False Positives and Transparency

Jonathan Libgober

Department of Economics, University of Southern California

July 8, 2019

Abstract. I develop a theoretical model of costly information acquisition in order to evaluate transparency requirements in empirical research. A sender chooses an experiment characterized by multiple dimensions, which determine its informativeness and bias toward positive results. I consider the sender's experiment choice when some dimensions need not be observed. The *receiver* may prefer to keep dimensions hidden, *even* those contributing to bias, despite preferring more informative experiments. This can occur if the perception of bias is lessened when the sender compensates along a dimension that *is* observed. I elucidate how complementarity between dimensions underlies this result.

Keywords. False positives, sender-receiver games, information acquisition, experimentation, transparency.
JEL Codes. D82, D83.

Contact. libgober@usc.edu. This paper comes from the second chapter of my PhD thesis, and supersedes various drafts with similar titles. I thank my committee, Drew Fudenberg, Eric Maskin, Jerry Green and Ben Golub, for their encouragement and advice. I am especially grateful to my discussant Isaiah Andrews for comments which improved the paper (particularly for comments which formed the basis of Section 4.2). Finally, I thank Philippe Aghion, Vivek Bhattacharya, Kirill Borusyak, Juan Carrillo, Yeon-Koo Che, Gonzalo Cisternas, Ashley Craig, Mira Frick, Matthew Gentzkow, Oliver Hart, Johannes Hörner, Emir Kamenica, Navin Kartik, Scott Kominers, Andrea Prat, Neil Thakral, Jesse Shapiro, Kathryn Spier, and Heidi Williams for helpful conversations, and seminar participants at Harvard, the BFI Media and Communication Conference, and the Southwest Economic Theory Conference for excellent feedback. Any remaining errors are my own.

Experimentation is multidimensional. In any empirical study, how much data were used, how precisely the data were recorded, and how many tests were run all influence the informational content of a given result. Some of these dimensions may involve costs, and though some may be naturally observable to outsiders (e.g., sample size), others may not be (e.g., the specifications considered by the researcher). This paper asks whether an interested party always prefers to observe *more* dimensions when observability is not all-or-none.

Costs and limited verifiability are important elements of a variety of settings featuring information acquisition. In this paper, I focus on one particular setting, namely empirical research. I maintain this focus because, for empirical disciplines, the question of optimal transparency over experimental conduct is particularly policy relevant. For instance, a common proposal aimed at establishing greater transparency is to promote pre-registration, whereby scientists describe (at least part of) their planned research activities prior to undertaking experimentation. Ioannidis et al. (2014) advocate this goal, arguing that a lack of transparency contributes to the incidence of false positives. Though some dimensions of an experiment may be verifiable in any case, pre-registration is viewed as one way of making others verifiable as well. Pre-registration has also been considered in economics (Miguel et al. (2014), Coffman and Niederle (2015), Olken (2015)), with similar motivation. But establishing transparency is the goal of various policies; others Ioannidis et al. (2014) mention include providing access to code or data, or strengthening review among funding agencies. Such proposals could be considered by any empirical discipline.

Understanding when transparency helps or hurts consumers of researcher is thus of real practical interest. And a contribution of this paper is in providing a number of concrete recommendations related to transparency policy, using a theoretical model of information acquisition I introduce. One may wonder, for instance, whether interested parties are necessarily worse off whenever *any* kind of biasing activity is unobservable. My analysis suggests the answer is no, and that while pre-specify explanatory variables may be beneficial, pre-specifying robustness checks need not be. This observation holds despite the fact that the temptation to bias under limited transparency is present in both instances. The formal model clarifies the properties of these two research activities which lead to opposite conclusions.

The theoretical model I study is a sender-receiver game that features the aforementioned multidimensionality in experiment choice. The sender conducts a costly experiment that is characterized by a *vector* of actions, producing an observable outcome that is seen by a receiver (he). My assumption that the choice of the sender is an experiment makes my model reminiscent of the Bayesian Persuasion literature, following Kamenica and Gentzkow (2011). But in contrast to this literature, I parameterize the sender's experiment choice, and do not assume that *all* experiments are feasible.[1] The experiment is of interest to the receiver because it provides information on his

[1]Experiment costs will also play a role in my model, and these are absent from Kamenica and Gentzkow (2011). But

returns to effort. The sender, in turn, is interested in influencing the receiver's effort level (or perhaps even his belief itself).

To fit this to a concrete story, take the sender to be a scientific researcher such as a cancer geneticist, deciding how much data to collect and whether to engage in specification searching. Suppose the experiment suggests whether a form of cancer can be attributed to a particular gene. A drug developer (with whom the geneticist has no *formal* relationship) observes the experiment's outcome, and decides how hard to work on curing the disease by targeting the gene. The more optimistic he is of a connection, the harder he works. The scientist may care about developer effort if, for instance, she receives an exogenous benefit (e.g., prestige) when the disease is successfully cured. In this case, the scientist cares about effort even in the absence of explicit incentives from the developer.

I model transparency as a policy, implemented by a third party (e.g., funding agency), that makes certain dimensions of the sender's experiment observable to the receiver. I show that if experiments have costs, then whether a transparency requirement benefits the receiver depends on the complementarity between different research actions. Simply put, the core insight is that *the need to establish experiment credibility under limited transparency can induce a sender to incur costs she would avoid under full transparency.*

This observation depends on twin features of the sender's problem: The incentive to acquire information, and the loss of credibility under limited transparency. The credibility loss arises because the sender does better (ex-post) following a positive result rather than a negative result—for instance, due to higher receiver effort. Unobserved choices will thus be assumed simply to maximize the probability of a positive result, even at the expense of informativeness. Due to her preference for more informative experiments, this lowers her expected payoff. So, if it is possible to increase an action along an observable dimension—e.g., via more sophisticated data collection—limited transparency can induce the sender to do so if this eliminates the perception of bias. If choices were *exclusively* along a dimension that increased bias and decreased informativeness, full transparency would be receiver optimal. With multiple (and varied) dimensions, a compensation channel emerges that may undo this conclusion.

My general results study the substitution that emerges as the observed dimensions change, considering more general preferences and experiment specifications. I highlight that the difference in equilibrium experiment choice across transparency regimes depends on two factors: First, whether transparency over a given dimension induces higher actions holding others fixed, and second, the complementarity between that dimension and others. I refer to the first as the *direct effect* and the second as the *indirect effect*. While the key forces are illustrated simply in Section 2, the general framework shows how the results vary across experimental settings and

Gentzkow and Kamenica (2014) introduce costs using similar techniques. My model of costs differs slightly.

possible incentives. In particular, if the environment features opposite complementarity from that described in the previous paragraph, the opposite conclusion emerges: the loss in credibility *discourages* information acquisition. I note that these characterizations are obtained via the belief-based approach, a common technique in the literature following Kamenica and Gentzkow (2011). However, this approach is complicated if the sender mixes along an unobserved dimension under the limited transparency benchmark. I discuss the necessary concavity assumptions to recover pure strategies and use this observation to articulate the differences in information preferences between sender and receiver.

I discuss the policy implications of my results more concretely in Section 4. While my main results tie the merits of transparency to the complementarity between research actions, it may not be immediately apparent how to take these ideas to practice. Above, I alluded to the argument that the merits of transparency over biasing activity (e.g., specification searching) differs for explanatory controls versus robustness checks. I obtain this conclusion by showing that the indirect effect goes in opposite directions in each instance. I also derive comparative statics regarding the relative importance of various incentives scientists might face. Using these comparative statics, I suggest that optimal transparency policy may depend on a researcher's career stage. More broadly, a lesson from this paper is that theoretical models of information acquisition can be helpful in formally evaluate policies related to scientific conduct.

Though I explain how my results expand on prior work in the conclusion, for now I mention that the dichotomy between the direct and indirect effects uses my multidimensional parameterization of the sender's problem. With only a single dimension (or all-or-none transparency), the direct effect is the only one present, and unobserved bias makes the receiver worse off whenever this lowers experiment informativeness. While non-transparency is optimal in other settings considered in the literature, the fact that my result supports *limited* transparency (versus non-transparency) distinguishes the contribution. My model also accommodates reasonably general preferences, whereas prior work on signal biasing is often more restrictive on this front. This facilitates discussion regarding the relative importance of various incentives for the results, and allows me to highlight some subtle factors influencing the incentive to bias. To the best of my knowledge, these features are missing from other models in the applied mechanism design literature. Still, while these messages are relevant in many applications, I only seek to claim that the counterfactual policies and assumptions I make are appropriate for empirical research. Modifying these for other contexts is left to future work.

I introduce the model in Section 1, and present the key ideas in an example in Section 2. General results are presented in Section 3, and the implications for empirical research are discussed in Section 4. I conclude by discussing assumptions, prior literature, and directions for future research.

1. MODEL

The basic setting is a sender-receiver game. My model resembles Bayesian Persuasion, adding a (multidimensional) parameterization over the sender's experiment set. The two other distinctions are that sender actions may have costs associated with them, and that it need not be the case that every experiment falls within the set of feasible (pure or mixed) sender actions. The notion of partial transparency in this paper rests upon this multidimensional parameterization, and experiment costs are the central (though potentially not exclusive) conflict between the sender and receiver's preferences over information.

Setting and Actions: The sender chooses an experiment that provides information on a state $\theta \in \Theta$, which is of interest to the receiver. I assume $|\Theta| < \infty$, with both parties initially sharing a full-support common prior over Θ, denoted by p_0. After observing the outcome of the experiment (as well as all or part of the experiment choice, as explained below), the receiver chooses an action $e \in [\underline{e}, \overline{e}]$.

The sender's experiment choice is parameterized by a tuple $a = (a_1, \ldots, a_n)$ with corresponding cost $c_S(a)$. Most of the intuition in this paper comes across with $n = 2$. When experiment a is chosen, an outcome y is drawn stochastically, with the distribution over y depending on θ and a. I focus on the case where $Y = \{0, 1\}$, and will refer to $y = 1$ as a positive result and $y = 0$ as a negative result. I write $y \sim a$ to denote that the outcome y is drawn as dictated by the experiment choice a, suppressing the implicit dependence on θ as well. I take each a_i to be chosen from a compact set $A \subset [0, 1]$, and refer to the set of signal distributions and the associated costs as the *experimentation technology*. I will call an action a_i *biasing* if increases make positive results more likely while decreasing experiment informativeness overall (holding other dimensions fixed).

When sender's experiment choice is explicit, I will write $\hat{p}_a(y)$ to denote the belief of the receiver when inferring the sender's choice as a and observing the outcome y. When necessary to emphasize the *particular* state that this belief is over, I write $\hat{p}_a(y)[\theta]$ to denote the receiver's belief that the state is θ.

Payoffs: The receiver's utility depends on both his action choice e and the state θ. I assume utility is given by a function $U^R(e, \theta)$ that is continuous in e with $U^R_{ee}(e, \theta) < 0$, for all θ. Note that this implies a uniquely optimal choice of e, for every possible receiver belief $\hat{p} \in \Delta(\Theta)$.

