False Positives and Transparency

Jonathan Libgober

Department of Economics, University of Southern California

April 15, 2020

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, while a receiver observes the experiment's outcome (though not necessarily all dimensions). I show that 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, Odilon Câmara, 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. For instance, Ioannidis et al. (2014) mention facilitating access to code or data, or strengthening review among funding agencies. Such proposals are relevant to any empirical discipline.

Understanding when transparency helps or hurts consumers of research is thus of real practical interest. And a contribution of this paper is to provide 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[1] is unobservable. My analysis suggests the answer is no, and that while pre-specifying 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 activities which lead to divergent conclusions.

The theoretical model I study is a sender-receiver game that features the aforementioned multidimensionality in experiment choice. The sender (she) 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

[1]My formal definition of a *biasing activity* is one that increases the probability of a positive result while decreasing experiment informativeness.

are feasible.[2] The experiment is of interest to the receiver because it provides information on his 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.[3] 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 more receiver effort. Unobserved choices will thus be assumed 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.

[2]Experiment costs will also play a role in my model, and these are absent from Kamenica and Gentzkow (2011). But Gentzkow and Kamenica (2014) introduce costs using similar techniques. My model of costs differs slightly.

[3]I introduce the third party to emphasize that the receiver has no direct control over sender actions, and to highlight the limited tools available to influence the sender's actions. In practice, third parties such as journals or funding agencies are cited as capable of implementing such requirements, although potentially with additional costs. The ability to choose effort as a function of the experiment would trivialize the analysis to a large extent; See Section 4.3 for a discussion. Abrams, Libgober and List (2020) discuss registration requirements for experiments in economics, and compare to other disciplines. I return to the question of enforcing transparency requirements in the conclusion, taking as given for the analysis that enforcement is feasible and costless for the third party.

My general results study the substitution that emerges as the observable dimensions change, considering more general preferences and experiment specifications. I refer to a particular specification of the set of observable dimensions as a *transparency regime*. I highlight that the difference in equilibrium experiment choice across transparency regimes depends on two factors: First, how transparency over a given dimension influence the action on that dimension holding others fixed, and second, the complementarity (or substitutability) 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 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.

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, something that may generally emerge under limited transparency. I discuss the necessary concavity assumptions to recover pure strategies and articulate the differences in information preferences between sender and receiver. I highlight that the characterization above makes use of a particular restriction on off-path beliefs which do not arise when the sender has full commitment. I discuss these issues in Section 3.5, relating my work to other limited commitment versions of Bayesian Persuasion (such as Felgenhauer and Loerke (2017), who also discuss the role of off-path beliefs in a similar setting).

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 suggested that the merits of transparency over biasing activity differs for choice of controls versus choice of robustness checks. I obtain this conclusion by showing that the indirect effect goes in opposite directions in each instance. The basic intuition is straightforward. As I discuss in Section 4.2, consumers of research may be less worried about discretionary choices regarding robustness checks (e.g., alternative assumptions on error distributions) when the underlying design is more informative. If the estimate is highly precise, then it might be unusual for a different set of robustness checks to significantly alter the result. This reasoning suggests the direction of the indirect effect; namely, more informative experiments are less susceptible to manipulation and bias. The conclusion differs when it comes to the choice of controls. Specifically, experiments with more potential covariates also have more scope for specification searching. With more controls, researchers inevitably have more choices to make regarding which regressions to consider, and thus more potential ways to obtain positive results. This flips the indirect effect; now, it is the more informative experiments that are more susceptible to manipulation and bias. My model articulates why this distinction

implies transparency may be beneficial in the latter case but not the former case. I highlight another subtlety emerging from the model by deriving comparative statics regarding the relative importance of various incentives researchers might face. I note that the compensation effect I highlight is more significant when impact (for instance, a larger amount of follow-on effort) is relatively more important than perceptions (for instance, the belief itself that the researcher's hypothesis is true). It is natural to conjecture that when researchers are less established, the latter is more significant. The incentive to compensate for the perception of bias would therefore be more significant when researchers are more established. Using these comparative statics, I suggest that optimal transparency policy may depend on career stage.

