{S}$ at least of length $\bar{n} - \bar{n}^{PC}$, where $S = s(X)$. For such an $\bar{a}$, the expected utility of the journal is

$$\begin{aligned}
E[\bar{a}(X)(\theta - \underline{\theta})] &= E[\mathbf{1}(S \in \mathcal{S})(E[\theta|S] - \underline{\theta})] \\
&= \sum_{s=0}^{\bar{n}} \mathbf{1}(s \in \mathcal{S}) P(S = s) \left(E[\theta|S = s] - \underline{\theta}\right) \\
&= \sum_{s=0}^{\underline{s}^{opt}(\bar{n})-1} \mathbf{1}(s \in \mathcal{S}) P(S = s) \overbrace{\left(E[\theta|S = s] - \underline{\theta}\right)}^{<0} \\
&\quad + \sum_{s=\underline{s}^{opt}(\bar{n})}^{\bar{n}} \mathbf{1}(s \in \mathcal{S}) P(S = s) \underbrace{\left(E[\theta|S = s] - \underline{\theta}\right)}_{\geq 0}.
\end{aligned}$$

Now suppose that $\mathcal{S}$ is optimal. It follows, first, that there is no interval in $\mathcal{S}$ that is fully included in $[0, \underline{s}^{opt}(\bar{n}))$, since dropping such an interval can only increase utility.

As a consequence, whenever $\mathcal{S} \neq \emptyset$, there must be some $s \in \mathcal{S}$ such that $s \geq \underline{s}^{opt}(\bar{n})$. Second, $\mathcal{S}$ will include all s with $\underline{s}^{opt}(\bar{n}) \leq s \leq \bar{n}$ or be empty. To see this, suppose that there are two s, s' with $\underline{s}^{opt}(\bar{n}) \leq s, s' \leq \bar{n}$, $|s - s'| = 1$, $s \in \mathcal{S}$ but $s' \notin \mathcal{S}$. If that is the case, then we can extend the interval that includes s to also include s' and obtain higher utility. Taken together, these claims imply that $\mathcal{S} = \emptyset$ or $\mathcal{S} = [\underline{s}, \bar{n}]$ with $\underline{s} \leq \bar{n}^{max}$, where $\underline{s}$ maximizes

$$E[\bar{a}(X)(\theta - \underline{\theta})] = \sum_{s=\underline{s}}^{\bar{n}} P(S = s) \underbrace{(E[\theta|S = s] - \underline{\theta})}_{<0 \Leftrightarrow s<\underline{s}^{opt}(\bar{n})}.$$

This, combined with the monotonicity of $E[\theta|S = s]$ in s, yields

$$\mathcal{S} = \begin{cases} \emptyset, & \underline{s}^{min} > \bar{n}^{max}, \\ [\min(\bar{n}^{max}, \underline{s}^{opt}(\bar{n})), \bar{n}], & \text{otherwise}, \end{cases}$$

as claimed. □

Proof of Corollary 1. Immediate from the identities for the uniform distribution described in Section 3.2, and Proposition 1. □

Proof of Corollary 2.

1. This holds trivially, given our assumption that $E[\theta] < \underline{\theta}$.

2. Proposition 1 implies that whenever $\underline{s}^{opt}(\bar{n}) \leq \bar{n}^{max}$, the reduced form decision rule $\bar{a}(X) = \mathbf{1}(S \geq \underline{s}^{opt}(\bar{n}))$ implements the first-best acceptance decision. The claim then follows from the fact that $X_{\{1,\ldots,\bar{n}+1\}}$ is Blackwell more informative than $X_{\{1,\ldots,\bar{n}\}}$.

3. Proposition 1 implies that whenever $\underline{s}^{min}(\bar{n}) \leq \min(\bar{n}, \bar{n}^{max})$, $E[\bar{a}(X)] > 0$ and $E[u^{\text{d-m}}] > 0$. This holds for some values of $\bar{n}$, under the given assumption.