For some of the discussion below, it will be helpful to allow for the sender's payoff to depend *both* on the receiver's action choice e, as well as the receiver's belief itself $\hat{p}$. Hence I write the sender's payoff from choosing experiment a, inducing receiver belief $\hat{p}$ and effort choice e in state θ to be $U^S(e, \hat{p}, \theta) - c_S(a)$. I assume $U^S(e, \hat{p}, \theta)$ is continuous in $\hat{p}$ and e, for all θ.[2]

[2]All results are maintained sender's payoffs are $\mathbf{1}[\hat{p}[\theta] > p_0[\theta]] \cdot U^S(e, \hat{p}, \theta)$, which can accommodate a discrete preference for positive results, for instance. See Appendix B.2 for a discussion of these preferences.

However, the reader should understand that, since any equilibrium e is a *function* of the receiver's belief alone, so is the sender's ex-post payoff, and so occasionally I will drop the dependence on e. I keep the dependence on e in the model since this will play a role when I discuss the sources of conflict between sender and receiver in Section 2 and Section 4.1.

(Partial) Transparency: My interest is in the problem of a third party, solely interested in the welfare of the receiver, that can require certain coordinates in a be observable to the receiver. I will study how equilibrium experiment choice changes depending on the third party's decision. The third party need not be able to make all dimensions observable, and also may not be able to make some dimensions unobservable. However, I assume that the sender cannot make these coordinates observable unless the third party requires it. Alternatively, one could impose the existence of fixed costs of transparency for the sender, making transparency prohibitive without a requirement. This would not change this paper's results. If some coordinates of a cannot be observed or verified, I refer to the regime as *limited transparency. Full transparency* refers to a regime where all dimensions are observed.

Discussion of the Model: The persuasion literature has made frequent use of the belief-based approach. Under this approach, sender's experiment choice is written as an optimization problem over a belief distribution. However, to adopt this approach, it is more common (e.g., Kamenica and Gentzkow (2011) or Lipnowski and Ravid (2019)) to select the sender-optimal equilibrium. In this paper, as I am focused on receiver-optimal transparency, it seems inconsistent to use such a selection argument. Instead, I focus on convex returns to receiver effort.

A subtler limitation of the belief-based approach is that the receiver must infer the experiment choice of the sender in equilibrium. In my model, the sender can influence this inference by verifiably changing the action along an observed dimension. If the sender mixes over experiments, preferences over information will depend on both the receiver's conjecture and the distribution over outcomes Y, and are not pinned down by the chosen experiment alone. Hence preferences over experiments are endogenous here in a way that is different from similar models. To avoid these complications, Lemma 2 describe the sender's preferences over information in this setting under the condition that a unique inference *is* possible. This characterization will thus apply across transparency regimes.

To fit the model to the story from the introduction, imagine θ determines whether a gene is responsible for a disease, with a geneticist's experiment providing information about this association. While some experiment dimensions may necessarily be verifiable (e.g., sample size), whether others are (e.g., the set of control variables considered) may depend on the policies of a funding agency or professional association. It may be impossible or difficult for the geneticist to *independently* make those verifiable (as it is often difficult without external oversight to prove that some variables were *not* considered). A developer may use the information provided by the

experiment to decide how many resources to expend on targeting the gene, knowing that the benefits depend on θ. While the developer may want to maximize returns, the geneticist may care about recognition associated with a cure, or simply the belief that the gene is important.

2. A SIMPLE EXAMPLE

In this Section, I walk through a numerical specification of the model illustrating the key forces at work which drive the gains to limited transparency. Suppose that $\theta \in \{T, F\}$ (e.g., a hypothesis is true or false) and $\mathbb{P}[\theta = T] = 1/4$. Let the receiver subsequently chooses e (thought of as effort) at cost e^2, which leads to a benefit of 2 with probability e if $\theta = T$ and no benefit if $\theta = F$. The sender obtains a benefit of 1 with probability e if $\theta = T$ and no benefit if $\theta = F$, and does not incur the receiver's effort costs. Ex-post payoffs are thus:

$$U^R(e, \theta) = 2 \cdot e \cdot \mathbf{1}[\theta = T] - e^2 \quad \text{and} \quad U^S(e, \hat{p}, \theta) = e \cdot \mathbf{1}[\theta = T].$$

The sender's action choice is $(a_1, a_2) \in \{0, 1\}^2$, with y being drawn as follows:

$$\begin{aligned}
&\mathbb{P}[y = 1 \mid \theta = T, a_1 = 0, a_2 = 0] = 2/5, && \mathbb{P}[y = 1 \mid \theta = F, a_1 = 0, a_2 = 0] = 0,\\
&\mathbb{P}[y = 1 \mid \theta = T, a_1 = 0, a_2 = 1] = 1/2, && \mathbb{P}[y = 1 \mid \theta = F, a_1 = 0, a_2 = 1] = 1/6,\\
&\mathbb{P}[y = 1 \mid \theta = T, a_1 = 1] = 4/5, && \mathbb{P}[y = 1 \mid \theta = F, a_1 = 1] = 0.
\end{aligned}$$

There are two important features of the experimentation technology. First, choosing $a_1 = 1$ leads to a more informative experiment than either experiment with $a_1 = 0$. Second, increasing a_2 increases the probability of $y = 1$ in both states if $a_1 = 0$. Hence a_2 is biasing when $a_1 = 0$ but has no impact if $a_1 = 1$. Take the sender's cost to be $c_S(a_1, a_2) = c \cdot a_1$.

Using these numbers, it is a straightforward (albeit slightly tedious) exercise to compute the receiver's payoff as a function of the sender's experiment, assuming full transparency. If the receiver's belief that $\theta = 1$ after seeing y following experiment a is is $\hat{p}_a(y)$, his optimal effort is $e(\hat{p}_a(y)) = \hat{p}_a(y)$, yielding payoff $\hat{p}_a(y)^2$. Let $\pi_R(a_1, a_2)$ denote the receiver's payoff as a function of the experiment choice. I compute:

$$\pi_R(1, a_2) \approx 1/5, \quad \pi_R(0, 0) = 1/8, \text{ and } \quad \pi_R(0, 1) = 1/12.$$

Not surprisingly, the receiver does best when the sender chooses $a_1 = 1$, and does worst when $a_1 = 0$ and $a_2 = 1$.

It turns out that sender's payoff can be written as $\pi_R(a_1, a_2) - c \cdot a_1$ when a is observed by the receiver. While perhaps striking, this follows immediately from Lemma 2 below, as I explain

following its statement. Though the preference for informative experiments is more general, the fact that the sender's preferences over information (gross experiment costs) coincides exactly with the receiver's is due to quadratic receiver cost for e. Even though the sender does not incur costs due to the choice of e, the fact that a benefit of e is only obtained if $\theta = T$ implies that payoffs are increasing functions of the variance of the receiver's posterior belief.[3]

I now show that partial transparency may be optimal when $c > 0$. Note that this clearly cannot be the case when $c = 0$, as then there is always an equilibrium where the sender chooses the receiver's most preferred experiment in both regimes. But if c is neither too high nor too low, the receiver strictly prefers to keep a_2 unobserved.[4] In an intermediate range of costs, the equilibrium experiment of the sender involves $a_1 = 1$ if *and only if* a_2 is *un*observable.

Why do gains to limited transparency arise? First suppose that a_2 were observable. In this case, the sender would only be hurting herself by setting $a_2 = 1$, as this does not influence experiment costs and simply lowers the quality of the experiment for the receiver. Hence the sender chooses $a_1 = 1$ if and only if $c < \pi_R(1, a_2) - \pi_R(0, 0)$ (in this case, $\approx \frac{3}{40}$).

Now suppose a_2 were unobservable, in which case the receiver must infer it in equilibrium. If $a_1 = 1$, then this inference is not relevant. But when $a_1 = 0$, receiver beliefs cannot respond to the choice of a_2 because now this is unobserved. And notice that the receiver always chooses higher effort following $y = 1$ compared to when $y = 0$. The sender, in turn, prefers that the receiver choose higher effort, since she does not internalize this cost. So if the receiver's beliefs cannot respond to the choice of a_2, then the sender simply chooses the action that makes a positive result (i.e., $y = 1$) most likely. This means choosing $a_2 = 1$. Intuitively, when a_2 is not observed, the sender *loses credibility for scrupulousness* when choosing a moderately informative experiment. *In equilibrium*, the receiver realizes that the sender does better when the result is positive rather than negative (but takes as given the *observable* dimension), and adjusts inferences accordingly. So under limited transparency, since the sender would choose $a_2 = 1$ whenever choosing $a_1 = 0$, she will choose $a_1 = 1$ if and only if $c < \pi_R(1, a_2) - \pi_R(0, 1) = \frac{1}{5} - \frac{1}{12} = \frac{7}{60}$.

To conclude that limited transparency can be beneficial, simply compare the threshold for c above which the sender will choose $a_1 = 0$. Since $\frac{7}{60} > \frac{3}{40}$, there is an interval of costs such that $a_1 = a_2 = 0$ is chosen when a_2 is observable to the receiver, and $a_1 = 1$ is chosen when it is not. If c is too low, $a_1 = 1$ is chosen in either transparency regime, and if c is too high, $a_1 = 1$ is too costly. In the latter case, partial transparency still induces the sender to add bias. But in the intermediate range, while the sender loses credibility for scrupulousness when bias is not observed, she *compensates by exerting costly effort.* Since the sender has a preference for more

[3] In addition, though receiver incurs the cost of e^2, sender only obtains a benefit of e in state θ whereas receiver's benefit is $2 \cdot e$. These two differences offset one another so that $\pi_R(a)$ is also the sender's payoff (gross costs).

[4] Keeping both dimensions hidden is never receiver-optimal in this example.

information, the only reason she would choose a less informative a_1 would be on account of costs. By making it impossible to commit to $a_2 = 0$, the sender takes a costly action which proves her scrupulousness. Even though having a_2 unobserved induces bias *given* $a_1 = 0$, the experiment overall is more informative in equilibrium.

A striking feature of this example—and an important distinguishing feature of this paper's setting—is that it is *precisely* the alignment over information preferences which provides a channel for partial transparency to benefit the receiver. To illustrate, suppose the sender instead simply wanted the receiver to exert as much effort as possible. Since $e(\hat{p}) = \hat{p}$, this gives us:

$$U^S(e, \hat{p}, \theta) = \hat{p}$$

The specification where the sender cares only about the belief of the receiver (independently of the state) is the focus of the career concerns literature (See Section 5.2; the fact that payoffs can be state-dependent in my model is a distinguishing feature). Since the sender does better when $y = 1$ than when $y = 0$, she would have an incentive to choose $a_2 = 1$ were she to choose $a_1 = 0$, as before. Now, however, benefits (gross of costs) are the same for all experiments because beliefs are a martingale. While the loss of scrupulousness arises, the compensation via costly effort does not. The fact that information preferences are aligned is precisely what induces the sender to incur the additional experiment cost.