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. This facilitates discussion regarding the relative importance of various incentives, and highlights some subtle factors influencing the incentive to bias. To the best of my knowledge, these features are missing elsewhere 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 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 the 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 $a = (a_1, a_2)$ with corresponding cost $c_S(a)$. 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 a_i to be chosen from a compact set $A_i \subset [0, 1]$, and refer to the set of signal distributions and the associated costs as the *experimentation technology*. For simplicity, I will take A_1 to be finite, writing $A_1 = \{\alpha_1, \cdots, \alpha_n\}$. The results in Section 3 will require $\mathbb{P}[y \mid a, \theta]$ and $c_S(a)$ to be differentiable in a_2 (and I introduce this assumption when providing conditions for pure strategy existence), although I illustrate the key ideas in Section 2 taking A_2 to be binary. Below, I call a_i *biasing* if increases make positive results more likely while decreasing the experiment's informativeness overall (holding the other dimension fixed), in a Blackwell sense.

When sender's experiment choice is explicit, I 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 probability the receiver assigns to the event 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, and strictly concave in e 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 the sender's payoff to depend *both* on the receiver's action choice e, as well as $\hat{p}$, the receiver's belief itself. Hence I write the sender's (ex-post) 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.[4]

However, the reader should understand that, since any equilibrium e is a *function* of the receiver's belief, 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, who determines whether a_2 is observed by the receiver. I will study how

[4]All results are maintained if the 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.

equilibrium experiment choice changes depending on the third party's decision. To emphasize, the only choice of the third party is whether the receiver observes a_2. Importantly, I assume that the sender *cannot* make these coordinates observable unless the third party requires it. I discuss this assumption further in the conclusion. Alternatively, one could impose the existence of fixed costs of transparency for the sender, making transparency prohibitive without a requirement.[5] If the (sender's) costs of making experiments observable are sufficiently large, this would not change this paper's results. If a_2 is not observed, I refer to the transparency regime as *limited transparency*. *Full transparency* refers to a transparency regime where both 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 describes 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.

While the above model assumes a two-dimensional action, in fact many of the results (e.g., Lemma 1) apply more generally. The main difficulty with the general case is that the comparisons across transparency regimes become more cumbersome to state. This is natural, since more possible dimensions can potentially adjust. Since the same intuition behind the welfare comparisons will apply even in this case, I focus on the simplest setting in this paper.

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

[5]For instance, Olken (2015) discusses the costs of writing pre-analysis plans, a common way for some researchers to establish transparency on their own. He suggests these costs may be significant, depending on the context.

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.

1.1. Examples of Experimentation Technologies

Before illustrating the main ideas, it is useful to introduce a few examples of experimentation technologies which are nested within the above framework. Both of these examples assume a binary state $\theta \in \{T, F\}$, and will be useful in interpreting and discussing the main results and implications of the model:

Example 1 (p-hacking)**.** *Suppose the sender chooses a number of times to run an (iid) experiment with a precision $a_1 \in A_1 \subset (1/2, 1]$, with the receiver observing whether at least one success is observed. To model this, take $A_2 = [\underline{a}_2, 1]$ with $\underline{a}_2 > 0$, and let:*

$$\mathbb{P}[y = 1 \mid \theta = T, a_1, a_2] = 1 - (1 - a_1)^{a_2/\underline{a}_2} \text{ and } \mathbb{P}[y = 1 \mid \theta = F, a_1, a_2] = 1 - a_1^{a_2/\underline{a}_2}.$$

When $a_2 = \underline{a}_2$, this corresponds to running the experiment a single time, with the experiment able to be repeated up to at most $1/\underline{a}_2$ times. Thus, a_2 is interpreted as the amount of p-hacking.[6]

Example 2 (Flipping Negatives)**.** *Suppose that a_1 determines the baseline probability that $y = 1$, in each state, but assume that a_2 denotes the probability that some outcome $y = 0$ is switched to an outcome $y = 1$—for example, as a result of either direct falsification or altering the data. Hence:*

$$\mathbb{P}[y = 1 \mid \theta = T, a_1, a_2] = f(a_1) + (1 - f(a_1))a_2 \text{ and } \mathbb{P}[y = 1 \mid \theta = F, a_1, a_2] = g(a_1) + (1 - g(a_1))a_2.$$

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 choose e 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:

[6] See Di Tillio, Ottaviani and Sørenson (2018) for a model of scientific experimentation similar to this specification. Felgenhauer and Schulte (2014) study a similar experimentation technology, where a_1 is exogenous and fixed and the sender can disclose the *number* of successes (and not just whether there is a success). See 5.2 for further discussion.