4. Again by Proposition 1, if $\bar{n}$ is such that $\bar{n}^{max} < \underline{s}^{min}(\bar{n})$, then $E[u^{\text{d-m}}] = 0$. Since $E[\theta \big| S \geq \underline{s}]$ is decreasing in $\bar{n}$ (for given $\underline{s}$), and since $\lim_{\bar{n}\to\infty} E[\theta \big| S \geq \underline{s}] = E[\theta] < \underline{\theta}$ (where the latter holds by assumption), it follows that $E[u^{\text{d-m}}] = 0$ for $\bar{n}$ large enough. □

Proof of Proposition 2. Take first J as given, i.e., not a choice variable for either the researcher or the journal. Conditional on J, the result of Proposition 1 applies. In particular, the optimal implementable decision rule that is symmetric given J (as required by Proposition 2) can be implemented by $a(s(X_{I\cap J}), t(X_{I\cap J})) = \mathbf{1}(s(X_{I\cap J}) \geq \min(\underline{s}^{opt}(|J|), \bar{n}^{max}))$ if $\bar{n}^{max} < \underline{s}^{min}(|J|)$, and otherwise by $a \equiv 0$.

Consider now the choice of J. The journal can effectively choose J for the researcher, as long as expected researcher utility is non-negative, by setting $a(J', I, X_I) = 0$ for $J' \neq J$. Expected journal utility, given the optimal symmetric rule conditional on J, is a function of $|J|$. By Corollary 2, expected decision-maker utility $E[u^{\text{d-m}}]$ is increasing in $|J|$ as long as $\underline{s}^{opt}(|J|) \leq \bar{n}^{max}$. The claim follows. □

Proof of Lemma 2. Suppose that $\bar{a}(x)$ is implementable. Implementability requires that there exist functions $I(x), a(I, x_I)$ which implement $\bar{a}(x)$, such that $I(x)$ is an incentive compatible reporting function.

Consider now some value of x such that $\bar{a}(x) = 1$. Let $I = I(x)$ and $w = x_{I(x)}$. The participation condition implies $I \leq \bar{n}^{max}$ for all x – otherwise the researcher would always prefer to report $I = \emptyset$. Incentive compatibility then requires that for all x' such that $x'_I = w$, we have $a(I(x'), x'_{I(x')}) = a(I, x'_I) = 1$; otherwise the researcher would have an incentive to report I, x'_I rather than $I(x'), x'_{I(x')}$. We therefore get that

$$C_{I,w} \subseteq \{x : \bar{a}(x) = 1\}.$$

This implies that $\bar{a}(x)$ is of the form $\bar{a}(x) = \mathbf{1}(x \in \cup_k C_{I_k,w_k})$, since we can find such an I and w for any $x \in \{x : \bar{a}(x) = 1\}$.

Lastly, for any I such that $|I| < \bar{n}^{max}$, we can write $C_{I,w}$ as a union of $C_{I',w'}$ with $|I'| = \bar{n}^{max}$, and it follows that all implementable $\bar{a}$ are of the form claimed in the theorem.

Reversely suppose now that $\bar{a}(x)$ is of the form $\bar{a}(x) = \mathbf{1}(x \in \cup_k C_{I_k,w_k})$ with $|I_k| = \bar{n}^{max}$. We have to show that $\bar{a}(x)$ is implementable. To see this, choose an arbitrary mapping $x \to j$ such that $x \in C_{I_k,w_k}$ for all $x \in \cup_k C_{I_k,w_k}$, and define $I(x) = I_k$ for such x, and $I(x) = \emptyset$ for all other x. Let furthermore $a(I, x_I) = \max_k \mathbf{1}(I = I_k, x_{I_k} = w_k)$. We claim that this choice implements $\bar{a}(x)$. Implementation is immediate, given the definition. The same holds for the condition $|I(x)| \leq \bar{n}^{max}$.