3. GENERAL RESULTS

3.1. On Pure Strategies

The results for the general model apply to a pure strategy equilibrium and therefore require their existence. Though this can be checked in particular examples, assumptions are necessary for existence more generally. The focus on pure strategies will be important because I characterize the direct and indirect effects using the first-order condition of the sender's problem. When mixing occurs along an unobserved dimension in equilibrium, two additional issues make this approach less straightforward: First, there are more first-order conditions to worry about. Second, it becomes necessary to distinguish the *inferred* and the *realized* experiment choice. On the second point, note that when mixing is present, observing the outcome of the sender's randomization is a signal that helps him interpret y. While this could influence the conclusion regarding the merits of transparency, it is not clear that randomization over methods is an important factor for the policies I wish to comment on.

In any case, a pure strategy equilibrium can be ensured with some reasonably innocuous

assumptions.[5] These assumptions are as follows:

Assumption 1. *For all $a \in A$, there is some θ such that $0 < \mathbb{P}[y = 1 \mid a, \theta] < 1$.*

Assumption 2. *In the dimensions that may be unobservable, the cost function $c_S(a)$ has continuous derivatives and is convex.*

Assumption 3. *For every a and all θ, $U^S(\hat{p}_a(1), \theta) \geq p_0 \geq U^S(\hat{p}_a(0), \theta)$. Furthermore, in any dimensions that may be unobservable $\mathbb{P}[y = 1 \mid a, \theta]$ is concave with continuous derivatives.*

The first assumption ensures that the result of an experiment does not, by itself, reveal the experiment that was undertaken by the sender. This eliminates multiplicity issues that are common in communication games where certain signals are observed with zero probability. Convexity of costs captures the idea that a convex combination of actions (e.g., running a fraction α of robustness checks from one list and a fraction $1 - \alpha$ from a second list) could only decrease costs relative to doing each separately. Perhaps the most economically substantive restriction is in Assumption 3, in that the label of "positive" and "negative" result make sense in terms of sender payoffs independently of the state or experiment. This assumption holds insofar as scientists uniformly prefer results which make followers more optimistic about particular hypotheses. I discuss this further in Section 5.1.

These assumptions deliver pure strategy existence, by invoking Kakutani's fixed point theorem:

Lemma 1. *Under Assumptions 1-3, there exists a pure strategy perfect Bayesian equilibrium, and no mixed strategy equilibria when convexity is strict in Assumption 2.*

The next Lemma rewrites the preferences of the sender in terms of the receiver beliefs, a consequence of being able to apply the belief-based approach with pure strategies:

Lemma 2. *In any pure strategy equilibrium, Sender's payoffs can be written:*

$$\mathbb{E}_{y \sim a}\left[\mathbb{E}_{\theta \sim \hat{p}_a(y)}\left[U^S(\hat{p}_a(y), \theta)\right]\right] - c_S(a).$$

The substantive part of the Lemma is that the expectation over states is taken with respect to the receiver's posterior beliefs. While this is essentially a version of the Law of Iterated Expectation, it differs because it distinguishes between the experiment *chosen* by the sender and the experiment *inferred* by the receiver. This is why it matters that the inference is correct. When the sender is mixing over *unobserved experiment dimensions*, sender's preferences over information depend on

[5] I am not aware of other papers for which unobserved mixing over *experiments* is important, even in cases where mixed strategies play a role. In Lipnowski and Ravid (2019), for instance, the sender sometimes *must* use a mixed strategy in order to achieve her optimal payoff. However, mixing is over the *messages chosen*, and the optimum is a pure strategy when viewed as mapping from states to distributions over outcomes.

both the realized experiment and the receiver's conjecture.[6] The pure-strategy requirement thus facilitates discussion of information preferences across transparency regimes.

Note that this Lemma rewrites the sender's objective in a way that facilitates relating sender and receiver's preferences over information.[7] For instance, while it may not be obvious that the sender and receiver share preferences over information (in the absence of costs) in Section 2, the Lemma makes this clear that the sender's preferences payoffs (gross costs) are:

$$\mathbb{E}_{\theta\sim\hat{p}_a}\left[U^S(\hat{p}_a(y),\theta)\right] = \hat{p}_a(y)\cdot e_a(\hat{p}(y)) + (1-\hat{p}_a(y))\cdot 0 = \hat{p}_a(y)^2.$$

3.2. Single-Dimensional Experiment Choice

Returning to the question of how experiment choice differs across transparency regimes, I first consider the incentive to change an action on some dimension when transparency over that dimension changes. I do this by focusing on the sender's choice when her action is single dimensional, thus eliminating the possibility of any adjustments along other dimensions. This yields a characterization of the *direct effect*. The Proposition follows from comparing the sender's first-order condition when beliefs adjust to first-order condition when they do not:

Proposition 1 (The Direct Effect)**.** *Suppose Assumptions 1-3 hold and that the experiment choice is one dimensional. Then making the experiment choice observable results in a weakly higher (resp. lower) choice of experiment whenever:*

$$E_{y\sim a}\left[\mathbb{E}_{\theta\sim\hat{p}_a(y)}\left[\nabla_{\hat{p}}U^S(\hat{p}_a(y),\theta)\cdot\left(\frac{d\hat{p}_a(y)[\theta]}{da}\right)_{\theta\in\Theta}\right]\right], \tag{1}$$

is positive (resp. negative). The change in experiment choice is strict whenever it is interior.

The Proposition shows whether higher actions are encouraged (or discouraged) hinges on whether the changes in preference are in the same direction as the changes in beliefs; formally, if $\nabla_{\hat{p}}U^S(\hat{p}_a(y),\theta)$ is positively correlated with $\left(\frac{d\hat{p}_a(y)[\theta]}{da}\right)_{\theta\in\Theta}$ (in expectation). If these changes in the payoffs move in the same direction on average as the changes in posterior beliefs, then the action is encouraged under observability.[8]

[6] The Appendix proves a version of Lemma 2 without correct inference requirement, illustrating its role.

[7] Though similar steps arise in Kamenica and Gentzkow (2011) and Dewatripont, Jewitt and Tirole (1999), I have not seen the problem being rewritten in this way elsewhere. While Dewatripont, Jewitt and Tirole (1999) provide a similar expression in the statement of their Proposition 2.1, this is an equilibrium condition and not a statement about preferences. While Kamenica and Gentzkow (2011) allow for state-dependent sender preferences, this Lemma is not in their paper. Their result takes the sender's payoff as a function of a posterior as given, showing a preference for information arises when this function is below its concavification at p_0. In contrast, Lemma 2 is not about optimal experiments and hence does not require the concavification to be computed.

[8] A similar argument is used in Dewatripont, Jewitt and Tirole (1999), who describe the marginal incentives of

To be a bit more concrete regarding the content of this observation, suppose $\theta \in \{0, 1\}$ with $U^S(\hat{p}, \theta) = \hat{p}[\theta = 1]$, i.e., the posterior belief that $\theta = 1$ (discussed at the end of Section 2). Then $\nabla_{\hat{p}} U^S(\hat{p}, \theta) = (1, -1)$, for all θ, and so the expectation over θ can be dropped. Hence whether the action is encouraged depends on the sign of:

$$\mathbb{E}_{y \sim a}\left[\mathbb{E}_{\theta \sim \hat{p}}\left[(1,-1) \cdot \begin{pmatrix} \frac{d\hat{p}_a(y)[T]}{da} \\ -\frac{d\hat{p}_a(y)[T]}{da} \end{pmatrix}\right]\right] = 2 \cdot \mathbb{E}_{y \sim a}\left[\frac{d\hat{p}_a(y)[T]}{da}\right]. \tag{2}$$

Suppose that higher a translates to more informative experiments. There are two ways experiments can become more informative: if positive results are more revealing that $\theta = 1$, or if negative results more revealing that $\theta = 0$. When the first case is more relevant, $\frac{d\hat{p}_a(y)}{da}$ is relatively larger (and positive)—so transparency increases a. If higher actions make experiments more informative by making positive results more revealing that the state is good, then transparency encourages informativeness. Mathematically, the changes in beliefs move in the same direction as $\nabla_{\hat{p}} U^S(\hat{p}, \theta)$, since they change more in the $\hat{p}[\theta = 1]$ coordinate more than in the $\hat{p}[\theta = 0]$ coordinate.

However, if the experiment becomes more informative because negative results are more revealing that $\theta = 0$, then transparency *discourages* informativeness. In that case, changes in the $\hat{p}[\theta = 0]$ coordinate are more significant, which moves in a direction opposite $\nabla_{\hat{p}} U^S(\hat{p}, \theta)$. So contrary to the case where actions uniformly make positive results more likely, informative actions may be either encouraged or discouraged under non-transparency.

If $U(\hat{p}, \theta) = \hat{p}[\theta = 1] \cdot \mathbf{1}[\theta = T]$ (as in the focus of Section 2), then $\nabla_{\hat{p}} U(\hat{p}, F) = 0$, making (2):

$$\mathbb{E}_{y}\left[\hat{p}_a(y)[T] \cdot \left[(1,-1) \cdot \begin{pmatrix} \frac{d\hat{p}_a(y)[T]}{da} \\ -\frac{d\hat{p}_a(y)[T]}{da} \end{pmatrix}\right]\right] = 2 \cdot \mathbb{E}_{y \sim a}\left[\hat{p}_a(y)[T] \cdot \frac{d\hat{p}_a(y)[T]}{da}\right]. \tag{3}$$

The same general intuition outlined above applies to this case as well, with the caveat that the direct effect now depends less on how beliefs change following signals which reveal that $\theta = F$ with higher probability (i.e., negative results). Thus a subtle difference emerges in the incentive to alter the experiment across these two specification of researcher's preferences. With state dependence, this incentive to alter an action is less influenced by the change in receiver beliefs following negative results, relative to the change in receiver beliefs following positive results (since $\hat{p}_a(1)[T] > \hat{p}_a(0)[T]$).

distorting a signal in terms of the covariance between the state and the likelihood ratio, the latter of which depends on actions. My expression features richer preferences, but requires finite states, and hence involves $\nabla_{\hat{p}} U^S$ (unlike theirs). Since their model focus on particular preferences, applying their results directly is difficult. See Section 5.2 for further comparison to this paper.

3.3. Multidimensional Actions and Complementarity

I now reintroduce multidimensionality to discuss the *indirect effect*. Proposition 2 shows how the incentive to adjust an action when it is unobserved differs depending on the choices along other dimensions. This result is quite simple. In the next section, I show that when this wedge is large, partial transparency can be receiver optimal.