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

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

$$\begin{aligned}
&\mathbb{P}[y=1 \mid \theta = T, a_1 = 0, a_2 = 0] = 2/5, && \mathbb{P}[y=1 \mid \theta = F, a_1 = 0, a_2 = 0] = 0,\\
&\mathbb{P}[y=1 \mid \theta = T, a_1 = 0, a_2 = 1] = 1/2, && \mathbb{P}[y=1 \mid \theta = F, a_1 = 0, a_2 = 1] = 1/6,\\
&\mathbb{P}[y=1 \mid \theta = T, a_1 = 1] = 3/4, && \mathbb{P}[y=1 \mid \theta = F, a_1 = 1] = 0.
\end{aligned}$$

Take the sender's cost to be $c_S(a_1, a_2) = c \cdot a_1$. Notice that choosing $a_1 = 1$ leads to a more informative experiment than either experiment with $a_1 = 0$. Furthermore, increasing a_2 increases the probability of $y = 1$ in both states when $a_1 = 0$. In fact, the $a = (0,1)$ experiment can be generated by first drawing y according to the $a = (0,0)$ experiment, and subsequently changing $y = 0$ to $y = 1$ with probability 1/6. It is thus a special version of Example 2 from Section 1.1. Hence the experiment is less (Blackwell) informative, and thus biasing (as defined in the previous section). On the other hand, a_2 has no impact when $a_1 = 1$.

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 $\theta = T$ 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:

$$\pi_R(1, a_2) = 5/26, \quad \pi_R(0,0) = 1/8, \quad \text{and} \quad \pi_R(0,1) = 1/12.$$

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.[7]

I now show that, if $c > 0$, the receiver may be better off when a_2 is unobservable. Note that this clearly cannot be the case when $c = 0$, as the sender chooses the receiver's most preferred

[7]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).

experiment in both transparency regimes (in the unique equilibrium under partial transparency). But if c is neither too high nor too low, the receiver strictly prefers to keep a_2 unobserved.[8] 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 $c < \pi_R(1, a_2) - \pi_R(0, 0)$.

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 it 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 $c < \pi_R(1, a_2) - \pi_R(0, 1)$.

To see that limited transparency can be beneficial, compare the threshold for c above which the sender chooses $a_1 = 0$. Since $\pi_R(0, 1) < \pi_R(0, 0)$, 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. This occurs when $c \in (\pi_R(1, a_2) - \pi_R(0, 0), \pi_R(1, a_2) - \pi_R(0, 1))$. 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 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.

I conclude by highlighting a feature of this example which distinguishes it from other settings (discussed in detail in Section 5.2) where signal biasing emerges. Specifically, I emphasize that here, the sender's preference for more informative experiments is crucial for the optimality of

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

partial transparency. The incentive to bias emerges simply due to the fact that the sender does better following $y = 1$ versus $y = 0$. But in order for partial transparency to be optimal, it is additionally important that the sender's payoff is convex in the receiver's belief (which, as alluded to above, I show formally in Lemma 2). To make this point concretely, keep the example otherwise unchanged, but suppose instead the sender's preferences were:

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

Since $e(\hat{p}_a(y)) = \hat{p}_a(y)$, the sender's payoff is strictly convex in $\hat{p}_a(y)$ whenever $\gamma > 1$, and linear when $\gamma = 1$. The previous sender preferences (where partial transparency is optimal for some range of c) turns out to coincide with the case of $\gamma = 2$. But convexity decreases as γ decreases, and thus so does the preference for more informative experiments. And when $\gamma = 1$, the sender has *no* preference for more informative experiments, obtaining the same payoff from $a = (1, 0)$ as $a = (0, 0)$ when $c = 0$.[9] However, the receiver would infer that $a_2 = 1$ after observing $a_1 = 0$ as long as $\gamma > 0$, since again, the sender obtains higher payoffs following $y = 1$ than following $y = 0$. When $\gamma = 1$, the incentive to bias emerges, but the incentive to compensate does not. And when $\gamma = 1$ and $c > 0$, the sender chooses $a = (0, 0)$ in the receiver-optimal equilibrium under full transparency, but chooses $a = (0, 1)$ in the unique equilibrium under partial transparency. Thus full transparency is optimal for all $c > 0$. Without the preference for information acquisition, the sender does not gain from counteracting the perception of bias. This preference is what provides scope for partial transparency to be beneficial.

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 consider. 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. This could provide some additional scope for transparency to benefit the receiver, although in that case the comparison between the direct and the indirect effect would

[9]This observation follows immediately from the martingale property of beliefs.

be insufficient to characterize welfare implications.

