Designing Incentives for Heterogeneous Researchers

Nathan Yoder

University of Georgia

A principal (e.g., the US government) contracts with a researcher with unknown costs (e.g., a vaccine developer) to conduct a costly experiment. This contracting problem has a novel feature that captures the difference between the form of an experiment and the strength of its results: researchers face a problem of information design rather than optimal effort. Using a novel comparative static for Bayesian persuasion settings, I characterize the optimal contract and show how experimentation is distorted by the need to screen researchers. Moreover, I show that there is no loss from contracting on the experiment's result rather than the experiment itself.

I. Introduction

Many contractual relationships focus on the procurement of information. For instance, a government may subsidize a pharmaceutical firm's

In memory of Bill Sandholm, who provided numerous invaluable comments on early drafts. For their helpful suggestions, I thank Nageeb Ali, Randy Beck, Alex Bloedel, Ben Brooks, Gabriel Carroll, Chris Cornwell, Alex Frankel, Drew Fudenberg, Roozbeh Hosseini, Emir Kamenica, Anton Kolotilin, Natalia Kovrijnykh, Ce Liu, Elliot Lipnowski, Gabriel Martinez-Roa, Konrad Mierendorff, Daehong Min, Antonio Penta, Dan Quint, Daniel Rappoport, Marzena Rostek, Maher Said, Ian Schmutte, Lones Smith, Joel Sobel, Lars Stole, Peter Norman Sørensen, Teck Yong Tan, John Turner, and Marek Weretka; seminar participants at Chicago Booth, Rochester, and Georgia; audience members at European Summer Symposium in Economic Theory 2017 and Southern Economic Association 2019; and anonymous referees. This paper previously circulated under the title "Designing Incentives for Academic Research." A previous version forms the first chapter of my dissertation, "Essays in Microeconomics and Market Design." I gratefully acknowledge the University of Wisconsin–Madison's fellowship support for my dissertation research as well as summer support from the University of Georgia through a Terry-Sanford Research Award. This paper was edited by Emir Kamenica.

Electronically published June 9, 2022

Journal of Political Economy, volume 130, number 8, August 2022.

https://doi.org/10.1086/720072

tests of a novel vaccine's efficacy, or a firm contemplating litigation may hire outside legal counsel to investigate the opposing party's liability and gather evidence for a potential trial. In such settings, the agent that conducts research is likely to be better informed about their costs of information acquisition than the principal who hires them. Consequently, the principal faces a screening problem when designing a contract.

This contracting problem is the focus of this paper. It has an important difference from those classically considered in the context of procurement (e.g., Baron and Myerson 1982; Laffont and Tirole 1986): instead of a monopolistic firm that chooses among effort or output levels, the agent with unknown cost is a researcher that chooses among *experiments*, or signals about the state of the world. It must decide not only how much information to gather but also how to gather it: the pharmaceutical firm must choose not only the size of its clinical trial but also the trial's control group (e.g., active or placebo) and primary endpoint (e.g., infection or severe disease); the attorney serving as outside counsel must choose not only the number of hours to devote to her investigation but also the witnesses she will question and the documents she will seek. Hence, when the principal decides which of these choices it wishes to induce, it is confronted with an information design problem: rather than a deterministic level of precision, it must select a distribution of posterior beliefs.

Moreover, the stochastic nature of experimentation may limit the principal's tools for incentivizing the researcher. Specifically, the principal may be unable to discriminate between experiments ex post when they differ only in their counterfactual outcomes, forcing it to contract on the experiment's result rather than its design.[1] For example, an attorney might conclude her investigation after retrieving just a few documents if they reveal grossly negligent conduct. The results of her investigation would then present the client—and the court, if the client opts to litigate—with strong evidence of liability. But they would not reveal how the attorney would have continued her investigation if the documents had failed to yield any relevant information. As a consequence, any reward for the attorney cannot depend on this counterfactual portion of her methodology.[2]

My analysis proceeds in a model with two states of the world. The principal seeks information about which of these is true either to improve its

[1] Rappoport and Somma (2017) also make this observation—and provide additional examples—in an environment without private information about cost.

[2] For a more abstract example, suppose that a researcher experiments by publicly observing a Wiener process whose drift depends on the binary state of the world, as in Wald (1945). The principal may see that the researcher has stopped experimenting after observing enough evidence in favor of state 1 to increase the probability that their beliefs place on its occurrence to $\bar{p}$. But this observation does not tell the principal the posterior $\underline{p}$ at which the researcher would have ceased experimentation had she instead observed evidence against state 1.

own decisions or to persuade a separate decision maker. To this end, it can contract with the researcher, who can conduct any experiment about the state. These experiments are costly for the researcher, and her marginal cost of a more informative experiment (in the sense of Blackwell 1953) depends on her type, which is drawn from a finite set and is private information.

The researcher's experiment produces *hard information*: its outcome cannot be falsified or concealed by the researcher. This is a key feature of many applications: US law requires that pharmaceutical firms register their clinical trials and report their results in a publicly available database;[3] rules of professional conduct require attorneys to provide their clients with the information necessary to make informed decisions, such as whether to litigate;[4] and if a case proceeds to litigation, an attorney is obligated to disclose evidence in their possession as part of the pretrial discovery process.[5] However, it also rules out some settings: for example, when they can selectively report their findings (as described in, e.g., Simmons et al. 2011), academic researchers do not produce hard information.[6]

[3] See 42 U.S.C. §282(j).

[4] See rule 1.4 in the American Bar Association's Model Code of Professional Conduct or, e.g., the State Bar of Georgia's Rules of Professional Conduct.

[5] In US courts, the rules of discovery require a civil litigant to disclose relevant evidence to the opposing party upon request and evidence that the litigant may rely on in court even without a request. (See Federal Rule of Civil Procedure 26.) Further, attorneys are prohibited from presenting false evidence to the court or concealing evidence from discovery. (See rule 3.4 in the American Bar Association's Model Code.) Hence, it seems reasonable to use a model without falsification or concealment to capture the incentives involved in pretrial information acquisition: litigants are obligated to truthfully disclose unfavorable evidence upon request and are incentivized to present favorable evidence at trial (and thus also disclose it in discovery). This argument is similar to Kamenica and Gentzkow's (2011) in the context of criminal prosecution. However, disclosure requirements are not quite as strong in civil litigation, and so the assumption of hard information may fit less precisely than Kamenica and Gentzkow (2011) show it fits the prosecution setting. In particular, the US Constitution requires prosecutors to turn over exculpatory evidence even without a specific request; see Brady v. Maryland, 373 U.S. 83 (1963). Further, while an attorney's opinion work product (e.g., notes, reports) is protected from disclosure in both civil and criminal cases, including Brady disclosure (Gershman 2006), civil courts at the federal level and in most states also protect other trial preparation materials (e.g., witness statements) that would be subject to disclosure under the Brady rule; see, e.g., Federal Rule of Civil Procedure 26(b)(3). However, this protection "does not protect the litigant from supplying her adversary with the underlying or harmful facts" (Thornburg 1991, 1528); see Thornburg (1991) for further discussion.

[6] More generally, the assumption of hard information rules out three broad classes of applications where information is acquired privately. First, it excludes *cheap talk* settings, where the researcher's result is not verifiable. Second, it excludes applications with *selective disclosure*, such as academic research, where the results of the researcher's experiment can be fully or partially concealed but not falsified. Finally, it excludes applications where the experiment's outcome is verifiable but its meaning depends on the researcher's hidden action, e.g., the design of a proprietary model to assess credit risk. This contrasts with several papers from the literature on contracting for information that consider settings where the researcher's result is unverifiable (Szalay 2005; Zermeño 2011; Carroll 2019; Häfner and Taylor 2019), can be selectively disclosed (Che and Kartik 2009), or has hidden meaning (Chade and Kovrijnykh 2016; Angelucci 2017; Bizzotto et al. 2020).

When the experiment itself is contractible, theorem 1 shows that the optimal contract has two properties specific to the context of information procurement. First, it induces Blackwell-monotone behavior by the various researcher types: lower-cost researchers conduct experiments that not only cost them more than those conducted by higher-cost researchers would but also are more informative in the Blackwell (1953) sense. Second, because there are two states of the world, each experiment that the contract induces has two possible outcomes.

Theorem 2 compares this contract with the one that would be offered in the absence of private information about cost. The familiar effects of hidden information—no distortion at the top and distortion elsewhere—are present but occur in a manner specific to the experimentation context. In particular, the distortion that affects all but the lowest-cost researcher's experiment takes the form of garbling: not only are these experiments less costly for the researcher and less valuable to the principal, they are less useful to any decision maker who might rely on their results.

Contracting on the *result* of the researcher's experiment instead—that is, based on the posterior belief that the experiment induces — eliminates degrees of freedom from the problem of contract design. To see why, suppose that we want a lower-cost researcher's experiment to either increase our belief that the state of the world is 1 (rather than 0) to $\bar{p}$ or decrease it to $\underline{p}$, and we want a higher-cost researcher to conduct an experiment with possible results $\bar{p}'$ and $\underline{p}'$. The rewards we specify for these four results pin down the expected reward for the experiments we want the higher- and lower-cost researchers to conduct. However, they also specify the expected reward for conducting an experiment that shifts our beliefs to either $\underline{p}$ or $\bar{p}'$ as well as one that yields a result of either $\underline{p}'$ or $\bar{p}$. This introduces additional incentive compatibility conditions to the screening problem.

Theorem 3 shows that these constraints are unimportant. Anything that the principal might actually want the researcher to do, it can get her to do by rewarding her on the basis of her results. In particular, suppose that the principal wants the researcher to experiment the same way she would when offered the optimal contract from theorem 1. It can induce this behavior with a results-based contract—and at the same expected cost as contracting on the experiment directly. In fact, this implementation result applies to any researcher behavior with the characteristics (Blackwell monotonicity and binary outcome experiments) described in theorem 1.[7]

[7] Note that theorem 3 differs from the implementation results introduced in concurrent work by Rappoport and Somma (2017): the additional incentive compatibility conditions that results-based contracting creates in the presence of private information about cost do not appear in their symmetric information environment.

The optimal results-based contract is described as follows. On each interval between results (i.e., posterior beliefs) produced by the experiments of two adjacent types, compensation is the sum of two parts. First, the researcher is reimbursed for the portion of her costs that can be attributed to her experiment's realized outcome at the lesser of the rates that those two types would incur them. Then, her transfer is adjusted by an affine function of the result that is constructed so as to make the type with lower costs indifferent between the two types' experiments.

This construction can only successfully implement researcher behavior that is Blackwell monotone. In fact, Blackwell monotonicity is necessary for implementation with any results-based contract: theorem 4 shows that if the behavior desired by the principal does not involve types with lower costs conducting more informative experiments, then there is no results-based contract that induces it.

The key insight behind theorems 1, 2, and 4 is a novel comparative statics result that has implications beyond the setting of this paper. Proposition 1 shows that in a setting with two states of the world, if one information design problem is more concave than another in the additive sense—that is, its objective differs by a concave function of the posterior—it will always have a Blackwell-less informative solution. Consequently, since the principal's contracting problem is separable into type-specific information design problems, which are additively more concave for types with higher costs, it will choose to induce behavior that is Blackwell monotone (theorem 1). Since those problems are additively more concave than the type-specific problems it would face with symmetric information, it instructs each researcher type to conduct a less informative experiment than it would in the symmetric information case (theorem 2). Finally, since any results-based contract will present higher-cost researcher types with additively more concave information design problems, they must conduct less informative experiments (theorem 4).

Related literature.—Within the contracting literature, this paper joins a number of others in which the agent's function is to gather costly information. These articles primarily consider environments without an information asymmetry regarding the agent's characteristics. Instead, most focus on problems of moral hazard: the form of the agent's experiment (Chade and Kovrijnykh 2016; Angelucci 2017; Bizzotto et al. 2020) and, in many cases, its outcome (Szalay 2005; Che and Kartik 2009; Zermeño 2011; Carroll 2019; Häfner and Taylor 2019) are at least partially unverifiable. Of these, the closest to my paper is concurrent work by Rappoport and Somma (2017). They analyze the problem of contracting on results in a setting without researcher heterogeneity. Instead, their focus is on the impact of risk aversion and limited liability.

Fewer papers consider contracts for research in a screening context. The earliest among these is Osband (1989), which considers a model with both screening and moral hazard. I innovate upon his setting by explicitly modeling experimentation and allowing experimenters to choose any experiment possible in principle as opposed to simply choosing a precision level. Closer to this paper is concurrent work by Min (2021), which considers a nontransferable utility setting with two researcher types that differ in the set of experiments available to them. His focus is on the form of the direct mechanism, whereas my paper is motivated in part by settings where the experiment is not contractible, and so the direct mechanism is unavailable. Also relevant is McClellan (2021), who considers dynamic incentives in the closely related Wald (1945) experimentation framework (see Morris and Strack 2019); in his setting, the researcher may have private information about the state of the world rather than her costs.

The literature on screening problems with an imperfectly observable choice variable is also related; Guesnerie et al. (1989) provides a summary. With a one-dimensional choice variable, the equivalence of perfect and noisy observation is known. With a multidimensional choice variable (such as an experiment), the sufficient conditions in the literature are much more stylized and do not accommodate the inability to observe an experiment's counterfactual results.

Finally, my setting is closely related to Bergemann et al. (2018), who consider the problem of selling experiments. In a sense, their problem is the reverse of mine: their researcher is a monopolist who screens the consumers of her experiments according to their private information.[8] Thus, the relationship between our papers is similar to the relationship between Mussa and Rosen (1978) and Baron and Myerson (1982), in that the roles of principal and agent are switched.

II. Environment

There is a state of the world $\omega \in \{0, 1\}$; a *researcher*, who is capable of gathering information about ω and a *principal*, who contracts with the researcher in order to profit from that information. Each agent places the same prior probability $p_0 \in (0, 1)$ on the event $\omega = 1$.

A. The Researcher

The researcher may investigate the state of the world ω by choosing a costly experiment τ to undertake. This experiment produces *hard information*: its outcome is verifiable and thus cannot be falsely reported by the researcher. Instead, it must be truthfully revealed to the principal. I

[8] A key difference lies in the nature of this private information: in Bergemann et al. (2018), consumers have private information about the state, whereas in my model, researchers have private information about their costs.

represent experiments as the distributions of posterior beliefs p they induce.[9] Accordingly, I call an experiment's realized posterior belief its *result* and refer to results $p < p_0$ as *negative results* and $p > p_0$ as *positive results.* All experiments possible in principle are available to the researcher, though their costs may be prohibitive. As Kamenica and Gentzkow (2011) show, this means the researcher chooses from the set of distributions $\tau \in \Delta([0, 1])$ that are *Bayes plausible*: $E_\tau p = p_0$.[10]

The researcher's cost of experimentation depends on how skilled she is at investigating the state of the world.[11] Specifically, her investigative ability is captured by her *type* $\theta \in \Theta \equiv \{\theta_1, \ldots, \theta_N\} \subset (0, \infty)$, where $\theta_1 > \cdots > \theta_N$. Her type is private information; the principal's beliefs about it are given by the distribution F, which has full support on Θ.[12] Smaller types (those with larger subscripts) correspond to lower costs and hence denote higher ability.

Furthermore, the cost of each experiment is *posterior separable*: that is, it is the expected reduction in a *measure of uncertainty*, or concave function of the posterior.[13] To simplify exposition, I also assume that costs are multiplicatively separable in type. Specifically, the cost $C(\tau, \theta)$ to a type θ researcher of undertaking the experiment τ is given by

$$C(\tau, \theta) \equiv E_\tau[\theta c(p)]; \; c(p) \equiv H(p_0) - H(p),$$

where $H : [0, 1] \to \mathbb{R}$ is continuous and strictly concave and is differentiable on (0, 1).