Consider the marginal benefit to higher a_i were receiver beliefs to not adjust:

$$M_{a_{-i}}(a_i) := \sum_{\theta} (U^S(\hat{p}_{(a_{-i},a_i)}(1), \theta) - U^S(\hat{p}_{(a_{-i},a_i)}(0), \theta)) \frac{\partial \mathbb{P}[y = 1 \mid \theta, a_{-i}, a_i]}{\partial a_i} \mathbb{P}[\theta]. \tag{4}$$

Together with marginal costs to higher a_i, this expression pins down the choice of a_i when unobserved, given the choice of a_{-i}. In turn, actions a_{-i} for which the marginal incentives are steeper are less susceptible to being altered:

Proposition 2 (The Indirect Effect)**.** *Suppose Assumptions 1-3 hold, and suppose, for all a_i:*

$$M_{\tilde{a}_{-i}}(a_i) - \frac{\partial c_S(\tilde{a}_{-i}, a_i)}{\partial a_i} \leq M_{a^*_{-i}}(a_i) - \frac{\partial c_S(a^*_{-i}, a_i)}{\partial a_i}. \tag{5}$$

Then the choice of a_i is (weakly) higher given an observable (or correctly inferred) choice of a^*_{-i} than it is given an observable (or correctly inferred) choice of $\tilde{a}_{-i}$.

The intuition is straightforward; actions which are less susceptible to this credibility loss (and hence favored under limited transparency) are those for which there is lower marginal benefit to altering actions along unobserved dimensions. Notice that this captures the key difference between the choice of $a_1 = 1$ and the choice of $a_1 = 0$ in Section 2; modifying the experimentation technology to have a continuous choice of $a_2 \in [0, 1]$ (treating this as the probability that $a_2 = 1$), we see that marginal cost to higher a_2 is 0, so (5) becomes $M_1(a_2) = 0 < M_0(a_2)$.

The simplicity of this result makes it amenable to discussion of practical examples of research activity, as in Section 4.2. However, showing that observed actions themselves change requires conditions on sender's preferences, as imposed in the welfare statement in Theorem 1. But these conditions are not necessary to appreciate the distinguishing feature of the indirect effect (relative to others in the literature), namely that it relates to incentives along dimensions that remains observable.

3.4. Welfare Implications

My last general result relates to receiver welfare with some additional structure on preferences, emphasizing that the general results so far focused exclusively on sender's incentives. Recall that

information preferences (absent costs) for both sender and receiver are:

$$\pi_i(a) = \mathbb{E}_{y\sim a}\left[\mathbb{E}_{\theta\sim\hat{p}_a(y)}\left[U^i(\hat{p}_a(y),\theta)\right]\right], \;\; i \in \{S, R\}.$$

Since receiver's payoffs come from a decision problem, the induced (complete) ordering over experiments is a refinement of the Blackwell order. Though this need not be true for the sender, the focus of this paper is the case where the primary conflict between producers and consumers is time and effort in experimentation.[9] As in Section 2 and highlighted again in Section 4.1, this emerges naturally if incentives for follow-on work are significant.

The Theorem below states that compensation emerges due to the combination of the forces highlighted so far. The restrictive conditions are the aforementioned alignment over information preferences, as well as there being scope for the highlighted effects to matter (as otherwise welfare comparisons follow from Proposition 1). This result articulates the forces yielding gains to limited transparency in Section 2, retaining the two-dimensional setup for simplicity:

Theorem 1. *Suppose Assumptions 1-3 hold and take $A = A_1 \times A_2$ with $A_1 = \{\alpha_1, \cdots, \alpha_n\}$. Suppose further that $\pi_R(a)$ and $\pi_S(a)$ are both increasing in a_1 and decreasing in a_2, and that if (a_1^{obs}, a_2^{obs}) is the sender's choice under full transparency, then $a_1^{obs} < \alpha_n$. Finally, suppose for all a_1:*

$$\pi_S(a_1, a_2^{obs}) - c_S(a_1, a_2^{obs}) > \pi_S(a_1^{obs}, \max A_2) - c_S(a_1^{obs}, \max A_2),$$

i.e., high a_2 is worse for the sender than changing a_1. Then if:

- *(1) is sufficiently large whenever $a_1 = a_1^{obs}$ under observability,*
- *$M(\alpha_{i+1}, a_2) - M(\alpha_i, a_2) - \left(\frac{\partial c_S(\alpha_{i+1}, a_2)}{\partial a_2} - \frac{\partial c_S(\alpha_i, a_2)}{\partial a_2}\right)$ is sufficiently negative relative to (1), for all i and a_2,*

then the receiver strictly prefers to keep a_2 unobserved. Conversely, if instead $a_1^{obs} > \alpha_1$ and:

- *$M(\alpha_{i+1}, a_2) - M(\alpha_i, a_2) - \left(\frac{\partial c_S(\alpha_{i+1}, a_2)}{\partial a_2} - \frac{\partial c_S(\alpha_i, a_2)}{\partial a_2}\right)$ is sufficiently positive relative to (1), for all i and a_2,*

then receiver strictly prefers to keep a_2 observed.

Within the general environment, the theorem formalizes that complementarity is the driving force behind the gains to limited transparency. The loss of credibility that emerges under limited transparency is due to the direct effect (as per the first bullet point). The receiver-benefiting

[9] This paper is not alone in positing these costs as the primary conflict in this application; the same is true in Glaeser (2006), for instance. And other models which do not impose this (e.g., Di Tillio, Ottaviani and Sørenson (2017, 2018)) may still satisfy it in particular specifications.

compensation is induced due to the indirect effect (as per the second bullet point). The Theorem also highlights that the opposite conclusion is obtained with the opposite complementarity. Were Section 2 to instead posit that the *less* informative experiment was more susceptible to bias, we would instead obtain a range of costs where transparency is *necessary* to induce the more informative choice of a_1. In this case, the sender requires credibility to be willing to invest in informativeness. When complementarity goes in the other direction, the incentive to re-establish credibility makes the experiment less informative.

I briefly comment on the Theorem conditions. The roles of most of these conditions are straightforward to see; for instance, the experiment choice should be bounded away from edge cases under observability to ensure scope for the direct and indirect effects to exist. The other conditions on sender preferences ensure that biasing is in fact costly, and that higher experiments are not too costly.[10] The discreteness in Theorem 1 greatly simplifies the necessary hypotheses. Though a similar result holds when the first dimension is continuous, a large degree of complementarity implies that only a *small* change in the observable dimension is necessary to compensate. The issue to address is if the necessary change in a_1 is small relative to the cost of higher a_2. On the other hand, Theorem 1 is maintained as long as the marginal cost does not change too steeply *near the experiment choice under observability*. Or, along the lines of this paper's take-away, as long as the compensation needed to overcome the direct effect is non-negligible.

4. POLICY IMPLICATIONS

This section is meant to guide how to bring the model to policy. Propositions 1 and 2 summarize how observability of some dimension influences the equilibrium experiment choice. Here, I seek to calibrate each of these to various components of the scientific research application. Before doing this, I emphasize that while I view the model setup and Assumptions 1-3 as reasonable for empirical research generally, the *conditions* behind the direct and indirect effect need not hold uniformly. In fact, they *should* vary widely across disciplines and experiments and hence may have varying relevance. The point of this section is to delineate when each condition should be significant or not.

4.1. Follow-on Work versus Perceptions

A comparative static that comes out of my analysis relates to how the impact of transparency requirements may differ depending on the stage of a researcher's career. To do this, I elaborate on the preferences in Section 2. Suppose $\theta \in \{T, F\}$, and the receiver's utility function is:

[10] Note that if any experiment is too costly to be chosen, it can essentially be removed from the consideration set for the researcher.

$$U^R(e,\theta) = b_R \cdot e \cdot \mathbf{1}[\theta = T] - c_R(e),$$

for some $b_R > 0$ and a convex cost function c_R (where as before, e can refer to the probability of successfully developing follow on work). Meanwhile, the sender cares about a combination of receiver's success and the belief that the hypothesis is true:

$$U^S(e,\theta,\hat{p}) = \lambda \overbrace{b_S \cdot e \cdot \mathbf{1}[\theta = T]}^{(1)} + (1-\lambda) \overbrace{g(\hat{p})}^{(2)}.$$

I refer to (1) as the *follow-on incentive* and (2) as the *perception incentive*. This latter refers to any benefit researchers may obtain in cases where they have a reputation for finding true hypotheses. Both of these are nested within the general model, and while the second term is reminiscent of the career-concerns literature, I am not aware of models in this literature that facilitate both kinds of incentives. Given that the relative importance of each may differ over the course of a career, I think of λ as reflecting differences in career stage and comment on the corresponding implications for transparency requirements.

First, I note that the follow-on incentive does generate incentives for information acquisition by the receiver, under conditions on $c_R(e)$:

Lemma 3. *A sufficient condition to ensure that the sender's follow-on benefit is convex in the receiver's belief is $c_R'''(e) \leq 0$, or if $c_R(e) = e^n/k$ for any $n > 1$.*

On the other hand, taking $g'' < 0$ is natural insofar as researchers tend to be risk averse over long term career outcomes. This captures, in a reduced form way, the tension between long-term and short-term incentives of researchers. In the short-term, researchers may very well have an incentive to add informativeness to experiments, as per Lemma 3. But in the long-term, it is harder to be unbiased if a negative result decreases the influence of the researcher's future experiments.

I make two points distinguishing follow-on incentives from information acquisition incentives. When $\lambda = 1$:

$$M_{a_{-i}}(a_i) = b_s(e(\hat{p}(1)) - e(\hat{p}(0))) \frac{\partial \mathbb{P}[y = 1 \mid a_{-i}, a_i, \theta = T]}{\partial a_k} \mathbb{P}[\theta],$$

but when $\lambda = 0$:

$$M_{a_{-i}}(a_i) = (g(\hat{p}(1)) - g(\hat{p}(0))) \frac{\partial \mathbb{P}[y = 1 \mid a_i, a_{-i}]}{\partial a_i}.$$

That is, the false positive rate influences the marginal benefit to higher a_2 for perception (or career-concerns) incentives, but not the follow-on incentives. When follow-on incentives matter more, biasing is done as a means to increase the true positive rate, even though more false positives may mean less experiment informativeness. The a_{-i} choices that are relatively favored under

limited transparency when λ is high are those less susceptible to unobserved changes in the *true* positive rate. When perceptions matter more, the distinction between false positives and true positives matters less.

Second, Lemma 3 shows that the incentive for follow-on research leads to gains to information acquisition, whereas the perception incentive discourages information acquisition (provided $g''(\hat{p}) < 0$). A simple corollary is that the compensation effect highlighted is more relevant for more established researchers, as these researchers would be more willing to incur additional costs to make experiments more credible. If perceptions matter more, then the fact that experiments are per se perceived less informative is not as significant of a problem, due to the risk associated with more informative experiments. This discussion suggests that transparency requirements may optimally vary depending on career stage, as the compensation channel is more relevant for late-career researchers.