In any case, a pure strategy equilibrium can be ensured with some reasonably innocuous assumptions. 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. *The cost function $c_S(a)$ has continuous derivatives in a_2 and is convex in a_2, with A_2 being a convex set.*

Assumption 3. *For every a and all θ, $U^S(\hat{p}_a(1), \theta) \geq U^S(\hat{p}_a(0), \theta)$ (with strict inequality for some θ). In addition, $\mathbb{P}[y = 1 \mid a, \theta]$ is concave in a_2 with continuous derivatives in a_2.*

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.

While convexity of A_2 in Assumption 2 seems to rule out examples such as the one in Section 2, the action set can be convexified by taking $a_2 \in (0, 1)$ to denote the probability of choosing $a_2 = 1$. However, note that this specification assumes that if $a_2 \in (0, 1)$ is chosen under full transparency, then the receiver observes this *probability* and not the realized experiment choice. Note that when $|A_2| = 2$, both cost and probability of $y = 1$ are linear in a_2, and hence satisfy the differentiability requirement as well.

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 where a_2 is correctly inferred (on and off path), and no mixing over a_2 when either convexity is strict in Assumption 2 or concavity is strict in Assumption 3.*

The condition that the inference is correct off-path means that the receiver's conjecture, following any a_1, must be the same as if the sender were instead *forced* to choose that particular a_1. Equivalently, this outcome would be part of any PBE with the sender choosing experiment dimensions sequentially, with a_1 being chosen before a_2 under partial transparency. See Section 3.5 for a discussion of this assumption, and equilibria where the choice of a_2 may be incorrect off-path.

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 the inference is required to be correct. When the sender is mixing over *an unobserved experiment dimension*, sender's preferences over information depend on both the realized experiment and the receiver's conjecture.[10] 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.[11] 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 clarifies that the sender's preferences payoffs (gross costs) are:

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

To emphasize, the assumptions are satisfied under many natural experimentation environments. Regarding preferences, the key assumptions are that the experiment costs are convex and differentiable, and that the sender prefers positive results to negative results. The latter holds whenever positive results are more indicative of "good states" for the sender (for instance, the receiver works harder when $\theta=T$ is more likely). Assumption 1 is satisfied in both examples in Section 1.1 (restricting a_2 in Example 2 to ensure not all results are flipped). The concavity assumption holds as well, provided a_2 is continuous.[12]

[10] The Appendix proves a version of Lemma 2 without correct inference requirement, illustrating its role. Note that by itself, the fact that sender and receiver share beliefs does not prevent the belief based approach from being applied, as shown by Alonso and Câmara (2016), who extend Kamenica and Gentzkow (2011) to the case where sender and receiver hold different priors. However, this is distinct from the case where the sender mixes over experiment choice, since then the disagreement is over *experiment chosen*. See the Appendix for further discussion.

[11] 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.

[12] For Example 1, I compute $\frac{\partial^2\mathbb{P}[y=1|\theta=T,a_1,a_2]}{\partial a_2}=-\frac{(1-a_1)^{a_2/\underline{a}_2}}{\underline{a}_2^2}(\log(1-a_1))^2<0$, as well as $\frac{\partial^2\mathbb{P}[y=1|\theta=F,a_1,a_2]}{\partial a_2}=$

3.2. Single-Dimensional Experiment Choice

Returning to the question of how transparency influences experiment choice, I first consider the incentive to change a_2 as transparency over a_2 changes. I do this by studying the sender's choice when her action along the first dimension is degenerate. That is, I take $|A_1| = 1$, which also describes a counterfactual where the choice along the first dimension is *fixed and exogenous*. Furthermore, the equilibrium I focus on requires that the inference made following any a_1 is the same inference that would be made if this choice were the only one possible. Hence this section also describes the choice of a_2 following an off-path choice of a_1.