Because of Jensen's inequality, the cost function's form ensures that more informative experiments (in the sense of Blackwell 1953) are more costly to generate. That is, if τ' is a mean-preserving spread of τ, then $C(\tau', \theta) \geq C(\tau, \theta)$. Further, the difference in the costs of τ' and τ is smaller

[9] This is without loss—the informational content of an experiment is fully characterized by the distribution of posteriors it induces.

[10] For a set S, let conv(S) denote its convex hull, and let $\Delta(S)$ denote the set of Borel probability measures on S. For a function v, let Gr$(v) \equiv \{(x, y) \mid y = v(x)\}$ denote its graph. For a Borel probability measure τ, let supp τ denote its support.

[11] Alternatively, ability could enter the model by allowing higher-ability researchers access to a larger set of experiments. This would make it unnecessary for the principal to screen researchers by ability; instead, it could implement the first-best choice function and capture all surplus by offering payment for each experiment equal to the researchers' common cost of conducting it. Since costs are posterior separable, the results-based implementation problem would then become trivial.

[12] This is without loss: types with a prior probability of zero can just be excluded from Θ.

[13] In Gentzkow and Kamenica (2014), the authors alternatively assume that costs are proportional to the expected reduction in a measure of uncertainty H for some agent with prior p_H, which need not be equal to p_0, so that the cost of an experiment does not depend on the prior of the agent who views it. They then show that we can write researcher cost as an expectation with respect to τ of $H(p_H) - \hat{H}(p)$ for appropriate $\hat{H}$. As long as $\hat{H}$ is itself concave, their functional form assumption is equivalent to the one I make here, with the addition of a fixed cost: in this case, one can simply take $\hat{H}$ for the measure of uncertainty.

for better researchers (those with lower θ): $C(\tau', \theta') - C(\tau, \theta') \geq C(\tau', \theta) - C(\tau, \theta)$ for $\theta < \theta'$.

The researcher is risk neutral: conducting the experiment τ and receiving a transfer of t from the principal yields a type θ researcher a payoff of $t - C(\tau, \theta)$.

B. *The Principal*

When the researcher's experiment induces the belief p, the principal obtains indirect utility $w(p)$, where $w : [0, 1] \to \mathbb{R}$ is upper semicontinuous. Hence, the value that the principal receives from the experiment τ is given by $W(\tau) \equiv E_\tau[w(p) - w(p_0)]$. The principal's payoff is quasilinear in transfers to the researcher: when he pays the researcher t for the experiment τ, he receives a payoff of $W(\tau) - t$.

C. *Contracting*

The principal maximizes his expected payoff by making the researcher a take-it-or-leave-it offer of one of two types of contracts.

When the experiment itself is contractible, the principal can offer a *methods-based contract* consisting of a transfer function $T : X \to \mathbb{R}$, where $X \equiv \{\tau \in \Delta([0, 1]) : E_\tau p = p_0\}$ is the set of all experiments. Under this type of contract, when the researcher conducts experiment τ, she is paid $T(\tau)$. The revelation and taxation principles ensure that this class of contracts is equivalent to the class of direct revelation contracts $(\chi : \Theta \to X, t : \Theta \to \mathbb{R})$. This makes it the most general among the classes of contracts I consider.

When the experiment's result is contractible, the principal can offer a *results-based contract* consisting of a transfer function $\psi : [0, 1] \to \mathbb{R}$. Under this type of contract, when the researcher produces p as the result of some experiment, she is paid $\psi(p)$.

A contract *implements* the experiment choice function $\chi : \Theta \to X$ if a type θ researcher who accepts that contract gets a nonnegative expected payoff from conducting the experiment $\chi(\theta)$ that is at least as large as the expected payoff she would get from conducting any other experiment.

D. *Discussion*

Instead of explicitly specifying how the principal benefits from the researcher's experiment, I take the principal's indirect utility function over beliefs as a primitive of the model. Depending on the setting, this indirect utility function might represent the principal's payoff from persuading a decision maker (as in Kamenica and Gentzkow 2011), his expected payoff from optimally choosing an action, a noninstrumental payoff from information (as in, e.g., Ely et al. 2015), or some combination of these.

The literature on information costs shows that posterior-separable cost functions arise from a wide variety of foundations. Examples of these include Shannon (1948) mutual information, Wald (1945) sequential sampling (Morris and Strack 2019), constant marginal cost of experimentation (Pomatto et al. 2020; Bloedel and Zhong 2021), or the instrumental value of information (Frankel and Kamenica 2019). Some of these foundations lead to infinite costs of certainty—that is, cost functions defined by measures of uncertainty that take extended real values—but are finite on the interior of the space of beliefs. In the online appendix, I show that my model can be extended to accommodate such cost functions without meaningfully affecting my results.

The online appendix also considers an extension of the model that allows the researcher to have an intrinsic motivation for experimentation. This feature is natural in many applications: for instance, a pharmaceutical firm may have an interest in the regulatory approval of their vaccine and so may undertake clinical trials even in the absence of a subsidy. Adapting a result from Jullien (2000), I show that my results are unaffected by this extension, so long as the principal asks each researcher type to conduct a more costly experiment than the next highest cost type would conduct in the absence of a subsidy.

Finally, note that if the researcher has additional costs that are contractible, we can include them in the model by assigning them to the principal. Then each type of contract considered becomes a cost-plus contract, but as long as the researcher's contractible costs are uninformative about her type, the analysis is otherwise unchanged. On the other hand, if these costs provide information about the researcher's type, then as in Laffont and Tirole (1986), the principal may be able to extract additional surplus by tying payment to contractible costs in a manner other than simply reimbursing them in full.

III. Methods-Based Contracting

If methods are contractible, we can write the principal's optimal (direct revelation) contracting problem as

$$\max_{\chi,t} E_F[E_{\chi(\theta)}[w(p)] - t(\theta)] \tag{1}$$

$$\text{such that } E_{\chi(\theta)} p = p_0 \qquad \forall\, \theta \in \Theta,$$

$$t(\theta) - C(\chi(\theta),\theta) \geq t(\theta') - C(\chi(\theta'),\theta) \qquad \forall\, \theta,\theta' \in \Theta, \qquad (\text{IC}\theta)$$

$$t(\theta) - C(\chi(\theta),\theta) \geq 0 \qquad \forall\, \theta \in \Theta. \qquad (\text{IR}\theta)$$

Fairly standard arguments following Maskin and Riley (1984) imply that the incentive constraints reduce to (IRθ) for $\theta = \theta_1$, which binds; a set of binding upward incentive compatibility conditions between adjacent types; and the monotonicity constraints $E_{\chi(\theta_{i-1})}[c(p)] \leq E_{\chi(\theta_i)}[c(p)],\ i \in \{2, \ldots, N\}$.[14] Hence, the principal's choice of χ pins down t as

$$t(\theta_1) = \theta_1 E_{\chi(\theta_1)}[c(p)], \tag{2}$$

$$t(\theta_i) = \theta_i E_{\chi(\theta_i)}[c(p)] + \sum_{j=1}^{i-1}(\theta_j - \theta_{j+1}) E_{\chi[\theta_j]}[c(p)] \ \ \forall\ i \in \{2, \ldots, N\}, \tag{3}$$

and we can rewrite the principal's problem as

$$\begin{aligned} &\max_{\chi} E_F[E_{\chi(\theta)}[w(p) - g(\theta)c(p)]] \\ &\text{such that } E_{\chi(\theta)} p = p_0 \qquad \forall\ \theta \in \Theta, \end{aligned} \tag{4}$$

$$E_{\chi(\theta_{i-1})}[c(p)] \ \leq\ E_{\chi(\theta_i)}[c(p)] \ \ \forall\ i \in \{2, \ldots, N\}, \tag{5}$$

where the *virtual type* $g(\theta)$ of a type θ researcher is given by

$$g(\theta_i) \equiv \theta_i + \frac{F(\theta_{i+1})}{f(\theta_i)}(\theta_i - \theta_{i+1}), i \in \{1, \ldots, N-1\};\ g(\theta_N) \equiv \theta_N, \tag{6}$$

letting f denote the probability mass function for F.[15]

Neglecting the monotonicity constraints (5) for now, the principal's problem (4) is separable into a set of N problems—one for each type θ—of the form

$$\max_{\tau}\{E_\tau v(p) \text{ such that } E_\tau p = p_0\}. \tag{7}$$

This is a fairly standard *Bayesian persuasion* or *information design* problem (Kamenica and Gentzkow 2011; Kamenica 2019). This class of problems consists of choosing a distribution to maximize the expectation of a *value function* v, subject to a Bayes plausibility constraint.

When v is upper semicontinuous, Kamenica and Gentzkow (2011) show that a solution to (7) exists whose support has no more elements than the number of states of the world. This has two immediate implications for methods-based contracting. First, when the monotonicity constraint (5) is disregarded, the principal's problem (4) has a solution χ^*. Second, that solution involves binary outcome experiments: for each type θ, the experiment

[14] For a proof, see lemma 4 in the appendix.

[15] Observe that (6) implies that the type θ_i researcher's information rent is proportional to the ratio $F(\theta_{i+1})/f(\theta_i)$. This contrasts with, e.g., Maskin and Riley (1984), where a type θ agent's information rent is instead proportional to the inverse hazard rate $(1 - F(\theta))/f(\theta)$. This is because better (i.e., lower-cost) types are smaller in my model, and so the probability of a type that is better than θ_i is $\Sigma_{j=i+1}^{N} f(\theta_j) = F(\theta_{i+1})$ rather than $1 - F(\theta_i)$.

$\chi^*(\theta)$ either gives one positive and one negative result or is totally uninformative.

A novel comparative statics result tells us how the experiments $\chi^*(\theta)$ that are part of this solution differ from one another. Say that a function f is *additively more concave* than a function g if $f = g + h$ for some continuous, strictly concave function h.[16] In the type θ Bayesian persuasion problem from (4), the principal's value function is additively more concave when the virtual type $g(\theta)$ is larger. Proposition 1 shows that this means that a solution to a problem featuring a smaller virtual type will be more informative (in the sense of Blackwell 1953) than a solution to a problem with a larger virtual type. Consequently, so long as higher-cost researchers have larger virtual types—as they do, for instance, when F has a monotone reversed hazard rate and Θ is a set of consecutive integers—the monotonicity constraints (5) can be safely ignored.[17]

Proposition 1 (Additive convexity and monotonicity). Suppose that $v_0 : [0, 1] \to \mathbb{R}$ is additively more concave than $v_1 : [0, 1] \to \mathbb{R}$. Then for any $p_0 \in (0, 1)$ and any solutions to the persuasion problems

$$\tau_0^* \in \arg\max_{\tau\in\Delta([0,1])}\{E_\tau v_0(p) \text{ such that } E_\tau p = p_0\},$$

$$\tau_1^* \in \arg\max_{\tau\in\Delta([0,1])}\{E_\tau v_1(p) \text{ such that } E_\tau p = p_0\},$$

τ_1^* is weakly Blackwell more informative than (i.e., a mean-preserving spread of) τ_0^*.

In a Bayesian persuasion problem, an additively more concave value function corresponds to a larger difference between the costs of more and less informative experiments, as a consequence of Jensen's inequality. Proposition 1 shows that with two states of the world, it also leads the problem to have a less informative solution. At first glance, this conclusion might appear to follow from Milgrom and Shannon (1994), since the objective functions $E_\tau v_i(p)$ have the single crossing property in (τ, i) in the Blackwell order. But Milgrom and Shannon's result requires quasisupermodularity in the choice variable τ for each i, which need not hold here. Consequently, proposition 1 must make use of the persuasion problem's structure in ways that are not captured by the single crossing property.

[16] This differs from the usual definition of *more concave* ($f = \phi \circ g$ for some increasing concave ϕ). This is easiest to see when g is not strictly increasing: if $g(x) = g(y)$ but $f(x) \neq f(y)$, then it is possible for f to be additively more concave than g (since we need not have $h(x) = h(y)$) but not more concave than g (since $\phi(g(x)) = \phi(g(y))$ for any ϕ). Alternatively, in the differentiable case, recall that f is more concave than g if it has a larger Arrow-Pratt coefficient of risk aversion; in contrast, f is additively more concave than g if it has a smaller second derivative.

[17] To see this, write $g(\theta_i) = \theta_i + [(F(\theta_i) - f(\theta_i))/f(\theta_i)](\theta_i - \theta_{i+1}) = \theta_{i+1} + [F(\theta_i)/f(\theta_i)](\theta_i - \theta_{i+1})$.

This is accomplished by way of another comparative statics result that extends to all finite-state Bayesian persuasion problems. In the general persuasion context, proposition 4 (see app. B) shows that when the sender's value function becomes additively less concave, the posteriors induced by the optimal signal are either (a) outside of or (b) extreme points of the convex hull of the posteriors induced by the original problem's solution. In the two-state setting, this implies the Blackwell ordering that I exploit in this paper (proposition 1) and is equivalent when both experiments have binary outcomes. In settings with more than two states, it is neither stronger nor weaker than the Blackwell order but does rule out the possibility that the Blackwell order is reversed. Proposition 4 thus formalizes and strengthens Gentzkow and Kamenica's (2014) observation that an additively more concave problem cannot have a more informative solution.

The intuition for the result is as follows. Suppose that a posterior p produced by a solution to the additively less concave problem could be represented as a convex combination $\Sigma_{i=1}^{d}\lambda_i p_i$ of posteriors $\{p_i\}_{i=1}^{d}$ produced by a solution to the additively more concave problem. Then the additively more concave problem's value function v_0 must be no higher at p than the convex combination $\Sigma_{i=1}^{d}\lambda_i v_0(p_i)$ of its values at the posteriors $\{p_i\}_{i=1}^{d}$. Otherwise, the value of the additively more concave problem could be increased by shifting probability from the posteriors $\{p_i\}_{i=1}^{d}$ to p. Then this must hold strictly for the additively less concave value function v_1; that is, we must have $\Sigma_{i=1}^{d}\lambda_i v_1(p_i) > v_1(p)$. Hence, the value of the additively less concave problem could be increased by shifting the probability that its solution places on p to $\{p_i\}_{i=1}^{d}$.

While the implications of these results for methods-based contracting are most straightforward when the researcher's virtual type $g(\theta)$ is increasing in θ, they do not require this. In general, standard ironing techniques (see app. A) can be applied to the virtual types $g(\theta)$ so as to separate the principal's problem (4) into a set of type-specific persuasion problems whose value functions are additively more concave for larger types. Consequently, together with the results of Kamenica and Gentzkow (2011), proposition 1 shows that regardless of the belief distribution F over researcher types, the principal has an optimal methods-based contract. Furthermore, these results show that this contract is a member of a particular class—one that will prove essential to the main results of this paper. Say that an experiment choice function $\chi : \Theta \to X$ is *binary* if for each θ, the support of $\chi(\theta)$ has at most two elements, and it is *Blackwell monotone* if for each $\theta' > \theta$, $\chi(\theta)$ is a mean-preserving spread of $\chi(\theta')$. Then the principal's optimal methods-based contract can be characterized as follows:

Theorem 1 (Optimal methods-based contracts). The principal has an optimal methods-based contract that implements a binary, Blackwell-monotone experiment choice function.

Proposition 1 also helps us understand the inefficiencies produced by the optimal contract. Observe that the lowest cost type, θ_N, is the same as its virtual type, $g(\theta_N)$. Consequently, the principal's type θ_N value function is simply the difference of its payoffs and the researcher's costs, and so it chooses $\chi(\theta_N)$ the same way it would if the researcher's type was common knowledge. However, for each $i < N$, the principal's type θ_i value function is the difference of its payoffs and the researcher's costs plus a strictly concave function representing information rent. Thus, it chooses a less informative $\chi(\theta_i)$ than it would if it knew the researcher's type and so did not face the incentive compatibility constraints (ICθ). These are, of course, the standard distortions associated with screening: there is no distortion at the top and distortion everywhere else. Here, proposition 1 tells us that this distortion takes the form of garbling: information asymmetry results in experiments that are not only less costly to the researcher and less valuable to the principal but also less valuable to any decision maker.