4.2. Which Kinds of Experiments Should Be Registered?

I now use my results to distinguish the implications of the results for two different kinds of research activity: specification searching versus robustness checks. I show that the model provides opposite recommendations regarding the merits of transparency over each. In both of these examples, assume that $a_1 \in \{\ell, h\}$ parameterizes an experiment with some underlying informativeness, with ℓ being less informative than experiment h. Both of these examples consider the setting where a_2 corresponds to an amount of p-hacking, with higher levels of a_2 denoting effort to "explain away" bad news. The difference between the two is what aspect of the experiment this biasing concerns:

Example 1 (Verification)**.** *Let a_1 be a parameter indexing the underlying informativeness of experiments, and a_2 denote effort in finding robustness checks which support an underlying result. In this case, $\mathbb{P}[y = 1 \mid \theta, a_1, a_2]$ is increasing in a_2. However, experiments with a higher degree of underlying informativeness may be less susceptible to the inclusion of specific robustness checks, since large models contain more variables that one might suspect were strategically omitted. In this case, $\frac{\partial \mathbb{P}[y=1|\theta,a_1,a_2]}{\partial a_2}$ is smaller when the underlying experiment when $a_1 = h$.*

Example 2 (Explanatory Controls)**.** *Let a_1 be a parameter indexing the underlying informativeness of experiments, and a_2 denote effort in searching over experiment controls. In this case, $\mathbb{P}[y = 1 \mid \theta, a_1, a_2]$ is increasing in a_2. However, experiments with more possible controls have more possible interactions and hence more scope for searching activity to change the result. In this case, $\frac{\partial \mathbb{P}[y=1|\theta,a_1,a_2]}{\partial a_2}$ is larger when the underlying experiment when the choice of a_1 corresponds to more informative experiments.*

I use Proposition 2 to articulate why a transparency requirement will have differing impacts for each of these kinds of research activities. To tell the clearest story, let us focus on the case where

the *cost* of each activity does not differ depending on the underlying informativeness (i.e., a_1), but the impact on whether the result is positive or negative does.[11] This means it suffices to compare $M_{a_1}(a_2)$ across a_1.

Let us start with Example 1. Here, experiments with *higher underlying informativeness* are less susceptible to discretionary choices regarding robustness checks. While large designs may be costlier overall, they reduce the scope for cherry-picking results. For instance, in illustrating for the importance of including "non-core variables" in regression specifications, Lu and White (2014) argue that the use of including more variables in models can lead to estimates that are less sensitive to ad-hoc discretionary choices [12] Since the marginal impact on the results depends less on a_2 when h is chosen, it follows that $M_\ell(a_2) > M_h(a_2)$. Proposition 2 says that the resulting equilibrium level of bias is higher for ℓ rather than h. Hence limiting transparency over this activity increases the incentive to choose the more informative experiment. This is an argument *against* pre-registration for these kinds of activities.

Now let's turn to Example 2. Here, the assertion that more informative experiments are easier to bias implies $M_\ell(a_2) < M_h(a_2)$—an experiment that includes more controls is more susceptible to bias, since there are more possible associations that can be found when there are more underlying controls, and any one interaction may lead to finding a positive result (see Vivalt (2018) for a reiteration of the point that experiments with more controls are more susceptible to biasing). While the researcher may be making the experiment more informative by collecting more controls, it is also easier to bias. This contrasts with robustness tests, since the researcher may not seek to claim a positive result because a particular robustness check is successful. Hence the opposite conclusion emerges, suggesting that pre-registration can be beneficial.

Taken together, this difference suggests the importance of distinguishing between explanatory controls and non-core regressors (in the language of Lu and White (2014)) in pre-registration. Of course, this observation requires the caveat that it is important that it is feasible to obtain a rich assembly of controls, something which may be more true in some fields than others. While to the best of my knowledge this reasoning is new, it is worth noting that many registration activities that been promoted in economics have focused on registering explanatory variables (Olken (2015)). The main punchline of this paper thus seems consistent with some existing policies, though the model's value is in formally clarifying their appropriate boundaries.

[11] In this case, costs are necessary only to ensure that the amount of p-hacking is not unlimited—otherwise, equilibrium experiments would always be maximally p-hacked.

[12] They write: "By submitting only results that may have been arrived at by specification searches designed to produce plausible results passing robustness checks, researchers can avoid having reviewers point out that this or that regression coefficient does not make sense or that the results might not be robust." Their argument that including more controls allows for easier and more precise hypothesis testing suggests that designs with more controls endow researchers with greater commitment, consistent with the complementarity I impose in this example.

4.3. Implications of Contractibility

The argument that limited transparency may be optimal requires the assumption that the receiver effort profile is chosen without commitment. This tends to be the case for the application; if the sender is a university researcher, she may not have a direct contracting relationship with the drug developing receiver who will use the results at some point. It may be too costly for an individual developer to invest in learning about the impact of a certain molecule on a certain biological pathway, even though such research may be helpful for a variety of different kinds of medicines.[13]

If direct contracting between the sender and the receiver were feasible, then the receiver may be able to commit to a particular effort profile. In this case, the receiver would be best off if he could observe the full experiment choice. For simplicity, I explain this in the context of Section 2, although the conclusion is more general. In this case, choosing $e = 0$ yields a payoff of 0 to the sender, and corresponds to the worst possible outcome. As long as the sender has positive payoff from choosing a given experiment a, then she would be willing to choose it to prevent $e = 0$. So, with contractible experiments, the receiver's surplus depends on the set of individually rational experiment under a given effort profile $e(\hat{p})$. However, the transparency regime *does not* influence the set of individually rational experiments. Hence observing the full experiment is optimal, since it increases the set of possible punishments.[14]

5. CONCLUSION

This paper seeks to make both a theoretical and an applied contribution. On the latter, it seeks to provide a formal framework to evaluate opposing viewpoints in the debate over transparency in scientific research. Arguments that biasing is solved endogenously are compelling in cases where a substitution channel exists, provided consumers of research are sophisticated and update as a rational receiver would. But the transparency requirements recommended by Ioannidis et al. (2014) are attractive when these channels are not present, or if there is insufficient incentive to invest in informativeness.

My model introduces a framework to discuss partial transparency over actions allowing for general preferences of senders (albeit while restricting the outcome space for experiments). I view this as a good approximation to the problem of experiment choice in empirical research. Therefore, I believe my results speak more directly to the relevant policy questions than much of the prior applied mechanism design literature, surveyed in Section 5.2. Still, the model I have developed is fairly general, and it does not seem to be a stretch to use this model to discuss information

[13] This argument often motivates the use of public funds for research activities in the first place.

[14] Of course, such an argument relies upon the fact that the sender's preferences respond to the receiver's effort choice (which is why it is easiest to illustrate for Section 4). If this were not the case, then this argument would not apply.

acquisition more generally. But in order to maintain focus, I have (until now) said much less on other applications, and leave this to future work.

5.1. Discussion of Assumptions 1-3 and Conditions for Result

I make two comments on Assumptions 1-3, which are used in order to ensure pure strategy equilibria across transparency regimes. First, the assumptions impose that dimensions that are potentially observable are continuous. Even in cases where they are not (such as Section 2), one can consider a version of the model where the choice of the sender along this dimension consists of a *probability* of choosing a given action. But the main results would implicitly assume that transparency would result in this probability being observable, and not the realized action.

Second, Assumption 3 imposes that results can be classified into *positives* (good realizations for the sender) and *negatives* (bad realizations for the sender), which are independent of the state and experiment. This property imposes enough uniformity on the sender's problem to avoid the non-existence of a pure equilibrium, with Appendix B.1 pointing out some assumption of this form is necessary. Though not conceptually difficult, classifying results as "favorable" or "unfavorable" is more cumbersome when $|Y| > 2$, without further restricting U^S. Still, such a separation seems natural for the application, as there does seem to be a consensus that researchers prefer statistically significant results (see Andrews and Kasy (2019) and Brodeur et al (2016)). And even when the state corresponds to a magnitude (and hence may be non-binary), hypothesis tests are often framed using a "significant or insignificant" dichotomy, with the dominant preference being for the former. So while Assumption 3 may be inappropriate in some settings where communication games have been applied, it does not seem out of line for my main application. In any case, even without this restriction, one could still distinguish the direct and indirect effects with more general outcomes, provided pure strategy existence could be maintained through other means.

Theorem 1 shows that limited transparency can be receiver-optimal when receiver and sender share similar preferences over information. While this alignment may seem artificially strong in general, it applies to many well-studied specifications in the communication literature. Under quadratic preferences a la Crawford-Sobel,[15] for instance, both sender and receiver do better (absent experiment costs) when the variance of the receiver's posterior is lower (as in Section 2). However, if the sender's bias is small (and, for instance, the state θ is binary), then under the experimentation technology of Section 2, there may be no incentive to choose $a_2 = 1$ following $a_1 = 0$ with these preferences, even if costless.

Of course, what matters most is whether alignment is a reasonable assumption for the application. Certainly time and effort in experimentation is a basic conflict. And it is not necessarily

[15] That is, taking $U^S(e, \theta) = -(e - (\theta + b))^2, U^R(e, \theta) = -(e - \theta)^2$.

inconsistent with others; for instance, an intrinsic preference for positive results by researchers.[16] But if limited transparency means researchers lose the ability to choose their preferred experiments, why do they not "self-impose" transparency? Answering this question convincingly is beyond the scope of this paper, but many actually do when straightforward. Venues such as the AEA hypothesis registry and aspredicted.org are commonly used even without formal requirements. However, this may not be enough with many possible contingencies to describe or if steps are difficult to verify. Planning costs appear to be the simplest explanation regarding why they are not used more thoroughly, at least for a first pass. Note that these planning costs *would* need to be taken into account in order to describe *sender* welfare across regimes.

I lastly comment that a richer model may allow for transparency to be stochastic. While I allow for dimensions to differ on their observability, I have still assumed that a dimension is either observable or not. This may be reductive for some policies. Making code and data available, for instance, may not *necessarily* lead to all research actions being observable, but only in the event that inspection occurs, which may itself be random. The forces in this paper would still be present, weighted by the probability that the dimension is observable. That said, stochastic verification suggests a number of other modifications which may be interesting to explore in different settings.

5.2. Relation to Prior Work

The limited transparency benchmark corresponds to a version of costly persuasion with partial verifiability. The sender commits to the dimensions of the experiment that are observed, but can manipulate the dimensions that are unobserved. Without costs or restrictions on experiment choice, a fully observable experiment would make my model a special case of Bayesian Persuasion (Kamenica and Gentzkow (2011)); a fully *un*observable experiment would make my model a special case of cheap talk (Crawford and Sobel (1982), Lipnowski and Ravid (2019)).[17] Costs have been introduced to both Bayesian Persuasion and cheap talk elsewhere,[18] as have various forms of intermediate commitment.[19] Note that *both* modifications are crucial; without costs and given alignment over information preferences (as in Section 2), the sender would always choose the most informative experiment under full transparency, but may add bias under limited transparency.

[16] I discuss these preferences in Appendix B.2; briefly, such preferences do not yield experiment choices that respond to transparency, at least not without preferences such as those present in the main model.

[17] The connection to cheap talk may not be obvious, since these models typically have the sender choose a *message* and not an *experiment*. However, Lipnowski and Ravid (2019) show that one can equivalently formulate the sender's problem in cheap talk as an experiment choice subject to a consistency condition.

[18] See Gentzkow and Kamenica (2014) for Bayesian Persuasion; Argenziano, Severinov, and Squintani (2014) and Pei (2014) for cheap talk.