This yields a characterization of the *direct effect*. The Proposition follows from a comparison between the (sender's) first-order condition when (receiver) beliefs adjust to when they do not:

Proposition 1 (The Direct Effect)**.** *Suppose Assumptions 1-3 hold and suppose $|A_1| = 1$, with the sender using a pure strategy under limited transparency. Then making a_2 observable results in a weakly higher (resp. lower) choice of a_2 whenever:*

$$E_{y\sim a}\left[\mathbb{E}_{\theta\sim\hat{p}_a(y)}\left[\nabla_{\hat{p}}U^S(\hat{p}_a(y),\theta)\cdot\left(\frac{d\hat{p}_a(y)[\theta]}{da_2}\right)_{\theta\in\Theta}\right]\right] \tag{1}$$

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 $\left(\frac{d\hat{p}_a(y)[\theta]}{da_2}\right)_{\theta\in\Theta}$ (in expectation). If changes in the payoffs move in the same direction on average as the changes in posterior beliefs, then the action is encouraged under observability.[13] The proposition presents a sufficient condition, as a function of model primitives (preferences and information structure) for general environments, which ensures the chosen a_2 is higher (or lower) when unobserved.

To be more concrete regarding the content of this observation, let $\theta \in \{0,1\}$ with $U^S(\hat{p},\theta) = \hat{p}[\theta = 1]$, i.e., the posterior belief that $\theta = 1$. Then $\nabla_{\hat{p}}U^S(\hat{p},\theta) = (1,-1)$, for all θ, allowing the expectation over θ to be dropped. Hence whether the action is encouraged depends on the sign of:

$$\mathbb{E}_{y\sim a}\left[\mathbb{E}_{\theta\sim\hat{p}}\left[(1,-1)\cdot\begin{pmatrix}\frac{d\hat{p}_a(y)[\theta=1]}{da_2}\\ -\frac{d\hat{p}_a(y)[\theta=1]}{da_2}\end{pmatrix}\right]\right] = 2\cdot\mathbb{E}_{y\sim a}\left[\frac{d\hat{p}_a(y)[\theta=1]}{da_2}\right]. \tag{2}$$

$-\frac{(a_1)^{a_2/a_2}}{a_2^2}(\log(a_1))^2 < 0$. For Example 2, $\mathbb{P}[y=1 \mid \theta, a_1, a_2]$ is linear in a_2 and hence weakly concave.

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

Suppose that higher a_2 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, $\frac{d\hat{p}_a(y)}{da_2}$ is relatively larger (and positive)—so transparency increases a_2. 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 $U(\hat{p}, \theta) = \hat{p}[\theta = 1] \cdot \mathbf{1}[\theta = 1]$ (as in the focus of Section 2, with $\theta = 1$ replacing $\theta = T$), then $\nabla_{\hat{p}} U(\hat{p}, 0) = 0$, making (2):

$$\mathbb{E}_{y \sim a}\left[\hat{p}_a(y)[\theta = 1] \cdot \left[(1, -1) \cdot \begin{pmatrix} \frac{d\hat{p}_a(y)[\theta=1]}{da_2} \\ -\frac{d\hat{p}_a(y)[\theta=1]}{da_2} \end{pmatrix}\right]\right] = 2 \cdot \mathbb{E}_{y \sim a}\left[\hat{p}_a(y)[\theta = 1] \cdot \frac{d\hat{p}_a(y)[\theta = 1]}{da_2}\right]. \tag{3}$$

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 a y signal that reveals that $\theta = 0$ (or $\theta = F$) 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 $\hat{p}_a(1)[\theta = 1] > \hat{p}_a(0)[\theta = 1]$).

In simple examples, there may be more immediate ways to infer that a_2 is higher when unobserved than by computing (1); for instance, this was deduced straightforwardly in Section 2. On the other hand, the discussion elucidates the properties of the experimentation technology which lead to higher actions being chosen, using reasoning that can be applied more generally.

3.3. Multidimensional Actions and Complementarity

I now reintroduce multidimensionality to discuss the *indirect effect*. Lemma 3 shows how the incentive to adjust an action when it is unobserved differs depending on the choices along another

dimension. This result is quite simple. However, I state it formally as an intermediate result to highlight the driving force behind the gains to partial transparency. 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_2 were receiver beliefs do not adjust:

$$M_{a_1}(a_2) := \sum_{\theta} (U^S(\hat{p}_a(1), \theta) - U^S(\hat{p}_a(0), \theta)) \frac{\partial \mathbb{P}[y = 1 \mid \theta, a]}{\partial a_2} \mathbb{P}[\theta]. \tag{4}$$

Together with the marginal costs for a_2, this expression pins down the choice of a_2 when unobserved, given the choice of a_1. In turn, the a_1 for which the marginal incentives are steeper are less susceptible to being altered:

Lemma 3 (The Indirect Effect)**.** *Suppose Assumptions 1-3 hold, and suppose, for all a_2:*

$$M_{\tilde{a}_1}(a_2) - \frac{\partial c_S(\tilde{a}_1, a_2)}{\partial a_2} \leq M_{a_1^*}(a_2) - \frac{\partial c_S(a_1^*, a_2)}{\partial a_2}. \tag{5}$$

Then the choice of a_2 is (weakly) higher given an observable choice of a_1^* than following an observable (or correctly inferred) choice of $\tilde{a}_1$.