Theorem 2 (Distortion from hidden information). Let (χ^*, t^*) be a solution to the principal's optimal contracting problem (1), and let $(\tilde{\chi}, \tilde{t})$ be a solution to the principal's full-information optimal contracting problem ([1] without [ICθ]). Then (χ^*, t^*) exhibits

i. no distortion at the top: $W(\chi^*(\theta_N)) - C(\chi^*(\theta_N), \theta_N) = W(\tilde{\chi}(\theta_N)) - C(\tilde{\chi}(\theta_N), \theta_N)$;
ii. garbling everywhere else: for each $i \in \{1, \dots, N-1\}$, $\chi^*(\theta_i)$ is Blackwell-less informative than $\tilde{\chi}(\theta_i)$; and
iii. distortion lowers cost to the researcher and value for the principal: for each $i \in \{1, \dots, N-1\}$, $C(\chi^*(\theta_i), \theta_i) \leq C(\tilde{\chi}(\theta_i), \theta_i)$ and $W(\chi^*(\theta_i)) \leq W(\tilde{\chi}(\theta_i))$.

IV. Results-Based Contracting

Now that we know what kind of choice functions a principal with methods-based contracts at its disposal will seek to implement—namely, binary, Blackwell-monotone ones—we can ask whether choice functions in this class are implementable using a results-based contract and, if so, whether their expected costs to the principal will be the same.[18] To see why they might not be, note that replacing a direct revelation contract (χ, t) with a results-based contract ψ in the principal's problem means more than just replacing $t(\theta)$ with $E_{\chi(\theta)}\psi(p)$. It also means expanding the set of incentive compatibility constraints: in addition to ensuring that a type θ researcher would prefer to perform $\chi(\theta)$ rather than any of the experiments that the principal has asked other types to perform, the principal must also

[18] Since the set of types is discrete, revenue equivalence need not hold—the incentive compatibility constraints do not pin down transfers up to a constant.

consider deviations to any experiment that might produce the same result as one of those experiments, that is, any experiment whose support overlaps with that of $\chi(\theta')$ for some $\theta' \in \Theta$.[19]

These *global incentive constraints* can be written as

$$\chi(\theta) \in \arg\max_{\tau}\{E_\tau[\psi(p) - \theta c(p)] \text{ such that } E_\tau p = p_0\}, \forall\, \theta \in \Theta. \quad (8)$$

Much like the principal's methods-based contracting problem, the researcher's problem in these constraints is one of Bayesian persuasion. However, instead of requiring us to find a solution given a value function, the results-based implementation problem is the reverse: we must find a transfer function ψ given solutions χ. It is best understood geometrically by building on the results of Kamenica and Gentzkow (2011).

Given a function $v : S \to \mathbb{R}$, define its *concavification* $V : S \to \mathbb{R}$ as the smallest concave function that is at least as large as v, that is, the upper envelope of the convex hull of v's graph.[20] Kamenica and Gentzkow (2011) show that the value of the persuasion problem (7) is $V(p_0)$: the concavification of its value function v evaluated at the prior. Proposition 2 shows an important consequence of their characterization: an experiment producing the pair of results $\{\underline{p}, \bar{p}\}$ solves the problem precisely when the affine function that coincides with v at both results—the secant line to v through $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$—is everywhere greater than v.

PROPOSITION 2 (A necessary and sufficient secant line condition). Suppose that $v : [0, 1] \to \mathbb{R}$ is upper semicontinuous and that $p_0 \in (0, 1)$. An experiment τ^* that produces the pair of results $\operatorname{supp} \tau^* = \{\underline{p}, \bar{p}\}$ solves the persuasion problem $\max_\tau\{E_\tau v(p)$ such that $E_\tau p = p_0\}$ if and only if the secant line to v through $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$ lies on or above the graph of v:

$$v(\underline{p}) + \frac{v(\bar{p}) - v(\underline{p})}{\bar{p} - \underline{p}}(p - \underline{p}) \geq v(p) \text{ for each } p \in [0, 1]. \quad (9)$$

Intuitively, since the convex hull of a set (such as the graph of v, $\mathrm{Gr}(v)$) is the intersection of all half-spaces containing that set, a point is on the boundary of the convex hull if and only if it lies in a supporting hyperplane of the original set. Consequently, since it is a convex combination of $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$, the point $(E_{\tau^*} p, E_{\tau^*} v(p))$ is on the boundary of the convex hull of $\mathrm{Gr}(v)$ if and only if the secant line through $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$ is a supporting hyperplane of $\mathrm{Gr}(v)$. In particular,

[19] For instance, a type θ_i researcher may wish to perform an experiment that produces positive results like those of $\chi(\theta_{i-1})$ but with a greater frequency because it produces negative results like those of $\chi(\theta_i)$.

[20] Formally, let $V(p) \equiv \sup\{z \mid (p, z) \in \mathrm{conv}(\mathrm{Gr}(v))\}$.

$(E_{\tau^*}p, E_{\tau^*}v(p))$ is on the upper boundary of Gr(v)'s convex hull—and hence $E_{\tau^*}v(p) = V(p_0) = \max_\tau\{E_\tau v(p)$ such that $E_\tau p = p_0\}$—if and only if the secant line through $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$ lies on or above the graph of v. Figure 1 illustrates.

If χ is binary and Blackwell monotone, proposition 2 lets us construct a results-based contract ψ_χ that implements χ. For simplicity, consider the case where χ instructs each type to conduct an informative experiment, that is, supp $\chi(\theta_i) = \{\underline{y}_i, \bar{y}_i\}$ for each i.[21]

The construction of ψ_χ proceeds recursively from the prior outward. In each step, transfers are set for the results of some type θ_i's experiment as well as those less extreme results that are still more extreme than those of any higher-cost type; that is, for $[\underline{y}_1, \bar{y}_1]$, if $i = 1$, or $[\underline{y}_i, \underline{y}_{i-1}) \cup (\bar{y}_{i-1}, \bar{y}_i]$, if $i > 1$.[22] For each of these results, the researcher is reimbursed for the cost to type θ_i of producing them, adjusted by an affine function $s_i(p)$ whose form depends on the transfers set in the previous step. Finally, for results stronger than those produced by any type's experiment, payments are set the same way they were for the results of the lowest-cost researcher's experiment.

In the first step, transfers are set for the highest-cost researcher's results as well as each less extreme result the same way that Rappoport and Somma (2017) set compensation in their setting with a single researcher type. That is, the affine function s_1 is defined so that it is tangent to the type θ_1 researcher's cost function at the prior:

$$\psi_\chi(p) \equiv s_1(p) + \theta_1 c(p) \text{ for each } p \in [\underline{y}_1, \bar{y}_1],$$
$$\text{where } s_1(p) \equiv -\theta_1 c'(p_0)(p - p_0).$$

Thus, $s_1(p)$ coincides with the highest-cost researcher's value function $\psi_\chi(p) - \theta_1 c(p)$ on $[\underline{y}_1, \bar{y}_1]$. Then since $s_1(p_0) = 0$, when we replace the transfer $t(\theta_1)$ with the expected results-based transfer $E_{\chi(\theta_1)}\psi_\chi(p)$, the participation constraint (IRθ) binds for $\theta = \theta_1$.

In subsequent steps, transfers are set so as to make the type θ_i researcher indifferent between her experiment and that of the next lowest-cost type θ_{i-1}. That is, for each $i > 1$, the affine function s_i is defined so that it coincides with the type θ_i researcher's value function at the results $\underline{y}_{i-1}$ and $\bar{y}_{i-1}$ generated by the next lowest-cost type θ_{i-1}'s experiment:

$$\psi_\chi(p) \equiv s_i(p) + \theta_i c(p) \text{ for each } p \in [\underline{y}_i, \underline{y}_{i-1}) \cup (\bar{y}_{i-1}, \bar{y}_i],$$
$$\text{where } s_i(p) \equiv \psi_\chi(\underline{y}_{i-1}) - \theta_i c(\underline{y}_{i-1}) + \frac{(\psi_\chi(\bar{y}_{i-1}) - \theta_i c(\bar{y}_{i-1})) - (\psi_\chi(\underline{y}_{i-1}) - \theta_i c(\underline{y}_{i-1}))}{\bar{y}_{i-1} - \underline{y}_{i-1}}(p - \underline{y}_{i-1}).$$

[21] When χ instructs some types not to perform an informative experiment, the procedure for constructing ψ_χ proceeds identically but starts from the highest-cost type that performs an informative experiment rather than the highest-cost type overall.

[22] Blackwell monotonicity ensures that $\underline{y}_i \le \underline{y}_{i-1}$ and $\bar{y}_{i-1} \le \bar{y}_i$.

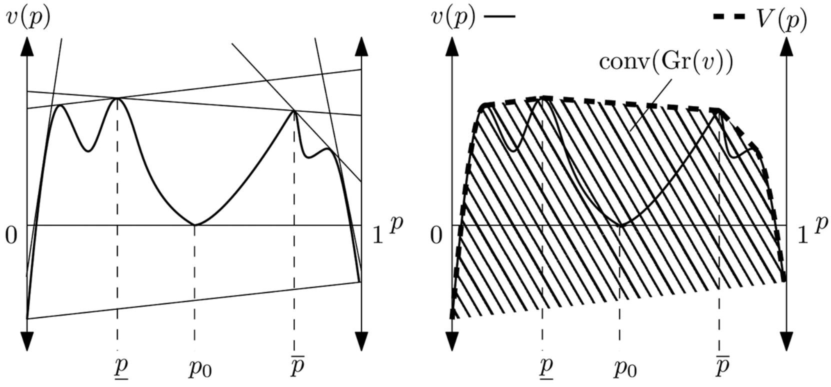

FIG. 1.—Secant lines and optimality. *Left*, the convex hull of v's graph is the intersection of all half-spaces that contain it. Note that the secant line through $(\underline{p}, v(\underline{p}))$ and $(\bar{p}, v(\bar{p}))$ is a supporting hyperplane of $\mathrm{Gr}(v)$. *Right*, the point $([(\bar{p} - p_0)/(\bar{p} - \underline{p})]\underline{p} + [(p_0 - \underline{p})/(\bar{p} - \underline{p})]\bar{p}, [(\bar{p} - p_0)/(\bar{p} - \underline{p})]v(\underline{p}) + [(p_0 - \underline{p})/(\bar{p} - \underline{p})]v(\bar{p}))$ is on the upper boundary of $\mathrm{conv}(\mathrm{Gr}(v))$—and hence coincides with $(p_0, V(p_0))$—precisely because it lies on a supporting hyperplane of $\mathrm{Gr}(v)$ that is on or above $\mathrm{Gr}(v)$.

This ensures that $s_i(p)$ coincides with the type θ_i researcher's value function $\psi_\chi(p) - \theta_i c(p)$ on $[\underline{y}_i, \underline{y}_{i-1}]$ and $[\bar{y}_{i-1}, \bar{y}_i]$. Consequently, both her experiment $\chi(\theta_i)$ and the next lowest cost type's experiment $\chi(\theta_{i-1})$ give her the same expected payoff, $s_i(p_0)$, and her upward incentive constraint binds: when we replace $t(\theta)$ with the expected results-based transfer $E_{\chi(\theta)}\psi_\chi(p)$, (IC$\theta$) binds for $(\theta, \theta') = (\theta_i, \theta_{i-1})$. Since the participation constraint (IRθ) binds for $\theta = \theta_1$, this means that ψ_χ's expected transfers to a researcher who performs $\chi(\theta)$ are pinned down the same way they were in the methods-based contracting problem. Hence, no contract that implements χ can offer lower expected payment (and thus lower cost to the principal) for any type's experiment $\chi(\theta)$.

Finally, for results more extreme than those produced by any type's experiment, ψ_χ is set so that the lowest-cost researcher's value function coincides with the affine function S_N:

$$\psi_\chi(p) \equiv s_N(p) + \theta_N c(p) \text{ for each } p \in [0, \underline{y}_N) \cup (\bar{y}_N, 1].$$

Figure 2 illustrates the process of constructing the results-based contract ψ_χ.

To see that the results-based contract ψ_χ satisfies the global incentive constraints (8), first note that for each i, $s_i(p)$ is the secant line to the type θ_i researcher's value function, $\psi_\chi(p) - \theta_i c(p)$, through the points on its graph corresponding to the results of her experiment. Consequently, proposition 2 tells us that the global incentive constraint for the type θ_i researcher (8) is satisfied if and only if her value function never exceeds

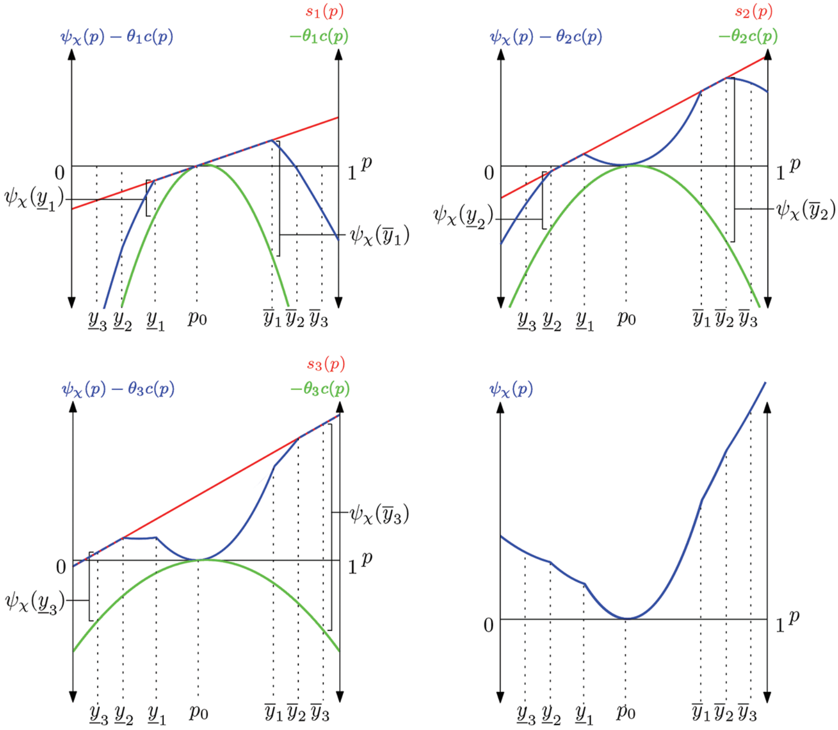

FIG. 2.—Constructing ψ_x when $N = 3$. *Top left*, on $[\underline{y}_1, \bar{y}_1]$, ψ_x is the difference between the highest-cost researcher's posterior-specific cost (green) and the affine function s_1 tangent to it at p_0 (red), so that the latter coincides with her value function (blue) there. *Top right*, on $[\underline{y}_2, \underline{y}_1)$ and $(\bar{y}_1, \bar{y}_2]$, ψ_x is constructed so that the middle-cost researcher's value function (blue) coincides with an affine function s_2 (red) on the closed intervals $[\underline{y}_2, \underline{y}_1]$ and $[\bar{y}_1, \bar{y}_2]$ by letting ψ_x be the difference of that affine function and the middle-cost researcher's posterior-specific cost (green). *Bottom left*, on $[0, \underline{y}_2)$ and $(\bar{y}_2, 1]$, ψ_x is constructed so that the lowest-cost researcher's value function (blue) coincides with an affine function s_3 (red) on the closed intervals $[0, \underline{y}_2]$ and $[\bar{y}_2, 1]$ by letting ψ_x be the difference of that affine function and the lowest-cost researcher's posterior-specific cost (green). *Bottom right*, results-based contract ψ_x.