[19] See Lipnowski, Ravid and Shishkin (2018), Nguyen and Tan (2018), Guo and Shmaya (2017) or Min (2017). Also relevant are papers where signals can be distorted in particular ways, such as fraud as in Lacetera and Zirulia (2008), or selective disclosure as in Henry (2009) and Felgenhauer and Schulte (2014).

A number of theoretical papers (with different applications) have illustrated that limiting transparency can be optimal in principal-agent settings with limited commitment. Transparency in these papers is typically "all-or-none," or of a different kind than here. Intuitively, under non-transparency, it becomes easier to commit to ex-post suboptimal actions that provide beneficial incentives ex-ante. Results of this form can be found in Prat (2005), Angelucci (2017), Cremer (1995) and Bergemann and Hege (2004). Indeed, in Section 2, the receiver's optimal policy with commitment would be full transparency with $e_a(\hat{p}) = 0$ for all $a \neq (1, 0)$.[20] But with single-dimensional effort (as these papers all feature), it is hard to see how their channels qualify as substitution. For me, the ability to *verifiably* compensate for a loss of credibility drives the result.

The career concerns literature (following the seminal work of Holmström (1999)) shows how limited observability leads to signal distortion (under particular assumptions on preferences), analogous to adding bias in this paper. As emphasized in Section 2, the fact that the sender cares about the receiver's belief differently in different state is a key feature of my model, and typically absent from this literature. More importantly, as far as I know, my limited transparency benchmark—and in particular, the corresponding compensation channel—has not been considered previously in this literature. Closest is Dewatripont, Jewitt and Tirole (1999), who consider an agent choosing a multidimensional action in a general informational environment, comparing marginal incentives as more information about *the state* becomes available. Instead, I compare incentives under different assumptions regarding observability of the agent's *actions*.[21] Their results thus only speak to the direct effect.

Lastly, several papers in economics study incentives in academic publication, cautioning against associating false positives with problems in scientific conduct. Glaeser (2006) studies the incentives behind false positives, and argues that eliminating them may be socially harmful. His focus is instead on the incentives to choose novel hypotheses with exogenous value. Di Tillio, Ottaviani and Sørenson (2017, 2018) similarly develop a Persuasion model to study publication bias. Di Tillio, Ottaviani and Sørenson (2018) in particular compares observed versus unobserved selection (in a model with a single-dimensional action) and show that the observer may be better off when selection is unobserved. Their focus is on the underlying results distribution, and not substitutability per se. Needless to say, this is an application deserving of much more work.

[20] As mentioned, I am primarily interested in cases where no formal relationship between sender and receiver exists, e.g., sender is an independent university researcher.

[21] These are certainly related, as additional information about ability may influence beliefs on how much effort was exerted. But it is not nested as a special case, since they require the additional information to be affiliated with the state conditional on the action.

5.3. Future Directions

This paper can help reconcile divergent views regarding whether the presence of biased research designs should prompt policy responses. For example, though Ioannidis et al. (2014) raise alarm regarding the prevalence of biased experimental designs, Glaeser (2006) argues that bias per se need not be problematic if it is correctly taken into account by consumers of research. Both of these perspectives are consistent with the result in my paper, depending on complementarity in the type of research activity. While I do not claim that my simple model perfectly captures this multifaceted application, I am able to provide some explicit policy guidelines that my model lends support to, but which I believe would be difficult to arrive at without it.

In terms of application, an obvious direction for future work is to develop approaches that are more aptly suited to study other policy proposals that have been advanced among researchers (besides transparency). For instance, how to optimally reward replication without discouraging researcher initiative (as defined in Glaeser (2006)) seems to be an important question left unanswered by the current paper. One could also attempt to speak more directly to what kinds of practices researchers would adopt, instead of treating these as exogenous. And describing how to adapt inference to the presence of publication bias (as in Andrews and Kasy (2019) or Furukawa (2017)) and its interaction with research incentives seems important.

Theoretically, the contribution of this paper is the analysis of limited transparency in a costly communication setting with multidimensional actions. Understanding complementarity of different research actions is necessary to determine how experiment choice changes across transparency regimes. There are many avenues for future work extending this insight.

Two approaches seem most promising. First, the model could be enriched in many of the same directions that follow-on work to Kamenica and Gentzkow (2011) has proceeded. For instance, one could allow for richer private information from the sender or receiver. Allowing for multiple senders also seem important and realistic modifications the machinery of this paper could be adapted to. Second, one could analyze these this kind of limited transparency in more general mechanism design settings. I have taken a stark approach regarding the tools at the designer's disposal, seeking to speak directly and practically to the merits of transparency requirements. In principle, this way of modelling limited verifiability could be taken to other contracting settings where information acquisition is endogenous.

References

[1] Abel Brodeur, Mathias L, Marc Sangnier and Yanos Zylberberg. Star Wars: The Empirics Strike Back American Economic Journal: Applied Economics, 8(1):1-32, January 2016.

[2] Isaiah Andrews and Maximilian Kasy. Identification of and correction for publication bias. American Economic Review, Forthcoming.

[3] Charles Angelucci. Motivating agents to acquire information. Working paper, Columbia Business School, November 2017.

[4] Rossella Argenziano, Sergei Severinov, and Francesco Squintani. Strategic information acquisition and transmission. Working paper, June 2014.

[5] Dirk Bergemann and Ulrich Hege. The financing of innovation: Learning and stopping. *RAND Journal of Economics*, 36(4):719–752, Winter 2005.

[6] Lucas C. Coffman and Muriel Niederle. Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives*, 29(3):61–80, Summer 2015.

[7] Vincent P. Crawford and Joel Sobel. Strategic Information Transmission. *Econometrica*, 50(6):1431–1451, November 1982.

[8] Jacques Cremer. Arm's length relationships. *The Quarterly Journal of Economics*, 110(2):275–295, May 1995.

[9] Alfredo Di Tillio, Marco Ottaviani and Peter Norman Sørenson. Persuasion Bias in Science: Can Economics Help? The Economic Journal, 127:266-304, October 2017.

[10] Alfredo Di Tillio, Marco Ottaviani and Peter Norman Sørenson. Strategic Sample Selection. Working Paper, Bocconi University, July 2018.

[11] Mathias Dewatripont, Ian Jewitt, and Jean Tirole. The economics of career concerns, part i: Comparing information structures. *Review of Economic Studies*, 66(1):183–198, January 1999.

[12] Mike Felgenhauer and Elisabeth Schulte. Strategic private experimentation. *American Economic Journal: Microeconomics*, 6(4):74–105, November 2014.

[13] Matthew Gentzkow and Emir Kamenica. Costly persuasion. *American Economic Review: Papers and Proceedings*, 104(5):457–462, May 2014.

[14] Edward Glaeser. Researcher incentives and empirical methods. In Andrew Caplin and Andrew Schotter, editors, *The Foundations of Positives and Normative Economics*, pages 300–319. Oxford University Press, Oxford, 2008.

[15] Yingni Guo and Eran Shmaya. Costly Miscalibration. Working paper, Northwestern University, July 2018.

[16] Emeric Henry. Strategic disclosure of research results: The cost of proving your honesty. *The Economic Journal*, (119):1036–1064, July 2009.

[17] Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182, January 1999.

[18] John P. A. Ioannidis et al. Increasing value and reducing waste in research design, conduct and analysis. *The Lancet*, 383(9912):166–175, January 2014.

[19] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, October 2011.

[20] Xun Lu and Halbert White. Robustness Checks and Robustness Tests in Applied Economics. *Journal of Econometrics*, 178(1):194–206, January 2014.

[21] Nicola Lacetera and Lorenzo Zirulia. The economics of scientific misconduct. *Journal of Law, Economics and Organization*, 27(3):215–260, October 2011.

[22] Elliot Lipnowski and Doron Ravid. Cheap Talk under Transparent Motives. Working paper, University of Chicago, May 2019.

[23] Elliot Lipnowski, Doron Ravid and Denis Shishkin. Persuasion via Weak Institutions. Working paper, University of Chicago, April 2018.

[24] Edward Miguel et al. Promoting transparency in social science research. *Science*, 343(6166):30–31, January 2014.

[25] Daehong Min Bayesian Persuasion under Partial Commitment. Working Paper, Korea Information Society Development Institute, October 2018.

[26] Anh Nguyen and Teck Yong Tan. Bayesian Persuasion with Costly Messages. Working paper, Carnegie Mellon University and Nanyang Technological University, December 2018.

[27] Benjamin A. Olken. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3):81–98, Summer 2015.

[28] Di Pei. Communication with endogenous information acquisition. Working paper, Massachusetts Institute of Technology, January 2015.

[29] Andrea Prat. The wrong kind of transparency. *The American Economic Review*, 95(3):862–877, June 2005.

[30] Eva Vivalt. Specification Searching and Significance Inflation Across Time, Methods and Disciplines. Working paper, Australian National University, October 2018.

A. PROOFS

I prove the following Lemma (which allows for more than two possible signals), and note that Lemma 2 follows immediately when pure strategy equilibrium is imposed.

Lemma 4. *In any equilibrium where the sender's choice is inferred as $\tilde{a}$, the sender's payoffs from experiment a are:*

$$\mathbb{E}_{y\sim\tilde{a}}\left[\mathbb{E}_{\theta\sim\hat{p}_{\tilde{a}}(y)}\left[U^S(e_{\tilde{a}}(y),\hat{p}_a(y),\theta)\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right]\right]-c_S(a).$$

Proof of Lemma 4. The sender's payoffs can be expressed as:

$$\mathbb{E}_{y,\theta}[U^S(e_{\tilde{a}}(y),\hat{p}_{\tilde{a}}(y),\theta)]-c_S(a),$$

where the expectation is taken over realizations of θ and y, and $e_{\tilde{a}}$ is the receiver's equilibrium effort as a function of y. Rewriting the expectation as over realizations of y and θ:

$$\begin{aligned}
\mathbb{E}[U^S(e_{\tilde{a}}(y),\hat{p}_{\tilde{a}}(y),\theta)] &= \sum_\theta\left(\sum_y U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\mathbb{P}[y\mid a,\theta]\right)\mathbb{P}[\theta]\\
&= \sum_\theta\left(\sum_y U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\cdot\mathbb{P}[y\mid\tilde{a},\theta]\right)\mathbb{P}[\theta]\\
&= \sum_\theta\left(\sum_y U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\cdot\frac{\mathbb{P}[y\mid\tilde{a},\theta]\mathbb{P}[\theta]}{\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]}\cdot\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right)\\
&= \sum_y\left(\sum_\theta U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\cdot\frac{\mathbb{P}[y\mid\tilde{a},\theta]\mathbb{P}[\theta]}{\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]}\cdot\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right)\\
&= \sum_y\left(\sum_\theta U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\cdot\mathbb{P}[\theta\mid\tilde{a},y]\cdot\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right)\\
&= \sum_y\left(\mathbb{E}_{\theta\sim p(\tilde{a},y)}\left[U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\right]\sum_{\tilde{\theta}}\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]\right)\\
&= \sum_{\tilde{\theta},y}\mathbb{E}_{\theta\sim p(\tilde{a},y)}\left[U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\cdot\left(\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right)\right]\mathbb{P}[y\mid\tilde{a},\tilde{\theta}]\mathbb{P}[\tilde{\theta}]\\
&= \mathbb{E}_{y\sim\tilde{a}}\left[\mathbb{E}_{\theta\sim p(\tilde{a},y)}\left[U^S(e_{\tilde{a}}(y),\hat{p}(y),\theta)\frac{\mathbb{P}[y\mid a,\theta]}{\mathbb{P}[y\mid\tilde{a},\theta]}\right]\right].
\end{aligned}$$