Incentive compatibility remains to be shown. For all x such that $\bar{a}(x) = 1$, the researcher has no incentive to deviate, since the size of reporting sets $I(x')$ is the same for all reporting sets that lead to acceptance. For all x such that $\bar{a}(x) = 0$, on the other hand, there exists no I_k, w_k such that $x_{I_k} = w_k$ and thus there is no mis-reporting that could lead to acceptance. The claim follows. □

Section 4

Lemma 3. *$p(\underline{\theta}; \bar{n}, \underline{s}^{test}(\bar{n}))$ is (weakly) monotonically increasing in $\bar{n}$ for $\theta > \underline{\theta}$, and is (weakly) monotonically decreasing in $\bar{n}$ for $\theta < \underline{\theta}$.*

Proof of Lemma 3. Consider the problem of testing the null $\theta \leq \underline{\theta}$ against the alternative $\theta > \underline{\theta}$, based on observation of X_I, where $I = \{1, \ldots, \bar{n}\}$ is non-random. A sufficient statistic for θ is given by $s(X_I)$, which satisfies the monotone likelihood ratio property. Therefore, a uniformly most powerful test of the null is given by Equation 16; cf. Theorem 3.4.1 in Lehmann and Romano (2006). This implies in particular that the alternative test which ignores $X_{\bar{n}}$ and applies (16) based on $s(X_{\{1,\ldots,\bar{n}-1\}})$ has uniformly weakly lower power. The claim for $\theta > \underline{\theta}$ follows. The claim for $\theta \leq \underline{\theta}$ can be shown by switching the role of null and alternative hypotheses. □

Proof of Proposition 3. The **first claim** immediately follows from our discussion of the first-best rule in Section 4.1.

For the **second claim**, note that any *uniformly most powerful test* (UMPT) has to control size for $\theta \leq \underline{\theta}$, and has to have maximal power for any given $\theta' > \underline{\theta}$. For any possibly randomized test, denote by $\bar{a}(x)$ the probability of rejecting the null conditional on $X = x$. Denote further

$$\beta(\bar{a}(\cdot), \theta) = \sum_x \theta^{s(x)} (1-\theta)^{\bar{n}-s(x)} \cdot \bar{a}(x)$$

the probability that the test $\bar{a}(\cdot)$ rejects given parameter θ. Fix now some $\theta' > \underline{\theta}$. Any UMPT has to solve $\max_{\bar{a}(\cdot)} \beta(\bar{a}(\cdot), \theta')$ subject to $\beta(\bar{a}(\cdot), \underline{\theta}) \leq \underline{\theta}$, and subject to $\bar{a}(x) \in [0, 1]$ for all x. Equivalently, any UMPT has to maximize the Lagrangian $\mathscr{L}(\bar{a}(\cdot)) = \beta(\bar{a}(\cdot), \theta') - \lambda \beta(\bar{a}(\cdot), \underline{\theta})$, subject to $\bar{a}(x) \in [0, 1]$ for all x, where λ is the multiplier on the size constraint. Writing $\delta(s) = \theta'^s (1-\theta')^{\bar{n}-s} - \lambda \underline{\theta}^s (1-\underline{\theta})^{\bar{n}-s}$, we have

$$\mathscr{L}(\bar{a}(\cdot)) = \sum_{x \in \{0,1\}^{\bar{n}}} \delta(s(x)) \cdot \bar{a}(x),$$

where the sign of $\delta(s)$ is monotonically increasing in s (equivalently, the likelihood of the sufficient statistic $s(X)$ satisfies the monotone likelihood ratio property). This Lagrangian can be maximized separately for each x, which immediately implies that the solution has to satisfy $\bar{a}(x) = 1$ if $\delta(s(x)) > 0$ and $\bar{a}(x) = 0$ if $\delta(s(x)) < 0$. Hence, there is some integer $\underline{s}$ such that $a(x) = 1$ for $s(x) > \underline{s}$ and $a(x) = 0$ for $s(x) < \underline{s}$. For the size constraint to be binding at $\underline{\theta}$, $\beta(\bar{a}(\cdot), \underline{\theta}) = \underline{\theta}$, we must have that $\underline{s} = \lfloor \underline{s}^{test}(\bar{n}) \rfloor$ by the definition of $\underline{s}^{test}$ in Equation 17.