$s_i(p)$ at any result p. For those results where $s_i(p)$ coincides with the type θ_i researcher's value function—$[\underline{y}_1, \bar{y}_1]$, if $i = 1$, or $[\underline{y}_i, \underline{y}_{i-1}] \cup [\bar{y}_{i-1}, \bar{y}_i]$, if $i > 1$—this is straightforward. Everywhere else, it is more subtle.

Consider that at the results $\bar{y}_k$ and $\underline{y}_k$ of the type θ_k researcher's experiment, the affine function s_k coincides with her value function, while the affine function s_{k+1} coincides with the next highest-cost type θ_{k+1}'s value function. Since type θ_k's value function is additively more concave than type θ_{k+1}'s, it follows that at more extreme results, it must cave downward relative to s_k more than type θ_{k+1}'s value function caves downward relative to s_{k+1}:

$$s_k(p) - (\psi_\chi(p) - \theta_k c(p)) \geq s_{k+1}(p) - (\psi_\chi(p) - \theta_{k+1} c(p)) \text{ for each } p \notin (\underline{y}_k, \bar{y}_k).$$

(Compare the top left and top right panels of fig. 2 on $[0, \underline{y}_1]$ and $[\bar{y}_1, 1]$). Conversely, at less extreme results, that is, $p \in [\underline{y}_k, \bar{y}_k]$, this inequality must be reversed: type θ_k's value function must cave downward relative to s_k less than type θ_{k+1}'s value function caves downward relative to s_{k+1}. (Compare the top right and bottom left panels of fig. 2 on $[\underline{y}_2, \bar{y}_2]$.) Applying this argument recursively yields

$$s_i(p) - (\psi_\chi(p) - \theta_i c(p)) \geq s_{j+1}(p) - (\psi_\chi(p) - \theta_{j+1} c(p)) \quad \text{for each } j \geq i \text{ and } p \notin (\underline{y}_j, \bar{y}_j), \tag{10}$$

$$s_\ell(p) - (\psi_\chi(p) - \theta_\ell c(p)) \leq s_i(p) - (\psi_\chi(p) - \theta_i c(p)) \quad \text{for each } \ell < i \text{ and } p \in [\underline{y}_\ell, \bar{y}_\ell]. \tag{11}$$

Observe that (11) applies on the intervals where the higher-cost type θ_ℓ's value function coincides with the affine function s_ℓ, and hence the left-hand side of (11) is zero. Likewise, (10) applies on the intervals where the lower-cost type θ_{j+1}'s value function coincides with s_{j+1}, and hence the right-hand side of (10) is zero. Since those intervals include every result where the type θ_i researcher's value function does not coincide with $s_i(p)$, it follows that the former is weakly smaller than the latter at each $p \in [0, 1]$—and hence that her global incentive constraint is satisfied.[23]

This construction of an implementing results-based contract ψ_χ is formalized in theorem 3.

Theorem 3 (Results-based implementation). Any binary, Blackwell-monotone experiment choice function $\chi : \Theta \to X$ can be implemented with a results-based contract ψ_χ. Furthermore, this contract ψ_χ cannot

[23] It is worth noting that the reasoning used here to rule out nonlocal deviations by the researcher has significant similarities to the logic underlying proposition 1: both rely on the fact that, relative to the secant line through the points on its graph corresponding to the posteriors $\underline{p}$ and $\bar{p}$, an additively more concave value function must cave downward strictly less on the interval $(\underline{p}, \bar{p})$ than an additively less concave value function does. Furthermore, both arguments use this fact by letting $\underline{p}$ and $\bar{p}$ be the lowest and highest posteriors, respectively, in the support of an optimal experiment for the additively more concave problem. The key difference is that proposition 1 relies on this experiment's optimality to ensure that the additively more concave function lies below its secant line through the points on its graph corresponding to $\underline{p}$ and $\bar{p}$ and hence that the additively less concave value function lies strictly below its secant line through the points on its graph corresponding to $\underline{p}$ and $\bar{p}$. In contrast, the argument here cannot rely on the optimality of an experiment inducing $\underline{p}$ and $\bar{p}$ in the additively more concave problem, since that is itself part of what we wish to show. Instead, it must rely on the construction of ψ_χ itself.

be outperformed by a methods-based contract T: If T implements χ, then $E_{\chi(\theta)}\psi_{\chi}(p) \leq T(\chi(\theta))$ for each $\theta \in \Theta$.

The binary character of the experiment choice function χ is not critical to the implementation result. As long as the experiments conducted by lower-cost types produce results at least as extreme as any of those produced by higher-cost types' experiments, they can be implemented with a results-based contract: we can simply replace $\bar{y}_i$ and $\underline{y}_i$ with the strongest negative and positive results, respectively, that type θ_i's experiment might produce and construct ψ_{χ} as before. However, this contract would underperform a methods-based contract, unless the only type conducting an experiment with more than two outcomes was the lowest-cost type. Consider that if χ is not binary, some type θ_i's experiment must place positive probability on results in $(\underline{y}_i, \bar{y}_i)$. On that interval, the type θ_{i+1} researcher's value function lies below the affine function s_{i+1} with which it coincides at each result produced by her experiment $\chi(\theta_{i+1})$. Consequently, type θ_{i+1}'s upward incentive constraint no longer binds: she must get a strictly higher payoff from conducting her own experiment than from conducting θ_i's. Hence, the expected cost of this contract must be greater than the cost of implementing χ with a methods-based contract where transfers are specified by (2) and (3), and so each upward incentive constraint binds.

Blackwell monotonicity of the experiment choice function χ, on the other hand, is not only crucial to constructing the results-based contract ψ_{χ}, it is required for implementation with any results-based contract: no matter what contract the principal offers, the value function of a higher-cost researcher will differ from that of a lower-cost researcher by a strictly concave function of p. Then proposition 1 tells us that the experiment chosen by the lower-cost researcher will be Blackwell more informative than the one chosen by the higher-cost researcher. Consequently, these experiments can match those prescribed by an experiment choice function only if that choice function is Blackwell monotone.[24]

Theorem 4 (Results-based implementation requires Blackwell monotonicity). If χ is not Blackwell monotone, it cannot be implemented by a results-based contract.

V. Applications

Here, I discuss how my main results can help inform our understanding of the real-world applications considered in the introduction.

[24] In fact, when the choice function is not binary, Blackwell monotonicity is not enough: proposition 4 (see app. B) can be invoked to show that for for results-based implementation to be possible, all of the results produced by a lower-cost researcher's experiment must be at least as extreme as any result produced by a higher-cost researcher's experiment.

A. Clinical Trial Subsidies

Consider the contracts signed with Moderna (2020) and Novavax (2020a, 2020b, 2021) as part of the US Government's Operation Warp Speed. These are cost-plus-fixed-fee contracts: the firms are paid a fixed fee plus reimbursement for the observable portion of their costs. Payment is explicitly contingent on methods: both contracts contain details of the trial designs that the vaccine developers are to use.[25] Unobservable costs appear to have played a role in contracting: for instance, during negotiations over trial protocols, Moderna behaved as if it had significant costs that were not reimbursable under the contract, initially resisting some elements of the study designs proposed by the government that it viewed as overly burdensome (Taylor and Respaut 2020). Such costs would have been difficult for the government to infer from the firms' past behavior, since at the time, Moderna had never run a stage 3 clinical trial, and Novavax had conducted only two of them.[26]

If the firms' observable costs do not reveal any of their private information about unobservable costs, then as discussed in section II, the model can accommodate such costs by assigning them to the principal. Under this accounting convention, the Moderna and Novavax contracts are both methods-based contracts, where the payment for conducting the contracted trials is specified by the fixed fee. While neither contract contains a screening mechanism, we can follow Laffont and Tirole (1993, 82) and interpret each as the outcome of a bargaining process in which the government screens the vaccine manufacturer on its unobservable costs. Given this interpretation, theorem 1 gives us a sense of the way the final contracts we observe depend on the firms' unobservable costs, while theorem 2 helps us understand the distortionary impact of the firms' private information about those costs.

B. Pretrial Investigations and Contingency Fees

This paper's analysis can also shed light on the impact of private information on contracts for legal services. When the principal is a potential plaintiff, such contracts often specify payment on a contingency basis, that is, as some function of any award or settlement.[27] A 2018 contract between the

[25] Some of these are heavily redacted, especially in the case of Moderna's contract; there, the least redacted descriptions of the trial designs can be found in the revised statement of work accompanying Modification P00003.

[26] See https://clinicaltrials.gov/ct2/results?term=novavax+OR+moderna&Search=Apply&recrs=e&age_v=&gndr=&type=&rslt=&phase=2.

[27] There is a significant law and economics literature focusing on contingency fees as a solution to agency issues between attorney and client, including private information about attorney ability (see, e.g., Dana and Spier 1993; Rubinfeld and Scotchmer 1993). In general, this literature does not explicitly model the attorney's information gathering process. Hence,

state of Georgia and outside counsel to pursue an investigation of and litigation against firms in the prescription opioid market is an illustrative (and publicly available) example (see State of Georgia 2018). This contract explicitly specifies compensation as a function of the amount recovered by the state, net of the attorneys' expenses (i.e., observable costs), which are reimbursed in full in the event that such a recovery occurs.

If we abstract away from the amount of the defendant's liability and think of the state of the world as its presence or absence, we can use the following example to consider such contingency fee schedules in the context of this paper. For simplicity, assume that a pretrial settlement is always reached.[28] Let the amount of the settlement net of the attorney's expenses be a strictly increasing function σ of the posterior probability that the defendant is liable. Finally, adopt the accounting convention discussed in section II that observable costs (here, attorney expenses) are assigned to the principal.

Then the expected payment (net of observable costs) to an attorney who accepts the contingency fee schedule ϕ and produces result p is $\psi_\phi(p) \equiv \phi(\sigma(p))$. Accordingly, the contingency fee schedule ϕ offers the same incentives for the attorney's investigation as the results-based contract ψ_ϕ. Since the net settlement amount is a one-to-one function of the result, we can also take any results-based contract ψ and find a contingency fee schedule that gives the same incentives: choose $\phi_\psi(x) \equiv \psi(\sigma^{-1}(x))$. Then $\psi_{\phi_\psi}(p) = \psi(p)$ for each p, and thus ϕ_ψ incentivizes pretrial information acquisition identically to ψ.

It follows that in this example, the classes of contingency fee schedules and results-based contracts are equivalent. Hence, by theorem 3, there is no loss in this setting from using a contingency fee to motivate the attorney's investigation rather than a methods-based contract. Consequently, theorem 1 shows how the attorney will respond to an optimal contingency fee schedule, and theorem 2 reveals the ways that private information about costs can distort her investigative decisions.

VI. Conclusion

This paper analyzes the design of incentives for flexible information acquisition in a tractable screening setting. In this environment, it offers two broad conclusions.

---

this example's analysis innovates by showing that contingency fee schedules can allow clients to optimally screen attorneys by ability in a setting with flexible information acquisition. Note that in contrast to, e.g., Rubinfeld and Scotchmer (1993), I do not require contingency fee schedules to be affine functions of the amount recovered from the defendant. In practice, contingency fees are often nonlinear (see, e.g., Ohio Attorney General 2019).

[28] In sec. S.2 of the online appendix, I show that the same conclusions hold in a version of this example that accounts for the possibility of going to trial.

First, the familiar rankings from the screening literature—higher types take higher actions, and each type takes a lower action than they would absent asymmetric information—extend to settings where those actions are experiments and are given the Blackwell (1953) order (theorems 1 and 2). This is true even though Blackwell informativeness is only a partial order, and the setting does not lend itself to application of Milgrom and Shannon (1994). In particular, asymmetric information about researcher costs not only causes underinvestment in experimentation, it actually causes less information to be gathered: any agent who relies on the researcher's experiment to make decisions is made worse off relative to the symmetric information case.

The second conclusion concerns a key difference between an experiment and other choices an agent might make: to describe an experiment, we need to know each of its potential outcomes, but we may be able to observe only the outcome that actually occurs. This can make it infeasible to contract on the form of an experiment (the entire distribution τ) but feasible to contract on the observed outcome of the experiment (the posterior p). This adds extra incentive constraints to the contracting problem. However, theorem 3 shows that it does not affect the problem's solution: any strategy for experimentation that the principal might actually want to induce, it can induce by contracting only on the experiment's outcome and at the same expected cost. As a corollary, the paper's first broad conclusion (theorems 1 and 2) continues to apply when only the experiment's result is contractible.

Appendix A

Ironing

When higher types do not correspond to higher virtual types—that is, when g is not monotone—separating the principal's problem (4) into type-specific problems requires the use of ironing techniques following, for example, Myerson (1981). Let $F^{-1}:[0,1]\to\Theta$ denote the quantile function of F: $F^{-1}(q)\equiv\min\{\theta\in\Theta|F(\theta)\geq q\}$. Define $R:[0,1]\to\mathbb{R}$ by

$$R(q)\equiv\begin{cases}-\sum_{\theta\leq F^{-1}(q)}g(\theta)f(\theta), & q>0;\\ 0, & q=0,\end{cases}$$

and let $P:[0,1]\to\mathbb{R}$ denote the concavification of R. Finally, define the *ironed virtual type* $\bar{g}(\theta)$ of a type θ researcher as

$$\bar{g}(\theta_i)\equiv\frac{P(F(\theta_{i+1}))-P(F(\theta_i))}{f(\theta_i)},\ i\in\{1,\dots,N-1\};\ \bar{g}(\theta_N)\equiv\frac{P(0)-P(F(\theta_N))}{f(\theta_N)}.$$

Proposition 3 (Ironing the principal's problem). If $\chi^*:\Theta\to\mathbf{X}$ solves the *ironed problem*

$$\max_{\chi} E_F[E_{\chi(\theta)}[w(p) - \bar{g}(\theta)c(p)]] \text{ such that } E_{\chi(\theta)}p = p_0 \ \ \forall\, \theta \in \Theta, \quad \text{(A1)}$$

and $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for each $i \in \{2, \ldots, N\}$ with $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$, then χ^* solves the principal's reduced problem (4). Moreover, if $\hat{\chi} : \Theta \to \mathbf{X}$ solves the principal's reduced problem (4), then $\hat{\chi}$ also solves the ironed problem (A1).

Ironing results in some adjacent types being bunched—or given equal ironed virtual types—so as to ensure that the ironed virtual types are ordered in the same way as the researcher types they correspond to. This results in a problem (A1) that is almost equivalent to the principal's reduced problem but can be decomposed into separate persuasion problems for each type, with those for larger types having Blackwell-less informative solutions.[29] Lemma 1 summarizes these consequences of the ironing process.

Lemma 1 (Bunching and monotonicity in the ironed problem).

i. Higher-cost researchers have higher ironed virtual types: $\bar{g}(\theta_{i-1}) \geq \bar{g}(\theta_i)$ for $i \in \{2, \ldots, N\}$. Hence, if χ^* solves the principal's ironed problem (A1), and $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for each $i \in \{2, \ldots, N\}$ with $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$, then χ^* is Blackwell monotone.
ii. Bunching occurs where R is below its concavification: if $P(F(\theta_i)) > R(F(\theta_i))$ for $i \in \{2, \ldots, N\}$, then $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$. Hence, there is a binary solution χ^* to the principal's ironed problem (A1) such that $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for any $i \in \{2, \ldots, N\}$ with $P(F(\theta_i)) > R(F(\theta_i))$.
iii. Bunching never involves the lowest-cost type: $\bar{g}(\theta_N) < \bar{g}(\theta_{N-1})$, and so $P(F(\theta_N)) = R(F(\theta_N))$ and $\bar{g}(\theta_N) = g(\theta_N) = \theta_N$.