Essentially, the argument follows from noting that under the full support condition, dividing and multiplying every term in the sum by $\mathbb{P}[y\mid\tilde{a},\theta]$ and $\mathbb{P}_{\tilde{a}}[y]$. The fifth line follows from an

application of Bayes rule, noting that this term is equal to the posterior belief that the state is θ when the chosen experiment is $\tilde{a}$. Finally, in any pure strategy equilibrium, we have both that the sender chooses a and that the receiver infers that the sender chose experiment a. Hence $\mathbb{P}[y \mid a, \theta] = \mathbb{P}[y \mid \tilde{a}, \theta]$, giving the second expression. □

Proof of Lemma 1. Suppose M is the index set of observable indices, and partition the sender's action into $a = (a_M, a_{-M})$. We show that there is some a^*_{-M} such that when the receiver conjectures that a^*_{-M} are the unobserved actions of the sender, the sender's best response is to follow action a^*_{-M}. Since p_0 is interior and, for any choice of experiment, some signal occurs with positive probability in some state, the receiver always puts non-negative probability on observing any $y \in \{0, 1\}$, for any conjecture regarding the sender's behavior. Therefore, there is a unique belief profile $(\hat{p}_a(y))_{y \in Y}$ formed after observing any signal, for any equilibrium strategy of the sender. In fact, since A is compact, we have that $\mathbb{P}[y]$ is bounded away from 0 for all y. This implies that beliefs are a continuous function of actions, and well-defined given any conjecture.

Define the function $\phi(a)$ as follows:

$$\phi(a) = \arg\max_{\tilde{a} \in A_{-M}} \overbrace{\sum_{\theta} \left(\sum_{y} U^S(e_{(a_M, a)}(y), \hat{p}_{(a_M, a)}(y), \theta) \mathbb{P}[y \mid \theta, a_M, \tilde{a}] \right) \mathbb{P}[\theta]}^{(\dagger\dagger)} - c(a_M, \tilde{a}).$$

Note that $\phi(a)$ gives the payoff maximizing response, assuming (observable) actions a_M are chosen and a conjecture of a. We first show this is function is upper hemicontinuous. Take $a_n \to a$, and $b_n \in \phi(a_n)$ with $b_n \to b$.

Note that beliefs are continuous in the sender's action choice, since $\mathbb{P}[y \mid a, \theta]$ is bounded away from 0 on a compact set. We further show that, $e(\hat{p})$ is continuous in $\hat{p}$; if $\hat{p}_n \to \hat{p}^*$, then $e(\hat{p}_n) \to e(\hat{p}^*)$; to see this, note that since effort is a compact set and the receiver's best response is unique, we can ensure $e(\hat{p}_n) \to e^*$, passing to a subsequence if necessary by compactness of the receiver's action set. If e^* does not maximize $\mathbb{E}_{\theta \sim \hat{p}}[U^R(e, \theta)]$, then there is some e^{**} where the receiver does strictly better when the induced belief is $\hat{p}$. But continuity of the receiver payoff function implies that $\mathbb{E}_{\theta \sim \hat{p}_n}[U^R(e(\hat{p}_n), \theta)] \to \mathbb{E}_{\theta \sim \hat{p}}[U^R(e^*, \theta)]$, which implies that e^{**} would be a preferred action choice to $e(\hat{p}_n)$ for some n sufficiently large.

From this, we conclude that $(\dagger\dagger)$ is simply the sum and product of terms that are continuous

in a, and so:

$$\sum_\theta \left(\sum_y U^S(e_{(a_M,a_n)}(y), \hat{p}_{(a_M,a_n)}(y), \theta) \mathbb{P}[y \mid (a_M, \tilde{a}), \theta] \right) \mathbb{P}[\theta] \to^n \sum_\theta \left(\sum_y U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) \mathbb{P}[y \mid (a_M, \tilde{a}), \theta] \right) \mathbb{P}[\theta].$$

If $b \notin \phi(a_{-M})$, then there exists some value δ such that a deviation to δ would result in a higher objective than b, namely we would have:

$$\sum_\theta \left(\sum_y U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) (\mathbb{P}[y \mid \theta, (a_M, \delta)] - \mathbb{P}[y \mid \theta, (a_M, b)]) \right) \mathbb{P}[\theta] > c(a_M, \delta) - c(a_M, b).$$

But since $a^n_{-M} \to a_{-M}$ and $b_n \to b$, by continuity we would be able to find some find some n sufficiently large such that this inequality would also be satisfied replacing b by a_n, which would contradict our assumption that b_n is a maximizer of $\phi(a_n)$. Hence the map ϕ is upper-hemicontinuous.

Furthermore, $\phi(a)$ is nonempty and closed because A_{-M} is compact (being the product of intervals) and the objective function is the difference between a function that is linear in $\mathbb{P}[y \mid a, a_m, \theta]$ and convex in $\mathbb{P}[y \mid a, a_m, \theta]$. Hence there exists some $\tilde{a}$ that maximizes the objective, and the set of maximizers forms a closed set.

Finally, to see that it is convex, suppose that a', a'' are both in $\phi(a_{-M})$. Note that the objective's maximizer is unchanged if we subtract the sender's payoff from the degenerate outcome where no information is conveyed, i.e. replacing (††) with

$$\sum_\theta \left(\sum_y (U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) - U^S(e(p_0), p_0, \theta)) \mathbb{P}[y \mid (a_M, \tilde{a}), \theta] \right) \mathbb{P}[\theta]$$

where $e(p_0)$ is the action chosen at the prior. Now, by Assumption 3, we have $\mathbb{P}[y = 1 \mid \theta, (\tilde{a}, a_M)]$ is concave and remains concave when multiplied by $U^S(e_{(a_M,a)}(1), \hat{p}_{(a_M,a)}(1), \theta) - U^S(e(p_0), p_0, \theta)$ (since this is positive). Likewise, $\mathbb{P}[y = 0 \mid \theta, (\tilde{a}, a_M)]$ and convex and is hence concave when multiplied by $U^S(e_{(a_M,a)}(0), \hat{p}_{(a_M,a)}(0), \theta) - U^S(e(p_0), p_0, \theta)$ (since this is negative). It follows that the objective in the definition of $\phi(a)$ is concave in $\tilde{a}$, meaning that if this expression is maximized at a' and a'', it must also be maximized at every $a''' = \alpha a' + (1 - \alpha) a''$, as desired. Having demonstrated that the conditions for Kakutani's fixed point theorem are satisfied, an

equilibrium exists when a_M is observed, for any choice of a_M.

I now show the claim on mixed strategy equilibria. If there were, then the first order condition must hold for two values of $\tilde{a}$, say $\tilde{a}^1 < \tilde{a}^2$. On the other hand, the receiver's beliefs do not depend on the choice of $\tilde{a}$. Using the last expression for the sender's benefit:

$$\nabla_{\tilde{a}} c(a_M, \tilde{a}^i) = \sum_\theta \left(\sum_y (U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) - U^S(e(p_0), p_0, \theta)) \nabla_{\tilde{a}} \mathbb{P}[y \mid (a_M, \tilde{a}^i), \theta] \right) \mathbb{P}[\theta],$$

and hence subtracting the equation for $i = 1$ from the equation for $i = 2$, and taking the dot product for some arbitrary α with $||\alpha|| = 1$:

$$\alpha \cdot \nabla_{\tilde{a}} (c(a_M, \tilde{a}^2) - c(a_M, \tilde{a}^1)) =$$
$$\sum_\theta \left(\sum_y (U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) - U^S(e(p_0), p_0, \theta)) \nabla_{\tilde{a}} (\mathbb{P}[y \mid (a_M, \tilde{a}^2), \theta] - \mathbb{P}[y \mid (a_M, \tilde{a}^1), \theta] \right) \mathbb{P}[\theta].$$

By the mean value theorem, applied to $\mathbb{P}[y \mid a, \theta]$ and c, for some $a_{y,\theta}$ and a_c which are all convex combinations of $\tilde{a}^1$ and $\tilde{a}^2$ such that:

$$\alpha \cdot (\nabla^2_{\tilde{a}} c(a_M, a_c)(\tilde{a}^2 - \tilde{a}^1)) =$$
$$\sum_\theta \left(\sum_y (U^S(e_{(a_M,a)}(y), \hat{p}_{(a_M,a)}(y), \theta) - U^S(e(p_0), p_0, \theta)) \nabla^2_{\tilde{a}} \mathbb{P}[y \mid (a_M, a_{y,\theta}), \theta] \cdot (\tilde{a}^2 - \tilde{a}^1) \right) \mathbb{P}[\theta].$$

But by the strictness of concavity or convexity, either the left hand side is strictly positive or the right hand side is strictly negative, with both being at least weakly so, a contradiction. Hence in equilibrium, there can only be pure strategies. □

Proof of Proposition 1. Note that the first order conditions (potentially as an inequality if non-interior) characterize the sender's equilibrium action choice, as per the proof of Lemma 1. The sender's payoff (in equilibrium) can be written:

$$\sum_y \sum_\theta U^S(\hat{p}_{\tilde{a}}(y), \theta) \mathbb{P}[y \mid a, \theta] \mathbb{P}[\theta] - c_S(a).$$

When a is unobservable, then the action and the inferred belief do not change as a changes. Using that the experiment choice is deterministic, the first order condition following a correct inference by the receiver is:

$$\frac{dc_S(a)}{da} = \sum_y \sum_\theta U^S(\hat{p}_a(y), \theta) \frac{d\mathbb{P}[y \mid a, \theta]}{da} \mathbb{P}[\theta].$$

When dimension a is observable, the added term corresponds to the change in the receiver's belief about the state. In that case, using that the beliefs are differentiable as a function of the action, chain rule gives us that the added term is:

$$\sum_y \sum_\theta \nabla_{\hat{p}} U^S(\hat{p}_a(y), \theta) \cdot \frac{d\hat{p}_a(y)[\cdot]}{da} \mathbb{P}[y \mid a, \theta] \mathbb{P}[\theta],$$

which is (1). (The brackets reflect that $\hat{p}$ is a belief over many states) The interiority assumption implies the first order condition holds when a is observed, say at action a^*_{obs}. If (1) is positive, then:

$$\sum_y \sum_\theta U^S(\hat{p}_a(y), \theta) \frac{d\mathbb{P}[y \mid a, \theta]}{da_2} \mathbb{P}[\theta] - \left. \frac{dc_S(a)}{da} \right|_{a=a^*_{obs}} < 0,$$

since this holds with equality when (1) is added. Note that the objective, as a function of a, is concave in every coordinate under Assumptions 1-3, meaning that the left hand side is decreasing in a. It follows that for the first order conditions to hold, the resulting a must be lower. Hence if (1) is positive, then keeping a hidden lowers it, so that the choice of a is higher under observability, as claimed. Other cases are analogous. □

For the previous proof, note that if $a = \underline{a}$, then the first order condition for the sender's actions hold as inequality; if (1) is sufficiently large, then these inequalities would be violated at $a = \underline{a}$. Hence the equilibrium action choice would necessarily be higher, noting that Lemma 1 implies that the first order condition must be satisfied for some choice of a.