Let us next consider the set of *implementable randomized* tests. Recall that we assume that any randomization device U used by the test is realized and common knowledge before the start of the game. Conditional on U, implementability is as characterized in Lemma 2. The implementable randomized rules are therefore given by the set of convex combinations of non-randomized rules that satisfy the conditions of Lemma 2.

To prove that *no UMPT is implementable*, we need to show that our characterizations of UMPT and of implementable tests are incompatible. Consider any mixture of non-randomized implementable tests $\bar{a}(x; u)$ such that $\bar{a}(x) = E[\bar{a}(x; U)]$. Suppose that $\bar{a}(x)$ is a UMPT. Then $\bar{a}(x) = 0$ for all x such that $s(x) < \lfloor \underline{s}^{test}(\bar{n}) \rfloor$, and $\bar{a}(x) = 1$ for all x such that $s(x) > \lfloor \underline{s}^{test}(\bar{n}) \rfloor$. This can only be true if the same conditions hold almost surely (in U) for $\bar{a}(x; U)$. But no non-randomized test of this form is implementable, by Lemma 2. Therefore, the randomized test $\bar{a}(x)$ is not implementable, which shows the claim.

To show the **third claim**, recall that Lemma 1 implies that for any symmetric test of the form $a(s(X_I), t(X_I))$, $\bar{a}(\cdot)$ is of the form $\bar{a}(X) = \mathbf{1}(s(X) \in \mathscr{S})$, where $\mathscr{S}$ is a union of intervals of length at least $\bar{n} - \bar{n}^{max}$. We can immediately discard the case $\mathscr{S} = \emptyset$, as it does not allow for power strictly bigger than size. The lowest possible size of a non-trivial test is achieved when $\mathscr{S} = [\underline{s}, \underline{s} + \bar{n} - \bar{n}^{max}]$ consists of only one such interval.

We show that no test which rejects for $s(X) \in \mathscr{S}$, where $\mathscr{S}$ is an interval of length $\bar{n} - \bar{n}^{max}$, can control size, under the stated assumptions. The number of successes $S = s(X)$ follows a Binomial distribution conditional on θ. This distribution is unimodal, so that the (conditional) probability mass function of S is monotonic on either side of its mode. This implies that any interior rejection interval of length $\bar{n} - \bar{n}^{max}$ cannot minimize the rejection probability; shifting such an interval by one unit to either the left or the right has to (weakly) reduce the rejection probability, otherwise we would have a contradiction to unimodality. The rejection probability $E[\bar{a}(X)]$ is therefore minimized when either $\mathscr{S} = [\bar{n}^{max}, \bar{n}]$, or $\mathscr{S} = [0, \bar{n} - \bar{n}^{max}]$. The first choice does not control size if $\underline{s}^{test}(\bar{n}) > \bar{n}^{max}$. The second choice leads to a rejection probability that is at least as large as that for the first choice, whenever $\theta \leq \underline{\theta} \leq 0.5$. The claim follows. $\square$

Proof of Proposition 4. Consider first the *decision-maker beliefs.* Bayesian updating given the analyst response implies $P(\theta|X_I, I) = P(\theta|s(X) \geq \underline{s})$ when $s(X_{I(X)}) = \underline{s}$, and $P(\theta|X, I_X) = P(\theta|s(X) < \underline{s})$ when $s(X_{I(X)}) = 0$.