Intuitively, both the ironed virtual types $\bar{g}(\theta)$ and the virtual types $g(\theta)$ can be thought of as the slopes of secant lines connecting a sequence of points on the graph of a function—P in the former case and R in the latter. The ironed virtual types are monotone because P is concave. The virtual types, on the other hand, may fail to be monotone if R is not concave—and so lies below its concavification, P—at the quantiles of their corresponding types. Where this occurs, P must be affine, and so bunching must result.

Appendix B

Proofs

For a set S, let $\text{cl}(S)$ be its closure and $\text{ri}(S)$ be its relative interior.[30] For a convex set S, let $\text{ext}(S)$ be the set of its extreme points. For a function $v : S \to \mathbb{R}$, let $\text{hypo}(v) \equiv \{(s, z) | z \leq v(s)\}$ be its hypograph.

[29] When adjacent types are bunched and have the same ironed virtual type, and there are multiple solutions to the persuasion problem specific to that ironed virtual type, some solutions to the ironed problem (A1) may not solve the reduced problem (4) and may even fail to be Blackwell monotone.

[30] That is, its interior relative to the affine hull of S.

B1. Results for General Bayesian Persuasion Problems

LEMMA 2. If $\tau \in \Delta([0,1])$ and $\text{supp}\,\tau = \{\underline{p}, \bar{p}\}$ with $\underline{p} < \bar{p}$, then $\tau(\{\underline{p}\}) = (\bar{p} - E_\tau p)/(\bar{p} - \underline{p})$, $\tau(\{\bar{p}\}) = (E_\tau p - \underline{p})/(\bar{p} - \underline{p})$, and $\underline{p} < E_\tau p < \bar{p}$.

Proof. Since $\text{supp}\,\tau = \{\underline{p}, \bar{p}\}$, we have $\tau(\{\underline{p}\}), \tau(\{\bar{p}\}) > 0$, $\tau(\{\underline{p}\}) + \tau(\{\bar{p}\}) = 1$, and $E_\tau p = \tau(\{\underline{p}\})\underline{p} + \tau(\{\bar{p}\})\bar{p}$. Hence, $E_\tau p = \tau(\{\underline{p}\})\underline{p} + (1 - \tau(\{\underline{p}\}))\bar{p} \Leftrightarrow E_\tau p - \bar{p} = \tau(\{\underline{p}\})(\underline{p} - \bar{p}) \Leftrightarrow \tau(\{\underline{p}\}) = (\bar{p} - E_\tau p)/(\bar{p} - \underline{p})$. Likewise, $\tau(\{\bar{p}\}) = (1 - \tau(\{\underline{p}\})) = (E_\tau p - \underline{p})/(\bar{p} - \underline{p})$. Since these are nonzero, we must have $\underline{p} < E_\tau p < \bar{p}$. QED

LEMMA 3 (Properties of the concavification at the optimum). Let $S \subseteq \mathbb{R}^n$ be convex and compact, let $v: S \to \mathbb{R}$ be upper semicontinuous, let $p_0 \in \text{ri}(S)$, and let V be the concavification of v. If $\tau^* \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v(p) \text{ such that } E_\tau p = p_0\}$, then $v(p) = V(p)$ for all $p \in \text{supp}\,\tau^*$, and V is affine on $\text{conv}(\text{supp}\,\tau^*)$.

Proof. Since $p_0 \in \text{ri}(S)$ and V is by definition concave, V is superdifferentiable at p_0:[31] there exists $x \in \mathbb{R}^n$ such that $V(p_0) + x \cdot (p - p_0) \geq V(p) \geq v(p)$ for each $p \in S$. Then

$$V(p_0) = \int_S V(p_0) + x \cdot (p - p_0)\, d\tau^*(p) \geq \int_S v(p)\, d\tau^*(p) \geq \int_S v(p)\, d\tau^*(p). \tag{B1}$$

By Kamenica and Gentzkow's (2011) online appendix proposition 3, $E_{\tau^*} v(p) = V(p_0)$, and each inequality in (B1) must bind.[32] Then $V(p_0) + x \cdot (p - p_0) = V(p) = v(p)$ τ^*-almost everywhere. Suppose that this does not hold at some $p' \in \text{supp}\,\tau^*$, and so $0 > v(p') - (V(p_0) + x \cdot (p' - p_0))$. Since v is upper semicontinuous and $V(p_0) + x \cdot (p - p_0)$ is continuous, $v(p) - (V(p_0) + x \cdot (p - p_0))$ is upper semicontinuous. Then there exists a neighborhood M of p' such that $0 > [v(p') - (V(p_0) + x \cdot (p' - p_0))]/2 > v(p) - (V(p_0) + x \cdot (p - p_0))$ for each $p \in M$. Then since $v(p) - (V(p_0) + x \cdot (p - p_0)) = 0$ τ^*-almost everywhere, we must have $\tau^*(M) = 0$, a contradiction since $p' \in \text{supp}\,\tau^*$.

Then $V(p_0) + x \cdot (p - p_0) = V(p) = v(p)$ for all $p \in \text{supp}\,\tau^*$. Now since V is concave and $x \cdot (p - p_0)$ is affine, $V(p) - x \cdot (p - p_0)$ is concave. Then it is also quasiconcave, and so the upper contour set $\{p \mid V(p) - x \cdot (p - p_0) \geq V(p_0)\} \supset \text{supp}\,\tau^*$ is convex. Then $V(p) - x \cdot (p - p_0) \geq V(p_0)$ for each $p \in \text{conv}(\text{supp}\,\tau^*)$; since x is a supergradient of V at p_0, this inequality must bind, and hence V must be affine on $\text{conv}(\text{supp}\,\tau^*)$. QED

PROPOSITION 4 (Additive convexity and monotonicity: general case). Suppose that $S \subseteq \mathbb{R}^n$, that $v_0, v_1 : S \to \mathbb{R}$ are upper semicontinuous, and that v_0 is additively more concave than v_1. Then for any $p_0 \in \text{ri}(S)$ and any solutions to the persuasion problems

$$\tau_0^* \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v_0(p) \text{ such that } E_\tau p = p_0\},$$
$$\tau_1^* \in \arg\max_{\tau \in \Delta(S)} \{E_\tau v_1(p) \text{ such that } E_\tau p = p_0\},$$

we have $\text{supp}\,\tau_1^* \cap \text{conv}(\text{supp}\,\tau_0^*) \subseteq \text{ext}(\text{conv}(\text{supp}\,\tau_0^*))$.

[31] See, e.g., theorem B.1.2.1 in Hiriart-Urruty and Lemaréchal 2001.

[32] I rely on proposition 3 in the online appendix of Kamenica and Gentzkow (2011) rather than the concavification result from their main text because the revelation principle argument they use there does not necessarily apply in my setting.

Proof. Denote the concavifications of v_0 and v_1 by V_0 and V_1, respectively, and let v_Δ be the continuous, strictly concave function such that $v_0 = v_1 + v_\Delta$. Suppose that there exists $p' \in (\text{supp } \tau_1^* \cap \text{conv}(\text{supp } \tau_0^*))\backslash\text{ext}(\text{conv}(\text{supp } \tau_0^*)) \subset \text{conv}(\text{supp } \tau_0^*)$. Since τ_0^* is a Borel measure on S, supp τ_0^* is closed in S; then since S is compact, supp τ_0^* is compact, and since $S \subseteq \mathbb{R}^n$, conv(supp τ_0^*) is also compact. Then by Minkowski's theorem (e.g., theorem A.2.3.4 in Hiriart-Urruty and Lemaréchal 2001), since it is an element of conv(supp τ_0^*), p' can be written as a convex combination of extreme points of conv(supp τ_0^*): there exist $\{p_i\}_{i=1}^d \subseteq \text{ext}(\text{conv}(\text{supp } \tau_0^*))$ and $\{\lambda_i\}_{i=1}^d \subset (0, 1]$ with $\Sigma_{i=1}^d \lambda_i = 1$ such that $p' = \Sigma_{i=1}^d \lambda_i p_i$.

Now recall that $\text{ext}(\text{conv}(\text{supp } \tau_0^*)) \subseteq \text{supp } \tau_0^*$ (e.g., corollary 18.3.1 in Rockafellar 1997). By lemma 3, V_0 coincides with v_0 on supp τ_0^* and is affine on conv(supp τ_0^*); it follows that

$$\sum_{i=1}^d \lambda_i v_0(p_i) = \sum_{i=1}^d \lambda_i V_0(p_i) = V_0(p') \geq v_0(p'). \tag{B2}$$

Since $p' \notin \text{ext}(\text{conv}(\text{supp } \tau_0^*))$ and $\{p_i\}_{i=1}^d \subseteq \text{ext}(\text{conv}(\text{supp } \tau_0^*))$, we must have $p' \neq p_i$, $i \in \{1, \ldots, d\}$; then since v_Δ is strictly concave, $\Sigma_{i=1}^d \lambda_i v_\Delta(p_i) < v_\Delta(p')$. Hence,

$$\sum_{i=1}^d \lambda_i v_1(p_i) = \sum_{i=1}^d \lambda_i v_0(p_i) - \sum_{i=1}^d \lambda_i v_\Delta(p_i) > v_0(p') - v_\Delta(p') = v_1(p'). \tag{B3}$$

Since $p' \in \text{supp } \tau_1^*$, $V_1(p') = v_1(p')$ by lemma 3. And since, by definition, $V_1(p) \geq v_1(p)$ for each $p \in S$,

$$\sum_{i=1}^d \lambda_i V_1(p_i) \geq \sum_{i=1}^d \lambda_i v_1(p_i) > v_1(p') = V_1(p'),$$

a contradiction since V_1 is concave. QED

Proof of proposition 1 (additive convexity and monotonicity).[33] For $i \in \{0, 1\}$, let $\bar{v}_i$ be the upper semicontinuous hull of v_i, that is, $\bar{v}_i(p) \equiv \max\{z | (p, z) \in \text{cl}(\text{hypo}(v))\}$. By lemma S.4 (available online), $\bar{v}_1 + v_\Delta = \overline{v_1 + v_\Delta} = \bar{v}_0$. Moreover, by lemma S.3 (available online),

$$\tau_0^* \in \underset{\tau \in \Delta([0,1])}{\arg\max}\{E_\tau \bar{v}_0(p) \text{ such that } E_\tau p = p_0\},$$
$$\tau_1^* \in \underset{\tau \in \Delta([0,1])}{\arg\max}\{E_\tau \bar{v}_1(p) \text{ such that } E_\tau p = p_0\}.$$

Since τ_0^* is a Borel measure on [0,1], supp τ_0^* is closed in [0,1]; then since [0,1] is a compact subset of $\mathbb{R}$, so is supp τ_0^*. Then min supp τ_0^* and max supp τ_0^* exist, and hence $\text{conv}(\text{supp } \tau_0^*) = [\min \text{ supp } \tau_0^*, \max \text{ supp } \tau_0^*]$. Then by proposition 4, $\text{supp } \tau_1^* \subseteq [0, \min \text{ supp } \tau_0^*] \cup [\max \text{ supp } \tau_0^*, 1]$. Then τ_1^* is a mean-preserving spread of τ_0^* in the sense of Machina and Pratt (1997): $\tau_1^*(I) \geq \tau_0^*(I)$ for each interval $I \subseteq [0, \min \text{ supp } \tau_0^*)$ and each interval $I \subseteq (\max \text{ supp } \tau_0^*, 1]$, and $\tau_1^*(I) \leq \tau_0^*(I)$ for each interval $I \subseteq (\min \text{ supp } \tau_0^*, \max \text{ supp } \tau_0^*)$. It follows from Machina

[33] The proof of proposition 1 relies on two lemmas (S.3 and S.4) in the online appendix to drop the assumption of upper semicontinuity from proposition 4. While this is not used in theorems 1 or 2, since the persuasion problems facing the principal each have upper semicontinuous value functions, it allows theorem 4 to rule out implementation of non-Blackwell-monotone choice functions with any results-based contract, not just an upper semicontinuous one.

and Pratt's (1997) theorem 3 that τ_1^* is a mean-preserving spread of τ_0^* in the usual (stochastic dominance) sense and hence is Blackwell more informative than τ_0^*. QED

Proof of proposition 2 (a necessary and sufficient secant line condition). (Only if) Without loss, assume that $\bar{p} > \underline{p}$. By lemma 3, V is affine on conv(supp τ^*) $= [\underline{p}, \bar{p}]$ and coincides with v at $\underline{p}$ and $\bar{p}$. Then $V(p) = v(\underline{p}) + [(v(\bar{p}) - v(\underline{p}))/(\bar{p} - \underline{p})](p - \underline{p})$ for each $p \in [\underline{p}, \bar{p}]$. By lemma 2, $p_0 \in (\underline{p}, \bar{p})$; then V is differentiable at p_0 with derivative $V'(p_0) = (v(\bar{p}) - v(\underline{p}))/(\bar{p} - \underline{p})$. Since V is concave and $V'(p_0)$ exists, it is the unique superderivative of V at $p_0 : V(p_0) + [(v(\bar{p}) - v(\underline{p}))/(\bar{p} - \underline{p})](p - p_0) \geq V(p)$ for each $p \in [0, 1]$. Writing $V(p_0) = v(\underline{p}) + [(v(\bar{p}) - v(\underline{p}))/(\bar{p} - \underline{p})](p_0 - \underline{p})$ and noting that $V(p) \geq v(p)$ by definition yields (9), as desired.

(If) For any Bayes-plausible $\tau \in \Delta([0, 1])$, taking expectations over either side of the secant line inequality (9) yields

$$E_\tau v(p) \leq E_\tau\left(v(\underline{p}) + \frac{v(\bar{p}) - v(\underline{p})}{\bar{p} - \underline{p}}(p - \underline{p})\right) = v(\underline{p}) + \frac{v(\bar{p}) - v(\underline{p})}{\bar{p} - \underline{p}}(p_0 - \underline{p})$$
$$= E_{\tau^*} v(p),$$

by lemma 2, and so $\tau^* \in \arg\max_{\tau \in \Delta([0,1])}\{E_\tau v(p)$ such that $E_\tau p = p_0\}$. QED

B2. *Methods-Based Contracting*

LEMMA 4 (Reducing the principal's problem).

i. If $\chi : \Theta \to X$ satisfies the monotonicity constraints (5) and $t : \Theta \to \mathbb{R}$ is given by (2) and (3), then (χ, t) satisfies (ICθ) and (IRθ).
ii. For any direct revelation contract (χ, t) that satisfies (ICθ) and (IRθ),
   a. $E_{\chi(\theta_{i-1})}[c(p)] \leq E_{\chi(\theta_i)}[c(p)]$ for all $i \in \{2, \ldots, N\}$; and
   b. if $\hat{t}$ is derived from χ according to (2) and (3), then $t(\theta) \geq \hat{t}(\theta)$ for each $\theta \in \Theta$; moreover, $(\chi, \hat{t})$ satisfies (ICθ) and (IRθ).