Proof of Proposition 2. Denote by $a_i^*(a_{-i})$ the equilibrium response of a_i, fixing the choice of a_{-i}. By Lemma 1, this is characterized by the first order condition:

$$\frac{\partial}{\partial a_i} c_S(a_{-i}, a_i^*(a_{-i})) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_i} \mathbb{P}[y \mid (a_{-i}, a_i^*(a_{-i})), \theta] \mathbb{P}[\theta], \tag{6}$$

with equality holding whenever a_i is interior. Using that Y is binary and evaluating at $a_{-i} = \tilde{a}_{-i}$, we rewrite this as:

$$\frac{\partial}{\partial a_i} c_S(\tilde{a}_{-i}, a_i^*(\tilde{a}_{-i})) \leq \sum_\theta (U^S(\hat{p}(1), \theta) - U^S(\hat{p}(0), \theta)) \frac{\partial}{\partial a_i} \mathbb{P}[y = 1 \mid (\tilde{a}_{-i}, a_i^*(\tilde{a}_{-i})), \theta] \mathbb{P}[\theta]. \tag{7}$$

First suppose that the first order condition defining $a_i^*(\tilde{a}_{-i})$ holds with equality. Then adding $\frac{\partial c_S(a^*_{-i}, a_i)}{\partial a_i} - \frac{\partial c_S(\tilde{a}_{-i}, a_i)}{\partial a_i} \leq M_{a^*_{-i}}(a_i) - M_{\tilde{a}_{-i}}(a_i)$ to both sides of (7) when it holds with equality yields:

$$\frac{\partial}{\partial a_i} c_S(a^*_{-i}, a^*_i(a^*_{-i})) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_i} \mathbb{P}[y \mid (a^*_{-i}, a^*_i(a^*_{-i})), \theta] \mathbb{P}[\theta], \tag{8}$$

Using the fact that $c(a_{-i}, a_i)$ is convex in a_i, and that $\mathbb{P}[y = 1 \mid (a_{-i}, a_i), \theta]$ is concave in a_i, it follows that given an inferred choice of a_i must be higher when a^*_{-i} is chosen, and strictly so when a_i is interior and the inequality is strict. If the first order condition holds as a strict inequality, then a_i is chosen as an edge case, and the same reasoning implies that a_i could only increase as well. □

Proof of Theorem 1. I first note that the receiver's payoff function is continuous in a_2, which follows immediately from continuity assumption on the sender's experiment choice. More precisely, since the proof of Lemma 1 shows that effort is continuous in posterior beliefs, as well as that beliefs are continuous in a_2, it follows that continuity of receiver's payoffs are maintained across transparency regimes.

Since the receiver's payoff increases in a_1, given any *fixed* choice of a_2, there exists a discrete increase in the receiver's payoffs, say α, when the sender chooses a higher a_1. By continuity of the receiver's payoff function, we can find some ε such that an increase from a_1 to the next largest value also delivers a higher payoff, whenever the increase in a_2 is no more than ε. Hence there is slack, in the sense that we need only show that a_2 increases at most slightly (and not necessarily exactly 0) when moving across regimes, given the conditions of the Theorem.

The rest follows immediately from the same arguments as in Propositions 1 and 2. Denoting the equilibrium response of a_2 given an (observable) choice of a_1 by $a^*_2(a_1)$, given (1) sufficiently positive, Proposition 1 shows that $a^*_2(a^{obs}_1)$ approaches $\max A_2$ when choosing a^{obs}_1. By assumption, this lower's sender's payoff relative to any other action. From Proposition 2 it follows that a^*_2 decreases when a_1 increases. On the other hand, a higher choice of a_1 yields an equilibrium choice of a_2 not significantly larger than that chosen under observability. Hence as discussed above, continuity of the receiver's payoff function implies this leads to an increase in payoffs, proving the theorem.

The reasoning for the converse case is identical and hence omitted. □

Proof of Lemma 3. This follows from demonstrating that the sender's payoffs, as a function of the receiver's beliefs, are convex, using Lemma 2 to determine this expression. This is immediate in the case of polynomial effort costs. Indeed, take the second derivative of $pe(p)$ (the benefit from follow-on effort) and observe that it is equal to:

$$\lambda(2e'(p) + pe''(p)).$$

Since $c_R(e)$ is strictly convex, $e'(p) > 0$ since the receiver's first order condition is:

$$bp = c_R'(e(p)).$$

Differentiating with respect to p gives:

$$b = c_R''(e(p))e'(p),$$

and differentiating again gives:

$$0 = c_R'''(e)(e'(p))^2 + c_R''(e)e''(p).$$

Since $e(p)$ is strictly increasing, the assumptions on $c_R'''(e)$ ensure that $e''(p) \geq 0$, and hence the objective is convex. □

In general, convexity of receiver effort by itself is not a strong enough assumption in order to ensure that $pe(p)$ is convex. To see this, suppose that:

$$c_R(e) = 1 - \sqrt{1-e} \Rightarrow c_R'(e) = \frac{1}{2\sqrt{1-e}} > 0 \Rightarrow c_R''(e) = \frac{1}{4(1-e)^{3/2}} > 0.$$

In that case:

$$e(p) = \max\{0, -\frac{1}{4b^2p^2} + 1\},$$

and observe that $pe(p)$ is concave whenever $e(p) > 0$.

B. MISCELLANEOUS

B.1. Counterexample to Lemma 2 when Assumptions are Violated

I briefly demonstrate, by example, on the possibility of a failure of pure strategy equilibrium existence when results cannot be classified into positive and negative results. The failure arises due to a failure of concavity in the objective stated in Lemma 2. But the point of this example is to show that the real technical issue arises when results cannot be classified into positives or negatives independently of the experiment choice. In this example, the experiment set is convex, so the failure of the pure strategy existence is driven by costs.

Let $|\Theta| = |Y| = 2$, with $\Theta = \{-1, 1\}$ and $\mathbb{P}[\theta = 1] = 1/2$. Consider the following sender preferences:

$$U^S(\hat{p}, \theta) = -\hat{p}[\theta = 1] \cdot \theta$$

And let:

$$\mathbb{P}[Y = 1 \mid \theta = 1] = \mathbb{P}[Y = 0 \mid \theta = -1] = 1 - \mathbb{P}[Y = 1 \mid \theta = -1] = 1 - \mathbb{P}[Y = 0 \mid \theta = 1] = a^2.$$

with $c(a) = a/4$. In this example, in state $\theta = 1$, the event $Y = 1$ is a positive result when $a < \frac{1}{\sqrt{2}}$ is inferred (in which case it is evidence for the state $\theta = -1$), and a negative result otherwise. In state $\theta = -1$, this is flipped. So when $a > \frac{1}{\sqrt{2}}$, the concavity assumption is satisfied in state $\theta = 1$ but violated in state $\theta = -1$, and the opposite is true when $a < \frac{1}{\sqrt{2}}$.

Write the payoff to the sender from an experiment a when it is inferred as $\tilde{a}$ (noting that $a = \tilde{a}$ in equilibrium). Since everything is symmetric, the probability of a positive result $1/2$ ex-ante, for any choice of experiment. Hence the payoff is:

$$-\frac{1}{4}a + \frac{1}{2}\left(-\tilde{a}^2 \cdot \left(a^2\right) + \tilde{a}^2\left(1 - a^2\right) - (1 - \tilde{a}^2)(1 - a^2) + (1 - \tilde{a}^2) \cdot a^2\right).$$

which reduces to:

$$-\frac{1}{4}a + \frac{1}{2}\left(-4\tilde{a}^2a^2 + 2\tilde{a}^2 + 2a^2 - 1\right).$$

Given a conjecture of $\tilde{a}$, sender chooses a to maximize $-\frac{a}{4} + a^2(1 - 2\tilde{a}^2)$. Note that if $\tilde{a}^2 \geq \frac{1}{2}$, this is maximized at $a = 0$, since the objective is negative for all other values of a. But if $\tilde{a}^2 < \frac{1}{2}$, I have that the second derivative of this expression is positive. Since this is a quadratic function, the optimum is either 0 or 1, for any choice of $\tilde{a}$. So the only choice of a that would be a best response when inferred correctly would be $a = 0$, but $a = 1$ is a best response to $\tilde{a} = 0$.

B.2. Preferences over y

In this appendix, I comment on a modification to the model where I allow for the sender to have preferences over y itself.

For simplicity, first consider for simplicity the case where the payoffs are separable, and the sender obtains an added benefit of $\lambda_y \cdot y$. from a positive result. In general, this model still is amenable to the belief-based approach, noting that any positive result leads to a higher belief and any negative result leads to a lower belief. Hence this setting is as if there were a jump in the sender's payoff function at the prior (as commented on in Footnote 2). That said, it is simplest to comment on this case simply by inspection. In this case, it is immediate that the sender is incentivized to maximize the biasing action in this case (whether higher informative actions will be taken depends on the prior):

Proposition 3. *As $\lambda_y \to \infty$, the sender's choice of experiment converges to the one which maximizes $\mathbb{P}[y = 1]$.*

However, I also comment that transparency does not interact with the experiment choice when all that matters is whether the result is positive or not (and would similarly expect a limited impact if this consideration itself was overwhelmingly dominant):

Proposition 4. *Suppose the sender's payoffs $U^S(y, e, \hat{p}, \theta)$ is constant in e and $\hat{p}$. Then the sender's experiment choice does not differ depending on transparency regime.*

This is immediate since the (ex-ante) probability of an outcome y, conditional on the experiment, does not depend on transparency regime. Hence neither do payoffs if all that matters for the scientist is the probability of a positive result.

While these results are theoretically immediate, one may still find it counterintuitive since the preference for positive results appears so widespread that it is tempting to think it is intrinsic. The paper takes the view that the definition of "positive" or "negative" are endogenous and depend on the belief movement. If a preference emerges entirely because negative results are harder to publish than positive results, then this would suggest an *interaction* between positive results and the other payoff terms. This would not change the results drastically in the paper (due to the added non-convexity), but would result in additional notation. Also, it would make it more difficult to characterize the preference for information in certain places (e.g., Section 2), without changing intuition. It may also be that positive results that are obtained "cheaply" (via bias) are less meaningful, but those that are achieved "scrupulously" (via informativeness) are more meaningful. This would suggest greater interdependence between the cost function and the benefit than what I have here. While these observations may call for more empirical commentary to determine the extent to which that's the case, this is left for future work.