Consider next the *decision-maker best response.* For $s(X_{I(X)}) = 0$, we necessarily have $a = 0$; this follows from $E[\theta|s(X) \leq \underline{s}] < E[\theta] < \underline{\theta}$. For $s(X_{I(X)}) = \underline{s}$, we have $a = 1$ iff $\underline{s} \geq \underline{s}^{min}(\bar{n})$; this follows from the definition of $\underline{s}^{min}(\bar{n})$. For $s(X_{I(X)}) \notin \{0, \underline{s}\}$ we can assume $a = 0$; this follows since we can assume arbitrary off-equilibrium path beliefs.

Let us now turn to the best *action of the analyst*, given X. If $s(X) < \underline{s}$, all possible actions result in $a = 0$. The best possible response of the analyst is $I(X) = \emptyset$. If $s(X) \geq \underline{s}$, the analyst utility maximizing response is such that $s(X_{I(X)}) = \underline{s}$ if $\underline{s} \leq \bar{n}^{max}$, and $I(X) = \emptyset$ otherwise.

The claim of the proposition regarding equilibria follows. The claim regarding decision-maker and analyst utility is immediate. $\square$

Proof of Corollary 3. This proof is based on the same argument as the proof of Proposition 4. Consider first the *decision-maker beliefs.* Bayesian updating given the analyst response implies $P(\theta|X_I, I) = P(\theta|s(X_J) \geq \underline{s})$ when $s(X_{I(X)}) = \underline{s}$, and $P(\theta|X_I, I) = P(\theta|s(X_J) < \underline{s})$ when $s(X_{I(X)}) = 0$.

Consider next the *decision-maker best response.* For $s(X_{I(X)}) = 0$, we necessarily have $a = 0$; this follows from $E[\theta|s(X_J) \leq \underline{s}] < E[\theta] < \underline{\theta}$. For $s(X_{I(X)}) = \underline{s}$, we have $a = 1$ iff $\underline{s} \geq \underline{s}^{min}(n)$ (recall $n = |J|$); this follows from the definition of $\underline{s}^{min}(n)$. For $s(X_{I(X)}) \notin \{0, \underline{s}\}$ we can assume $a = 0$; this follows since we can assume arbitrary off-equilibrium path beliefs.

Let us now turn to the best action of the analyst, given X. If $s(X_J) < \underline{s}$, all possible actions result in $a = 0$. The best possible response of the analyst is $I(X) = \emptyset$.

If $s(X_J) \geq \underline{s}$, the analyst utility maximizing response is such that $s(X_{I(X)}) = \underline{s}$ if $\underline{s} \leq \bar{n}^{max}$, and $I(X) = \emptyset$ otherwise.

Lastly, it is always optimal for the analyst to preregister a set J such that $|J| = n$; this results in non-negative expected utility, while any other choice results in $E[u^{\text{an}}] = 0$.

The claim of the corollary regarding equilibria follows. $\square$

Proof of Proposition 5. We prove the claim by finding $\bar{n}_1, \bar{n}_2$ such that (i) for a PAP that restricts the analyst to only use observations $J = \{1, \ldots, \bar{n}_1\}$, the decision-maker has utility greater than $\frac{1}{2}E[(\theta - \underline{\theta})^+]$, and (ii) without a PAP, for any symmetric cutoff rule, the decision-maker has utility less than $\frac{1}{2}E[(\theta - \underline{\theta})^+]$. From this the claim follows.

To see (i), note that for every $\eta > 0$ there exists some $\bar{n}_1$ such that the PAP that restricts the analyst to only use observations $J = \{1, \ldots, \bar{n}_1\}$ has utility at least $(1-\eta)E[(\theta - \underline{\theta})^+]$. To see this, consider the rule $a(X_I, I) = \mathbf{1}(s(X_{I \cap J}) \geq \underline{\theta} \cdot \bar{n}_1)$. For almost all (fixed) θ, we have that $P(s(X_J)/\bar{n}_1 \geq \underline{\theta}|\theta) \to \mathbf{1}(\theta \geq \underline{\theta})$ by the (weak) law of large numbers, as $\bar{n}_1 \to \infty$. Hence, $E[a(X_I, I)(\theta - \underline{\theta})] = E[P(s(X_J)/\bar{n}_1 \geq \underline{\theta}|\theta)(\theta - \underline{\theta})] \to E[(\theta - \underline{\theta})^+]$, as $\bar{n}_1 \to \infty$, by the dominated convergence theorem.