Proof.

i. For each $i \in \{2, \ldots, N\}$ and $\ell < i$, (2) and (3) yield

$$t(\theta_i) - C(\chi(\theta_i), \theta_i) - (t(\theta_\ell) - C(\chi[\theta_\ell], \theta_i))$$
$$= \sum_{j=\ell}^{i-1}(\theta_j - \theta_{j+1})E_{\chi[\theta_j]}[c(p)] - (\theta_\ell - \theta_i)E_{\chi[\theta_\ell]}[c(p)]$$
$$= \sum_{j=\ell}^{i-1}(\theta_j - \theta_{j+1})\left(E_{\chi[\theta_j]}[c(p)] - E_{\chi[\theta_\ell]}[c(p)]\right) \geq 0,$$

(by [5]) and for each $i \in \{1, \ldots, N-1\}$ and $\ell > i$,

$$t(\theta_i) - C(\chi(\theta_i), \theta_i) - (t(\theta_\ell) - C(\chi[\theta_\ell], \theta_i))$$
$$= (\theta_i - \theta_\ell)E_{\chi[\theta_\ell]}[c(p)] - \sum_{j=i}^{\ell-1}(\theta_j - \theta_{j+1})E_{\chi[\theta_j]}[c(p)]$$
$$= \sum_{j=i}^{\ell-1}(\theta_j - \theta_{j+1})\left(E_{\chi[\theta_\ell]}[c(p)] - E_{\chi[\theta_j]}[c(p)]\right) \geq 0,$$

(by [5]), implying (ICθ). From (2), $t(\theta_1) - C(\chi(\theta_1), \theta_1) = 0$; by Jensen's inequality, since c is convex, $E_\tau[c(p)] \geq 0$ for any $\tau \in X$. Then for each $i \in \{1, \ldots, N\}$,

$$\begin{aligned} t(\theta_i) - C(\chi(\theta_i), \theta_i) &= t(\theta_i) - C(\chi(\theta_i), \theta_i) - (t(\theta_1) - C(\chi(\theta_1), \theta_1)) \\ &= t(\theta_i) - C(\chi(\theta_i), \theta_i) - (t(\theta_1) - C(\chi(\theta_1), \theta_i)) \\ &\quad + (\theta_1 - \theta_i) E_{\chi(\theta_1)}[c(p)] \\ &\geq 0, \end{aligned}$$

(by [ICθ]), implying (IRθ).

ii.

a. From (ICθ), for each $i \in \{2, \ldots, N\}$,

$$\begin{aligned} \theta_{i-1}(E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)]) &\geq t(\theta_i) - t(\theta_{i-1}), \\ \theta_i(E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)]) &\leq t(\theta_i) - t(\theta_{i-1}); \\ \Rightarrow (\theta_{i-1} - \theta_i)(E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)]) &\geq 0; \\ \Leftrightarrow E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)] &\geq 0, \end{aligned}$$

as desired.

b. Since $\hat{t}$ is derived from (2) and (3), we have $\hat{t}(\theta_1) = C(\chi(\theta_1), \theta_1)$, and for each $i \in \{2, \ldots, N\}$, $\hat{t}(\theta_i) - \hat{t}(\theta_{i-1}) = \theta_i(E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)]) = C(\chi(\theta_i), \theta_i) - C(\chi(\theta_{i-1}), \theta_i)$. Since (χ, t) satisfies (IRθ) for $\theta = \theta_1$, we have $t(\theta_1) \geq C(\chi(\theta_1), \theta_1) = \hat{t}(\theta_1)$. For each $i \in \{2, \ldots, N\}$, since (χ, t) satisfies (ICθ) for $\theta = \theta_i$ and $\theta' = \theta_{i-1}$, we have $t(\theta_i) - t(\theta_{i-1}) \geq C(\chi(\theta_i), \theta_i) - C(\chi(\theta_{i-1}), \theta_i) = \hat{t}(\theta_i) - \hat{t}(\theta_{i-1})$. Then for each $j \in \{2, \ldots, N\}$, we have

$$t(\theta_j) = t(\theta_1) + \sum_{i=2}^{N}(t(\theta_i) - t(\theta_{i-1})) \geq \hat{t}(\theta_1) + \sum_{i=2}^{N}(\hat{t}(\theta_i) - \hat{t}(\theta_{i-1})) = \hat{t}(\theta_j).$$

Finally, from parts iia and i, $(\chi, \hat{t})$ satisfies (ICθ) and (IRθ). QED

Corollary 1. (χ^*, t^*) solves the principal's problem if and only if χ^* solves the reduced problem (4) and t^* is given by (2) and (3).

Proof. By lemma 4, part ii, when solving the principal's problem, we need only consider (χ, t) such that χ satisfies the monotonicity constraints (5) and t is given by (2) and (3). By lemma 4, part i, any χ satisfies the monotonicity constraints (5) and t is given by (2) and (3) also satisfies (ICθ) and (IRθ). It follows that replacing the incentive constraints with (2), (3), and the monotonicity constraints (5), as in the reduced problem (4), does not change the set of solutions to the principal's problem. QED

Lemma 5. $R(F(\theta_i)) = -\theta_i F(\theta_i)$ for each $i \in \{1, \ldots, N\}$.

Proof. We have

$$
\begin{aligned}
R(F(\theta_i)) = -\sum_{j=i}^{N} g(\theta_j) f(\theta_j) &= -\left(\sum_{j=i}^{N} \theta_j f(\theta_j)\right) - \left(\sum_{j=i}^{N} \sum_{\ell > j} f(\theta_\ell)(\theta_j - \theta_{j+1})\right) \\
&= -\left(\sum_{j=i}^{N} \theta_j f(\theta_j)\right) - \left(\sum_{\ell=i+1}^{N} \sum_{j=i}^{\ell-1} f(\theta_\ell)(\theta_j - \theta_{j+1})\right) \\
&= -\left(\sum_{j=i}^{N} \theta_j f(\theta_j)\right) - \left(\sum_{\ell=i+1}^{N} f(\theta_\ell)(\theta_i - \theta_\ell)\right) \\
&= -\left(\sum_{j=i}^{N} \theta_i f(\theta_j)\right) = -\theta_i F(\theta_i),
\end{aligned}
$$

as desired. QED

Lemma 6. If $P(F(\theta_i)) > R(F(\theta_i))$, then P is affine on $[F(\theta_{i+1}), F(\theta_{i-1})]$ if $i < N$ or on $[F(0), F(\theta_{N-1})]$ if $i = N$.

Proof. First note that R is upper semicontinuous: observe that F^{-1}, and therefore R, is constant on each of the intervals $(0, F(\theta_N)]$ and $(F(\theta_{i+1}), F(\theta_i)]$, $i \in \{1, \dots, N-1\}$. Then since R is decreasing, for each $q \in (0, 1]$, we have $\limsup_{y \to q} R(y) = R(q)$; and further, $\limsup_{y \to 0} R(y) = R(F(\theta_N)) \le 0 = R(0)$.

Now suppose $P(F(\theta_i)) > R(F(\theta_i))$ for $i \in \{2, \dots, N\}$. Since $F(\theta_i) \in (0, 1)$ and R is upper semicontinuous, by lemma 2 and Kamenica and Gentzkow's (2011) propositions 3 and 4 (in their online appendix), there exists $\tau^* \in \arg\max\{E_\tau R(q) \text{ such that } E_\tau q = F(\theta_i)\}$ such that $\operatorname{supp} \tau^* = \{q', q''\}$ with $q' < F(\theta_i) < q''$. Then by lemma 3, P is affine on $[q', q'']$ and coincides with R at q' and q''.

Further, if $i < N$, we must have $q' \le F(\theta_{i+1})$, and if $i = N$, we must have $q' = 0$. Otherwise, $R(q') = R(F(\theta_i))$, and since R is decreasing, $R(q'') \le R(F(\theta_i))$, and we have $P(F(\theta_i)) = \tau^*(\{q'\})R(q') + \tau^*(\{q''\})R(q') \le R(F(\theta_i))$, a contradiction.

Finally, we must have $q'' \ge F(\theta_{i-1})$: suppose instead that $q'' < F(\theta_{i-1})$. Then $R(q'') = R(F(\theta_{i-1}))$, and we have

$$
\begin{aligned}
P(F(\theta_i)) &= R(q') + \frac{R(F(\theta_{i-1})) - R(q')}{q'' - q'}(F(\theta_i) - q') \\
&< R(q') + \frac{R(F(\theta_{i-1})) - R(q')}{F(\theta_{i-1}) - q'}(F(\theta_i) - q') \\
&= \frac{F(\theta_i) - q'}{F(\theta_{i-1}) - q'} R(F(\theta_{i-1})) + \frac{F(\theta_{i-1}) - F(\theta_i)}{F(\theta_{i-1}) - q'} R(q') \\
&\in \{z \mid (F(\theta_i), z) \in \operatorname{conv}(\operatorname{Gr}(R))\},
\end{aligned}
$$

a contradiction.

It follows that P is affine on $[F(\theta_{i+1}), F(\theta_{i-1})]$ if $i < N$ or on $[F(0), F(\theta_{N-1})]$ if $i = N$. QED

Proof of lemma 1 (bunching and monotonicity in the ironed problem).

i. For $i < N$, we have $\bar{g}(\theta_i) = -([P(F(\theta_i)) - P(F(\theta_{i+1}))]/[F(\theta_i) - F(\theta_{i+1})])$ and $\bar{g}(\theta_{i-1}) = -([P(F(\theta_{i-1})) - P(F(\theta_i))]/[F(\theta_{i-1}) - F(\theta_i)])$. Likewise, we have $\bar{g}(\theta_N) = -([P(F(\theta_N)) - P(F(0))]/[F(\theta_N) - F(0)])$ and $\bar{g}(\theta_{N-1}) = -([P(F(\theta_{N-1})) - P(F(\theta_N))]/[F(\theta_{N-1}) - F(\theta_N)])$. In either case, since $F(\theta_{i-1}) > F(\theta_i) > F(\theta_{i+1})$ for $i < N$, and $F(\theta_{N-1}) > F(\theta_N) > F(0) = 0$, it follows from concavity of P that $\bar{g}(\theta_{i-1}) \geq \bar{g}(\theta_i)$.[34] Now observe that the ironed problem (A1) is separable into type-specific Bayesian persuasion problems. Then whenever the value function in the type $\theta_{i\text{———}1}$ problem is distinct from that in the type θ_i problem, it is additively more concave, since it differs by the continuous, strictly concave function $(\bar{g}(\theta_i) - \bar{g}(\theta_{i-1}))c(p)$; the rest of the claim then follows from proposition 1.

ii. By lemma 6, $\bar{g}(\theta_i) = -([P(F(\theta_i)) - P(F(\theta_{i+1}))]/[F(\theta_i) - F(\theta_{i+1})]) = -([P(F(\theta_{i-1})) - P(F(\theta_i))]/[F(\theta_{i-1}) - F(\theta_i)]) = \bar{g}(\theta_{i-1})$, if $i < N$, or $\bar{g}(\theta_N) = -([P(F(\theta_N)) - P(F(0))]/[F(\theta_N) - F(0)]) = -([P(F(\theta_{N-1})) - P(F(\theta_N))]/[F(\theta_{N-1}) - F(\theta_N)]) = \bar{g}(\theta_{N-1})$, if $i = N$. Since the ironed problem (A1) is separable into type-specific Bayesian persuasion problems, and those problems have upper semicontinuous value functions (by continuity of H and upper semicontinuity of w), the rest of the claim then follows from Kamenica and Gentzkow (2011)'s proposition 4 (in their online appendix), choosing the same solution for the type θ_i and type θ_{i-1} persuasion problems whenever $P(F(\theta_i)) > R(F(\theta_i))$.

iii. Suppose that $\bar{g}(\theta_N) = \bar{g}(\theta_{N-1})$, and let $\bar{n} = \min\{i \in \{1, \ldots, N\} \mid \bar{g}(\theta_N) = \bar{g}(\theta_i)\}$; then $\bar{n} \leq N - 1$. If $\bar{n} \geq 2$, we must have $\bar{g}(\theta_{\bar{n}}) \neq \bar{g}(\theta_{\bar{n}-1})$, and so by part ii, $P(F(\theta_{\bar{n}})) = R(F(\theta_{\bar{n}}))$; if $\bar{n} = 1$, $P(F(\theta_1)) = P(1) = R(1) = R(F(\theta_1))$ by definition of concavification. Now by definition of concavification, $P(0) = R(0) = 0$; then we can write $P(F(\theta_i)) = -\Sigma_{j=i}^{N} f(\theta_j)\bar{g}(\theta_j)$. Then by lemma 5, we have

$$-\theta_{\bar{n}}F(\theta_{\bar{n}}) = R(F(\theta_{\bar{n}})) = P(F(\theta_{\bar{n}})) = -\sum_{j=\bar{n}}^{N} f(\theta_j)\bar{g}(\theta_j) = -F(\theta_{\bar{n}})\bar{g}(\theta_N)$$

$$\Rightarrow \theta_{\bar{n}} = \bar{g}(\theta_N) = -\frac{P(F(\theta_N))}{F(\theta_N)} \leq -\frac{R(F(\theta_N))}{F(\theta_N)}$$

$$= \theta_N,$$

a contradiction since $\theta_{\bar{n}} \geq \theta_{N-1} > \theta_N$. Then $\bar{g}(\theta_N) < \bar{g}(\theta_{N-1})$; the rest follows from part ii and the definition of $\bar{g}(\theta_N)$ and $g(\theta_N)$. QED

Lemma 7. Let $\chi^* : \Theta \to X$ be an experiment choice function.

i. If χ^* solves the principal's reduced problem (4), then

$$E_F\left[E_{\chi^*(\theta)}[w(p) - \bar{g}(\theta)c(p)]\right] \geq E_F\left[E_{\chi^*(\theta)}[w(p) - g(\theta)c(p)]\right].$$

ii. If χ^* solves the ironed problem (A1), and $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for each $i \in \{2, \ldots, N\}$ with $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$, then

$$E_F\left[E_{\chi^*(\theta)}[w(p) - \bar{g}(\theta)c(p)]\right] = E_F\left[E_{\chi^*(\theta)}[w(p) - g(\theta)c(p)]\right].$$

[34] See, e.g., proposition 6.1 in Hiriart-Urruty and Lemaréchal 2001.

Proof. By definition of concavification, we must have $P(F(\theta_1)) = P(1) = R(1) = R(F(\theta_1))$ and $P(0) = R(0) = 0$. Moreover, $R(F(\theta_N)) = -g(\theta_N)f(\theta_N)$, and for each $i \in \{1, \dots, N-1\}$, we have $R(F(\theta_{i+1})) - R(F(\theta_i)) = g(\theta_i)f(\theta_i)$. For any χ, we have

$$
\begin{aligned}
E_F[E_{\chi(\theta)}[w(p) - \bar{g}(\theta)c(p)]] &- E_F[E_{\chi(\theta)}[w(p) - g(\theta)c(p)]] \\
&= \sum_{i=1}^{N} f(\theta_i)(g(\theta_i) - \bar{g}(\theta_i))E_{\chi(\theta_i)}[c(p)] \\
&= \sum_{i=1}^{N-1} (R(F(\theta_{i+1})) - R(F(\theta_i)) - (P(F(\theta_{i+1})) - P(F(\theta_i))))E_{\chi(\theta_i)}[c(p)] \\
&\quad + (P(F(\theta_N)) - R(F(\theta_N)))E_{\chi(\theta_N)}[c(p)] \\
&= \sum_{i=2}^{N} (P(F(\theta_i)) - R(F(\theta_i)))(E_{\chi(\theta_i)}[c(p)] - E_{\chi(\theta_{i-1})}[c(p)]). \qquad \text{(B4)}
\end{aligned}
$$

i. By definition, $P(q) \geq R(q)$ for each $q \in [0, 1]$. If χ^* solves the principal's reduced problem (4), χ^* must satisfy the monotonicity constraints (5). Then each term of the sum in (B4) must be weakly positive; the claim follows.
ii. Suppose that χ^* solves the ironed problem (A1). By lemma 1, part ii, for each $i \in \{2, \dots, N\}$, either $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$, and so by assumption $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$, or (since $R(F(\theta_i)) \leq P(F(\theta_i))$ by definition) $R(F(\theta_i)) = P(F(\theta_i))$. The claim then follows from (B4). QED

Proof of proposition 3 (ironing the principal's problem). Suppose that χ^* solves the ironed problem (A1); $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for each $i \in \{2, \dots, N\}$, with $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$; and that $\hat{\chi}$ solves the reduced problem (4). Then we have

$$
\begin{aligned}
E_F\left[E_{\chi^*(\theta)}[w(p) - g(\theta)c(p)]\right] &\leq E_F[E_{\hat{\chi}(\theta)}[w(p) - g(\theta)c(p)]] && \text{(B5)} \\
&\leq E_F[E_{\hat{\chi}(\theta)}[w(p) - \bar{g}(\theta)c(p)]] \text{ (by lemma 7, part i)} && \text{(B6)} \\
&\leq E_F\left[E_{\chi^*(\theta)}[w(p) - \bar{g}(\theta)c(p)]\right] && \text{(B7)} \\
&= E_F\left[E_{\chi^*(\theta)}[w(p) - g(\theta)c(p)]\right] \text{ (by lemma 7, part ii).} && \text{(B8)}
\end{aligned}
$$

Hence, the expectations on the right-hand sides of (B5)–(B8) are equal. Furthermore, by lemma 1, part i, χ^* is Blackwell monotone and so satisfies the monotonicity constraints (5). Then χ^* solves the reduced problem (4) (by equality of [B5] and [B8]) and $\hat{\chi}$ solves the ironed problem (A1) (by equality of [B6] and [B7]). QED

Proof of theorem 1 (optimal methods-based contracts). By lemma 1, part ii, there is a binary χ^* that solves the principal's ironed problem (A1) such that $\chi^*(\theta_i) = \chi^*(\theta_{i-1})$ for each $i \in \{2, \dots, N\}$, with $\bar{g}(\theta_i) = \bar{g}(\theta_{i-1})$. By lemma 1, part i, χ^* is Blackwell monotone. By proposition 3, χ^* solves the principal's reduced problem (4). The claim then follows immediately from corollary 1. QED

LEMMA 8 (Reducing the principal's full-information problem). (χ, t) solves the principal's full-information optimal contracting problem ((1) without (ICθ) if and only if $\chi(\theta) \in \arg\max_\tau \{E_\tau[w(p) - \theta c(p)]$ such that $E_\tau p = p_0\}$ and $t(\theta) = C(\chi(\theta), \theta)$ for each $\theta \in \Theta$.