Let us now show (ii). Consider a symmetric cutoff rule with cutoff $\underline{s}$. For $S = s(X)$ we can write

$$P(S \geq \underline{s}|\theta) = \frac{1}{2}\left(P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_1) + P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_2)\right).$$

For all $\delta, \varepsilon > 0$ and $\bar{n}_1$ there exists some $M \geq 1$ such that, for all $\bar{n}_2$ with $\bar{n}_2 - \bar{n}_1 \geq M$ we have that

1. $P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_1) = 0$ for all $\underline{s} > \bar{n}_1$, and
2. $P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_2) > 1 - \epsilon$ for all $\underline{s} \leq \bar{n}_1$ and $\theta > \delta$.

The first claim is immediate. The second claim follows from Chebyshev's inequality, using $\text{Var}(S/E[S|\theta, \bar{n}_2]|\theta, \bar{n}_2) = \frac{1-\theta}{\theta \bar{n}_2} \leq \frac{1-\delta}{\delta M}$. For such $\delta, \varepsilon, \bar{n}_1$ and any $\bar{n}_2 \geq M + \bar{n}_1$ it then follows from the two claims that

$$\begin{aligned}
P(S \geq \underline{s}|\theta) &= \frac{1}{2}P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_2) \leq \frac{1}{2} && \text{for all } \underline{s} > \bar{n}_1, \\
P(S \geq \underline{s}|\theta) &\geq \frac{1}{2}\left(P(S \geq \underline{s}|\theta, \bar{n} = \bar{n}_1) + 1 - \epsilon\right) \geq \frac{1-\varepsilon}{2} && \text{for all } \underline{s} \leq \bar{n}_1 \text{ and } \theta > \delta.
\end{aligned}$$

In words, either $a = 1$ with probability 1 whenever $\bar{n} = \bar{n}_1$ (this happens for large cutoffs), or $a = 0$ with probability close to 1 whenever $\bar{n} = \bar{n}_2$ (this happens for small cutoffs) – (almost) independently of θ.

Hence, for any cutoff rule of the form $a(S) = \mathbf{1}(S \geq \underline{s})$,

$$\begin{aligned}
E[u^{\text{d-m}}] &= E[a(S)(\theta - \underline{\theta})] = E[(\theta - \underline{\theta}) \cdot \mathbf{1}(S \geq \underline{s})] = \\
&\leq \min \big(E[(\theta - \underline{\theta})^{+} \cdot P(S \geq \underline{s}|\theta)], \\
&\qquad E[(\theta - \underline{\theta})^{+}] - E[(\theta - \underline{\theta})^{-} \cdot P(S \geq \underline{s}|\theta)] \big) \\
&\leq \begin{cases} \frac{1}{2} E[(\theta - \underline{\theta})^{+}] \leq \frac{1}{2} E[(\theta - \underline{\theta})^{+}], & \underline{s} > \bar{n}_1, \\ E[(\theta - \underline{\theta})^{+}] - P(\theta > \delta) E[(\theta - \underline{\theta})^{-} | \theta > \delta] \frac{1-\varepsilon}{2}, & \underline{s} \leq \bar{n}_1. \end{cases}
\end{aligned}$$

Choose now some $\eta \in (0, \frac{1}{2})$ and $\varepsilon, \delta > 0$ with $\frac{1-\varepsilon}{2} P(\theta > \delta) E[(\theta - \underline{\theta})^{-} | \theta > \delta] > \eta \, E[(\theta - \underline{\theta})^{+}]$. For the resulting choices of $\bar{n}_1$ and $\bar{n}_2$ above, the PAP on the restricted set yields strictly better utility than any cutoff rule. □