Proof. Clearly, at any solution (χ, t) to the principal's full-information optimal contracting problem ([1] without [ICθ]), (IRθ) must bind for each θ. Otherwise, the value of the principal's objective can be increased by setting $t(\theta) = C(\chi(\theta), \theta)$ for each θ. The claim then follows immediately. QED

Proof of theorem 2 (distortion from hidden information).

i. By lemma 1, part iii, $\theta_N = g(\theta_N) = \bar{g}(\theta_N)$; the claim then follows from proposition 3, corollary 1, and lemma 8.

ii. We first show that $\bar{g}(\theta_i) > \theta_i$ for each $i \in \{1, \ldots, N-1\}$. If $\bar{g}(\theta_i) = g(\theta_i)$, this follows from the definition of $g(\theta_i)$. Suppose $\bar{g}(\theta_i) = ([P(F(\theta_{i+1})) - P(F(\theta_i))]/f(\theta_i)) \neq g(\theta_i) = ([R(F(\theta_{i+1})) - R(F(\theta_i))]/f(\theta_i))$. Then either $P(F(\theta_{i+1})) > R(F(\theta_{i+1}))$ or $P(F(\theta_i)) > R(F(\theta_i))$. Let $\bar{n} = \min\{j > i \mid P(F(\theta_j)) = R(F(\theta_j))\}$ and $\underline{n} = \max\{j \leq i \mid P(F(\theta_j)) = R(F(\theta_j))\}$; $\underline{n}$ exists since $P(F(\theta_1)) = P(1) = R(1) = R(F(\theta_1))$ by definition of concavification, while $\bar{n}$ exists since $P(F(\theta_N)) = R(F(\theta_N))$ by lemma 1, part iii. Then we have $\underline{n} \leq i \leq \bar{n} - 1$. By lemma 6, P is affine on $[F(\theta_{\bar{n}}), F(\theta_{\underline{n}})]$. It follows that

$$
\begin{aligned}
\bar{g}(\theta_i) &= -\frac{P(F(\theta_i)) - P(F(\theta_{i+1}))}{F(\theta_i) - F(\theta_{i+1})} = -\frac{P(F(\theta_{\underline{n}})) - P(F(\theta_{\bar{n}}))}{F(\theta_{\underline{n}}) - F(\theta_{\bar{n}})} = \\
&\quad -\frac{R(F(\theta_{\underline{n}})) - R(F(\theta_{\bar{n}}))}{F(\theta_{\underline{n}}) - F(\theta_{\bar{n}})} \\
&= \frac{\theta_{\underline{n}} F(\theta_{\underline{n}}) - \theta_{\bar{n}} F(\theta_{\bar{n}})}{F(\theta_{\underline{n}}) - F(\theta_{\bar{n}})} \quad \text{(by lemma 5)} \\
&= \theta_{\underline{n}} + \frac{(\theta_{\underline{n}} - \theta_{\bar{n}}) F(\theta_{\bar{n}})}{F(\theta_{\underline{n}}) - F(\theta_{\bar{n}})} > \theta_{\underline{n}} \geq \theta_i.
\end{aligned}
$$

Then for each $i \in \{1, \ldots, N-1\}$, $(\theta_i - \bar{g}(\theta_i))c(p)$ is a continuous and strictly concave function of p, and so the claim follows from proposition 3, corollary 1, lemma 8, and proposition 1.

iii. $C(\chi^*(\theta_i), \theta_i) \leq C(\tilde{\chi}(\theta_i), \theta_i)$ follows directly from part ii. Now suppose that $W(\chi^*(\theta_i)) > W(\tilde{\chi}(\theta_i))$. Then $E_{\chi^*(\theta_i)}[w(p) - \theta_i c(p)] = W(\chi^*(\theta_i)) - C(\chi^*(\theta_i), \theta_i) > W(\tilde{\chi}(\theta_i)) - C(\tilde{\chi}(\theta_i), \theta_i) = E_{\tilde{\chi}(\theta_i)}[w(p) - \theta_i c(p)]$, a contradiction since $\tilde{\chi}(\theta_i) \in \arg\max_\tau \{E_\tau[w(p) - \theta_i c(p)]$ such that $E_\tau p = p_0\}$ by lemma 8. QED

B3. Results-Based Contracting

Proof of theorem 3 (results-based implementation). Suppose χ is binary, interior, and Blackwell monotone. Let $\lambda = \min\{i \in \{1, \ldots, N\} \mid \operatorname{supp} \chi(\theta_i) \neq \{p_0\}\}$. Then since χ is binary and Blackwell monotone, for each $i \in \{\lambda, \ldots, N\}$, $\operatorname{supp} \chi(\theta_i) = \{\underline{y}_i, \bar{y}_i\}$, and by lemma 2, $\underline{y}_i \leq \underline{y}_{i-1} < p_0 < \bar{y}_{i-1} \leq \bar{y}_i$ for each $i \in \{\lambda + 1, \ldots, N\}$.

Now define $s_\lambda(p) \equiv \Delta_\lambda(p - p_0)$, where $\Delta_\lambda \equiv \theta_\lambda H'(p_0)$, and for each $i \in \{\lambda + 1, \ldots, N\}$, define s_i recursively as

$$s_i(p) \equiv s_{i-1}(\underline{y}_{i-1}) + (\theta_{i-1} - \theta_i)c(\underline{y}_{i-1}) + \Delta_i(p - \underline{y}_{i-1}),$$

$$\text{where } \Delta_i \equiv \frac{(s_{i-1}(\bar{y}_{i-1}) + (\theta_{i-1} - \theta_i)c(\bar{y}_{i-1})) - (s_{i-1}(\underline{y}_{i-1}) + (\theta_{i-1} - \theta_i)c(\underline{y}_{i-1}))}{\bar{y}_{i-1} - \underline{y}_{i-1}}.$$

Then define ψ_χ as

$$\psi_\chi(p) \equiv \begin{cases} s_\lambda(p) + \theta_\lambda c(p), & p \in [\underline{y}_\lambda, \bar{y}_\lambda]; \\ s_i(p) + \theta_i c(p), & p \in [\underline{y}_i, \underline{y}_{i-1}) \cup (\bar{y}_{i-1}, \bar{y}_i], \forall\ i \in \{\lambda + 1, \dots, N\}; \\ s_N(p) + \theta_N c(p), & p \in [0, \underline{y}_N) \cup (\bar{y}_N, 1]. \end{cases}$$

Note that for each $i \in \{\lambda, \dots, N\}$, $s_i(p)$ coincides with the secant line through $(\underline{y}_i, \psi_\chi(\underline{y}_i) - \theta_i c(\underline{y}_i))$ and $(\bar{y}_i, \psi_\chi(\bar{y}_i) - \theta_i c(\bar{y}_i))$: we have

$$\begin{aligned} \psi_\chi(\underline{y}_i) - \theta_i c(\underline{y}_i) + \frac{(\psi_\chi(\bar{y}_i) - \theta_i c(\bar{y}_i)) - (\psi_\chi(\underline{y}_i) - \theta_i c(\underline{y}_i))}{\bar{y}_i - \underline{y}_i}(p - \underline{y}_i) &= \psi_\chi(\underline{y}_i) - \theta_i c(\underline{y}_i) + \frac{s_i(\bar{y}_i) - s_i(\underline{y}_i)}{\bar{y}_i - \underline{y}_i}(p - \underline{y}_i) \\ &= \psi_\chi(\underline{y}_i) - \theta_i c(\underline{y}_i) + \Delta_i(p - \underline{y}_i) \qquad \text{(B9)} \\ &= s_i(p). \qquad \text{(B10)} \end{aligned}$$

ψ_χ is continuous.—First note that $\{s_i\}_{i=\lambda}^N$ are linear, hence continuous. By definition, we have $s_i(\underline{y}_{i-1}) + \theta_i c(\underline{y}_{i-1}) = s_{i-1}(\underline{y}_{i-1}) + \theta_{i-1} c(\underline{y}_{i-1})$ and $s_i(\bar{y}_{i-1}) + \theta_i c(\bar{y}_{i-1}) = s_{i-1}(\bar{y}_{i-1}) + \theta_{i-1} c(\bar{y}_{i-1})$ for each $i \in \{\lambda + 1, \dots, N\}$; since H (and thus c) is continuous, it follows that ψ_χ is continuous.

ψ_χ satisfies the global incentive compatibility condition (8) for $\theta \leq \theta_\lambda$.—Since ψ_χ is upper semicontinuous, so is $\psi_\chi(p) - \theta c(p)$ for each θ. Then by (B10) and proposition 2, $\chi(\theta_i) \in \arg\max_\tau\{E_\tau[\psi_\chi(p) - \theta_i c(p)]$ such that $E_\tau p = p_0\}$ if (i) $\psi_\chi(p) - \theta_i c(p) \leq s_i(p)$ for each $p \in [\underline{y}_j, \underline{y}_{j-1})$, $j \in \{\lambda + 1, \dots, N\}$, and for each $p < \underline{y}_N$; (ii) $\psi_\chi(p) - \theta_i c(p) \leq s_i(p)$ for each $p \in (\bar{y}_{j-1}, \bar{y}_j]$, $j \in \{\lambda + 1, \dots, N\}$, and for each $p > \bar{y}_N$; and (iii) $\psi_\chi(p) - \theta_i c(p) \leq s_i(p)$ for each $p \in [\underline{y}_\lambda, \bar{y}_\lambda]$. To see that part i holds, first note that for each $k \in \{\lambda + 1, \dots, N\}$, we have

$$\begin{aligned} \Delta_k &= \frac{c(\bar{y}_{k-1}) - c(\underline{y}_{k-1})}{\bar{y}_{k-1} - \underline{y}_{k-1}}(\theta_{k-1} - \theta_k) + \frac{s_{k-1}(\bar{y}_{k-1}) - s_{k-1}(\underline{y}_{k-1})}{\bar{y}_{k-1} - \underline{y}_{k-1}} \\ &= \frac{c(\bar{y}_{k-1}) - c(\underline{y}_{k-1})}{\bar{y}_{k-1} - \underline{y}_{k-1}}(\theta_{k-1} - \theta_k) + \Delta_{k-1}. \qquad \text{(B11)} \end{aligned}$$

Hence, for each $k \in \{\lambda + 1, \dots, N\}$,

$$\begin{aligned} s_k(p) + \theta_k c(p) &= s_{k-1}(\underline{y}_{k-1}) + (\theta_{k-1} - \theta_k)c(\underline{y}_{k-1}) + \Delta_k(p - \underline{y}_{k-1}) + \theta_k c(p) \\ &= (\theta_{k-1} - \theta_k)\left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1})\right) \\ &\quad + s_{k-1}(\underline{y}_{k-1}) + \Delta_{k-1}(p - \underline{y}_{k-1}) + \theta_{k-1} c(p) \\ &= (\theta_{k-1} - \theta_k)\left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1})\right) \\ &\quad + s_{k-1}(p) + \theta_{k-1} c(p) \qquad \text{(B12)} \\ &= (\theta_{k-1} - \theta_k)\frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}}\left(c(\underline{y}_{k-1}) - \frac{\underline{y}_{k-1} - p}{\bar{y}_{k-1} - p} c(\bar{y}_{k-1}) - \frac{\bar{y}_{k-1} - \underline{y}_{k-1}}{\bar{y}_{k-1} - p} c(p)\right) \\ &\quad + s_{k-1}(p) + \theta_{k-1} c(p). \qquad \text{(B13)} \end{aligned}$$

It follows from (B13) that when $p \in [\underline{y}_j, \underline{y}_{j-1})$ for $j > i$ or $p < \underline{y}_j$ for $j = N$,

$$\begin{aligned}\psi_\chi(p) - \theta_i c(p) &= s_j(p) + (\theta_j - \theta_i)c(p) \\ &= s_i(p) + \sum_{k=i+1}^{j} (\theta_{k-1} - \theta_k) \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} \left(c(\underline{y}_{k-1}) - \frac{\underline{y}_{k-1} - p}{\bar{y}_{k-1} - p} c(\bar{y}_{k-1}) - \frac{\bar{y}_{k-1} - \underline{y}_{k-1}}{\bar{y}_{k-1} - p} c(p) \right) \\ &\le s_i(p),\end{aligned} \tag{B14}$$

since c is convex and $p < \underline{y}_{j-1} \le \underline{y}_{k-1} < p_0 < \bar{y}_{k-1}$ for each $k \le j$. Furthermore, when $p \in [\underline{y}_j, \underline{y}_{j-1})$ for $\lambda + 1 \le j < i$, it follows from (B12) that

$$\begin{aligned} s_i(p) &= s_j(p) + (\theta_j - \theta_i)c(p) \\ &\quad + \sum_{k=j+1}^{i} (\theta_{k-1} - \theta_k) \left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1}) \right) \\ &= \psi_\chi(p) - \theta_i c(p) \\ &\quad + \sum_{k=j+1}^{i} (\theta_{k-1} - \theta_k) \left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1}) \right) \\ &\ge \psi_\chi(p) - \theta_i c(p), \end{aligned}$$

since c is convex and $\underline{y}_{k-1} \le \underline{y}_j \le p < \underline{y}_{j-1} < p_0 < \bar{y}_{k-1}$ for each $k > j$. Since $\psi_\chi(p) - \theta_i c(p) = s_i(p)$ for $p \in [\underline{y}_i, \underline{y}_{i-1})$ when $i \in \{\lambda + 1, \dots, N\}$, part i follows. Part ii then follows symmetrically. For part iii, note that it follows from (B12) that for $i > \lambda$ and $p \in [\underline{y}_\lambda, \bar{y}_\lambda]$,

$$\begin{aligned} s_i(p) &= s_\lambda(p) + (\theta_\lambda - \theta_i)c(p) \\ &\quad + \sum_{k=\lambda+1}^{i} (\theta_{k-1} - \theta_k) \left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1}) \right) \\ &= \psi_\chi(p) - \theta_i c(p) \\ &\quad + \sum_{k=\lambda+1}^{i} (\theta_{k-1} - \theta_k) \left(\frac{p - \underline{y}_{k-1}}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\bar{y}_{k-1}) - c(p) + \frac{\bar{y}_{k-1} - p}{\bar{y}_{k-1} - \underline{y}_{k-1}} c(\underline{y}_{k-1}) \right) \\ &\ge \psi_\chi(p) - \theta_i c(p), \end{aligned}$$

since c is convex and $\underline{y}_{k-1} \le \underline{y}_\lambda \le p \le \bar{y}_\lambda \le \bar{y}_{k-1}$ for each $k \ge \lambda + 1$. Furthermore, by definition, $s_\lambda(p) = \psi_\chi(p) - \theta_\lambda c(p)$ on $[\underline{y}_\lambda, \bar{y}_\lambda]$; part iii follows.

ψ satisfies the global incentive compatibility condition (8) for $\theta > \theta_\lambda$.—Since $E_{\chi[\theta_\lambda]} p = p_0$, we have $E_{\chi[\theta_\lambda]}[\psi_\chi(p) - \theta_\lambda c(p)] = E_{\chi[\theta_\lambda]}[s_\lambda(p)] = 0$. Since $\chi[\theta_\lambda] \in \arg\max_\tau \{E_\tau[\psi(p) - \theta_\lambda c(p)]$ such that $E_\tau p = p_0\}$, this implies $E_\tau[\psi_\chi(p) - \theta_\lambda c(p)] \le 0$ for each $\tau \in X$. It follows that for each $\theta > \theta_\lambda$ and $\tau \in X$, $E_\tau[\psi_\chi(p) - \theta c(p)] \le 0$, since $E_\tau[c(p)] \ge 0$ by Jensen's inequality. Then for each $\theta > \theta_\lambda$, since $E_{\chi(\theta)}[\psi_\chi(p) - \theta c(p)] = 0$, we have $\chi(\theta) \in \arg\max_\tau \{E_\tau[\psi(p) - \theta c(p)]$ such that $E_\tau p = p_0\}$.

If $\hat{t} : \Theta \to \mathbb{R}$ is derived from χ according to (2) and (3), then $E_{\chi(\theta)} \psi_\chi(p) = \hat{t}(\theta)$ for each $\theta \in \Theta$.—We have $\psi_\chi(p_0) = c(p_0) = 0$. Then for each $\theta > \theta_\lambda$, we have $E_{\chi(\theta)}[\psi_\chi(p) - \theta c(p)] = \psi_\chi(p_0) - \theta c(p_0) = 0 = \hat{t}(\theta)$. Further, since $E_{\chi[\theta_\lambda]} p = p_0$, we have $E_{\chi[\theta_\lambda]}[\psi_\chi(p) - \theta_\lambda c(p)] = E_{\chi[\theta_\lambda]}[s_\lambda(p)] = E_{\chi[\theta_\lambda]}[\theta_\lambda H'(p_0)(p - p_0)] = 0$, and so

$$E_{\chi[\theta_\lambda]}[\psi_\chi(p)] = E_{\chi[\theta_\lambda]}[\theta_\lambda c(p)] = E_{\chi[\theta_\lambda]}[\theta_\lambda c(p)] + \sum_{j=1}^{\lambda-1}(\theta_j - \theta_{j+1})E_{\chi[\theta_j]}[\theta_j c(p)] = \hat{t}(\theta_\lambda). \quad \text{(B15)}$$

For each $j > \lambda$, by lemma 2, we have

$$\begin{aligned}
E_{\chi[\theta_j]}[\psi_\chi(p) - \theta_j c(p)] &= \frac{\bar{y}_j - p_0}{\bar{y}_j - \underline{y}_j} s_j(\underline{y}_j) + \frac{p_0 - \underline{y}_j}{\bar{y}_j - \underline{y}_j} s_j(\bar{y}_j) \\
&= \frac{\bar{y}_j - p_0}{\bar{y}_j - \underline{y}_j}(s_{j-1}(\underline{y}_{j-1}) + (\theta_{j-1} - \theta_j)c(\underline{y}_{j-1}) + \Delta_j(\underline{y}_j - \underline{y}_{j-1})) \\
&\quad + \frac{p_0 - \underline{y}_j}{\bar{y}_j - \underline{y}_j}(s_{j-1}(\underline{y}_{j-1}) + (\theta_{j-1} - \theta_j)c(\underline{y}_{j-1}) + \Delta_j(\bar{y}_j - \underline{y}_{j-1})) \\
&= \psi_\chi(\underline{y}_{j-1}) - \theta_j c(\underline{y}_{j-1}) \\
&\quad + \Delta_j \frac{(\bar{y}_j - p_0)(\underline{y}_j - \underline{y}_{j-1}) + (p_0 - \underline{y}_j)(\bar{y}_j - \underline{y}_{j-1})}{\bar{y}_j - \underline{y}_j} \\
&= \psi_\chi(\underline{y}_{j-1}) - \theta_j c(\underline{y}_{j-1}) + \Delta_j(p_0 - \underline{y}_{j-1}).
\end{aligned}$$

Then from (B11), we have

$$\begin{aligned}
E_{\chi[\theta_j]}[\psi_\chi(p) - \theta_j c(p)] &= \psi_\chi(\underline{y}_{j-1}) - \theta_{j-1}c(\underline{y}_{j-1}) + \Delta_{j-1}(p_0 - \underline{y}_{j-1}) \\
&\quad + (\theta_{j-1} - \theta_j)\left(c(\underline{y}_{j-1}) + (c(\bar{y}_{j-1}) - c(\underline{y}_{j-1}))\frac{p_0 - \underline{y}_{j-1}}{\bar{y}_{j-1} - \underline{y}_{j-1}}\right) \\
&= \psi_\chi(\underline{y}_{j-1}) - \theta_{j-1}c(\underline{y}_{j-1}) + \frac{(\psi_\chi(\bar{y}_{j-1}) - \theta_{j-1}c(\bar{y}_{j-1})) - (\psi_\chi(\underline{y}_{j-1}) - \theta_{j-1}c(\underline{y}_{j-1}))}{\bar{y}_{j-1} - \underline{y}_{j-1}}(p_0 - \underline{y}_{j-1}) \\
&\quad + (\theta_{j-1} - \theta_j)\left(\frac{p_0 - \underline{y}_{j-1}}{\bar{y}_{j-1} - \underline{y}_{j-1}}c(\bar{y}_{j-1}) + \frac{\bar{y}_{j-1} - p_0}{\bar{y}_{j-1} - \underline{y}_{j-1}}c(\underline{y}_{j-1})\right) \text{ (by [B9])} \\
&= E_{\chi[\theta_{j-1}]}[\psi_\chi(p) - \theta_{j-1}c(p)] + (\theta_{j-1} - \theta_j)E_{\chi[\theta_{j-1}]}[c(p)] \text{ (by lemma 2).}
\end{aligned}$$

Then by (B15), for all $i > \lambda$, we have

$$E_{\chi(\theta_i)}[\psi_\chi(p)] = E_{\chi(\theta_i)}[\theta_i c(p)] + \sum_{j=1}^{i-1}(\theta_j - \theta_{j+1})E_{\chi[\theta_j]}[c(p)] = \hat{t}(\theta_i),$$

as desired.

ψ_χ cannot be outperformed by a methods-based contract that implements χ.—If T implements χ, then by the revelation principle, so does the direct revelation contract (χ, t) with $t(\theta) = T(\chi(\theta))$. By lemma 4, part iib, we have $E_{\chi(\theta)}\psi_\chi(p) = \hat{t}(\theta) \le t(\theta) = T(\chi(\theta))$. QED

Proof of theorem 4 (results-based implementation requires Blackwell monotonicity). Suppose that χ is not Blackwell monotone. Then for some $\theta, \theta' \in \Theta$ with $\theta > \theta'$, $\chi(\theta')$ is not a mean-preserving spread of $\chi(\theta)$. Now for any ψ, $\psi(p) - \theta' c(p) + (\theta' - \theta)c(p) = \psi(p) - \theta c(p)$; since c is strictly convex, $(\theta' - \theta)c(p)$ is a strictly concave function. Then by proposition 1, if $\chi(\theta)$ solves the type θ researcher's problem $\max_\tau\{E_\tau[\psi(p) - \theta c(p)]$ such that $E_\tau p = p_0\}$ for some ψ, $\chi(\theta')$ cannot solve the type θ' researcher's problem $\max_\tau\{E_\tau[\psi(p) - \theta' c(p)]$ such that $E_\tau p = p_0\}$ for that ψ. It follows that no results-based contract ψ can implement χ. QED

References

Angelucci, C. 2017. "Motivating Agents to Acquire Information." Working paper.

Baron, D. P., and R. B. Myerson. 1982. "Regulating a Monopolist with Unknown Costs." *Econometrica* 50:911–30.

Bergemann, D., A. Bonatti, and A. Smolin. 2018. "The Design and Price of Information." *A.E.R.* 108:1–48.

Bizzotto, J., E. Perez-Richet, and A. Vigier. 2020. "Information Design with Agency." CEPR Discussion Paper no. DP13868, Center Econ. Policy Res., Washington, DC.

Blackwell, D. 1953. "Equivalent Comparisons of Experiments." *Ann. Mathematical Statis.* 24:265–72.

Bloedel, A. W., and W. Zhong. 2021. "The Cost of Optimally-Acquired Information." Working paper.

Carroll, G. 2019. "Robust Incentives for Information Acquisition." *J. Econ. Theory* 181:382–420.

Chade, H., and N. Kovrijnykh. 2016. "Delegated Information Acquisition with Moral Hazard." *J. Econ. Theory* 162:55–92.

Che, Y.-K., and N. Kartik. 2009. "Opinions as Incentives." *J.P.E.* 117:815–60.

Dana, J. D., Jr., and K. E. Spier. 1993. "Expertise and Contingent Fees: The Role of Asymmetric Information in Attorney Compensation." *J. Law, Econ., and Org.* 9:349–67.

Ely, J., A. Frankel, and E. Kamenica. 2015. "Suspense and Surprise." *J.P.E.* 123:215–60.

Frankel, A., and E. Kamenica. 2019. "Quantifying Information and Uncertainty." *A.E.R.* 109:3650–80.

Gentzkow, M., and E. Kamenica. 2014. "Costly Persuasion." *A.E.R.* 104 (5): 457–62.

Gershman, B. L. 2006. "Litigating Brady v. Maryland: Games Prosecutors Play." *Case Western Reserve Law Rev.* 57:531–65.

Guesnerie, R., P. Picard, and P. Rey. 1989. "Adverse Selection and Moral Hazard with Risk Neutral Agents." *European Econ. Rev.* 33:807–23.

Häfner, S., and C. R. Taylor. 2019. "Contracting for Research: Moral Hazard and the Incentive to Overstate Significance." Working paper.

Hiriart-Urruty, J.-B., and C. Lemaréchal. 2001. *Fundamentals of Convex Analysis.* Heidelberg: Springer.

Jullien, B. 2000. "Participation Constraints in Adverse Selection Models." *J. Econ. Theory* 93:1–47.

Kamenica, E. 2019. "Bayesian Persuasion and Information Design." *Ann. Rev. Econ.* 11:249–72.

Kamenica, E., and M. Gentzkow. 2011. "Bayesian Persuasion." *A.E.R.* 101:2590–615.

Laffont, J.-J., and J. Tirole. 1986. "Using Cost Observation to Regulate Firms." *J.P.E.* 94:614–41.

———. 1993. *A Theory of Incentives in Procurement and Regulation.* Cambridge, MA: MIT Press.

Machina, M., and J. Pratt. 1997. "Increasing Risk: Some Direct Constructions." *J. Risk and Uncertainty* 14:103–27.

Maskin, E., and J. Riley. 1984. "Monopoly with Incomplete Information." *RAND J. Econ.* 15:171–96.

McClellan, A. 2021. "Experimentation and Approval Mechanisms." Working paper.

Milgrom, P., and C. Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62:157–80.

Min, D. 2021. "Screening for Experiments." Working paper.

Moderna. 2020. "Agreement No. HHSO100201600029C, by and between the Company and the Biomedical Advanced Research and Development Authority, Dated as of April 16, 2020, as Amended on May 24, 2020, June 16, 2020, July 25, 2020, August 31, 2020 and September 15, 2020." Exhibit 10.1 in Q3 2020 Form 10-Q, https://www.sec.gov/Archives/edgar/data/1682852/000168285220000023/exhibit101.htm.

Morris, S., and P. Strack. 2019. "The Wald Problem and the Equivalence of Sequential Sampling and Ex-Ante Information Costs." Working paper.

Mussa, M., and S. Rosen. 1978. "Monopoly and Product Quality." *J. Econ. Theory* 18:301–17.

Myerson, R. B. 1981. "Optimal Auction Design." *Mathematics Operations Res.* 6:58–73.

Novavax. 2020a. "Base Agreement, Entered into June 25, 2020, between the Company and Advanced Technology International." Exhibit 10.1 in Q3 2020 Form 10-Q, https://www.sec.gov/Archives/edgar/data/1000694/000155837020013462/nvax-20200930xex10d1.htm.

———. 2020b. "Undefinitized Project Agreement No. 1, Entered into July 6, 2020, between the Company and Advanced Technology International." Exhibit 10.2 in Q3 2020 Form 10-Q, https://www.sec.gov/Archives/edgar/data/1000694/000155837020013462/nvax-20200930xex10d2.htm.

———. 2021. "Modification No. 04 to Undefinitized Project Agreement No. 01, Entered into December 23, 2020, between the Company and Advanced Technology International." Exhibit 10.43 in FY 2020 Form 10-K, https://www.sec.gov/Archives/edgar/data/1000694/000100069421000004/exhibit1043.htm.

Ohio Attorney General. 2019. "Second Renewal of Retention Agreement for Opioid Manufacturers Fraudulent Marketing Litigation." https://www.ohioattorneygeneral.gov/Files/Legal/Outside-Counsel/Contingency-Fee-Information/Opioid-Manufacturers-Litigation.aspx.

Osband, K. 1989. "Optimal Forecasting Incentives." *J.P.E.* 97:1091–112.

Pomatto, L., P. Strack, and O. Tamuz. 2020. "The Cost of Information." Working paper.

Rappoport, D., and V. Somma. 2017. "Incentivizing Information Design." Working paper.

Rockafellar, R. T. 1997. *Convex Analysis.* Princeton, NJ: Princeton Univ. Press.

Rubinfeld, D. L., and S. Scotchmer. 1993. "Contingent Fees for Attorneys: An Economic Analysis." *RAND J. Econ.* 24:343–56.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Tech. J.* 27:379–423.

Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Sci.* 22:1359–66.

State of Georgia. 2018. "Agreement with Outside Counsel for the Georgia Opioid Crisis Investigation and Litigation." https://law.georgia.gov/document/publication/agreementoutside-counselopioid-investigation-and-litigationpdf/download.

Szalay, D. 2005. "The Economics of Clear Advice and Extreme Options." *Rev. Econ. Studies* 72:1173–98.

Taylor, M., and R. Respaut. 2020. "Exclusive: Moderna Spars with U.S. Scientists over COVID-19 Vaccine Trials." *Reuters*, July 7. https://www.reuters.com/arti cle/us-health-coronavirus-moderna-exclusive/exclusive-moderna-spars-with-u-s-scientists-over-covid-19-vaccine-trials-idUSKBN2481EU.

Thornburg, E. 1991. "Rethinking Work Product." *Virginia Law Rev.* 77:1515–83.

Wald, A. 1945. "Sequential Tests of Statistical Hypotheses." *Ann. Mathematical Statis.* 16:117–86.

Zermeño, L. 2011. "A Principal-Expert Model and the Value of Menus." Working paper.