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 the unobserved dimension. 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 $M_1(a_2) = 0 < M_0(a_2)$.

This result highlights the importance of the complementarity (or substitutability) for the implications of partial transparency. Here, complementarity simply refers to whether the incentive to choose higher a_2 is higher when a_1 is higher. The simplicity of this observation makes it amenable to discussion of practical examples of research activity, as in Section 4. However, showing that observed actions themselves change requires conditions on sender's preferences, as imposed in the welfare statement in Proposition 2. But these conditions are not necessary to appreciate the distinguishing feature of the indirect effect, namely that it relates to incentives along a dimension 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:

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

Note that the definition of $\pi_i(a)$ assumes that a is *observed* by the receiver. Since receiver effort given beliefs is unique (by the strict convexity of $U^R(e,\theta)$ in e), this can be computed directly from the payoff function and experimentation technology. 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 main case of interest is where the primary conflict between producers and consumers is time and effort in experimentation.[14] As in Section 2 and highlighted again in Section 4.1, this emerges naturally if incentives for receiver (or follow-on) effort are significant.

The Proposition 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 a sufficiently strong indirect effect—otherwise, welfare comparisons follow from Proposition 1. This result articulates the forces yielding gains to limited transparency in Section 2. Recall that $A_1 = \{\alpha_1, \cdots, \alpha_n\}$:

Proposition 2. *Let (a_1^{obs}, a_2^{obs}) be the sender's choice under full transparency. Suppose Assumptions 1-3 hold, that $\pi_R(a)$ and $\pi_S(a)$ are both increasing in a_1 and decreasing in a_2, and that $a_1^{obs} < \alpha_n$. 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, relative to $\frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}$, for $a_1 = a_1^{obs}$ and all a_2 (i.e., the incentives to increase a_2 are sufficiently strong),*
- *$M_{\alpha_i}(a_2) - M_{a_1^{obs}}(a_2) - \left(\frac{\partial c_S(\alpha_i, a_2)}{\partial a_2} - \frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}\right)$ is sufficiently negative relative to (1), for all i and all a_2,*

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

- *$M_{\alpha_i}(a_2) - M_{a_1^{obs}}(a_2) - \left(\frac{\partial c_S(\alpha_i, a_2)}{\partial a_2} - \frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}\right)$ is sufficiently positive relative to (1), for all i and all a_2,*

then receiver strictly prefers to keep a_2 observed.

[14] 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.

I emphasize that $\pi_R(a)$ and $\pi_S(a)$ can be computed directly from primitives, under the assumption that the receiver's payoff is strictly concave in e for all θ. With stronger assumptions on preferences and the experimentation technology, one could parameterize environments under which the conditions are jointly satisfied.[15] On the other hand, as the point of the proposition is to highlight the economic properties of the experimentation environment that can be used to make welfare predictions in particular applications, I avoid imposing any additional assumptions directly.

Within the general environment, Proposition 2 formalizes that the complementarity between experiment dimensions (i.e., how an increase in a_1 incentivizes the sender to change the choice of a_2) 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. The receiver-benefiting compensation is induced due to the indirect effect. The Proposition also highlights that the opposite conclusion is obtained if instead the incentive to increase a_2 were *stronger* following higher choice of a_1. Were Section 2 to instead posit that the more informative choice for a_1 were *more* susceptible to changes in a_2 (instead of independent of a_2), we would instead obtain a range of costs where $a_1 = 1$ were chosen only under full transparency. 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 conditions. The roles of most of these conditions are straightforward to see; for instance, the a_1 choice under transparency should be bounded away from edge cases to ensure scope for the direct and indirect effects to have impact. The other conditions on sender preferences ensure that biasing is in fact costly, and that higher experiments are not too costly. The discreteness in Proposition 2 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, Proposition 2 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.

[15] For instance, one could consider specifications where sender and receiver share information preferences, as in Section 2. Then, the conditions on information preferences would be satisfied if informativeness increases in a_1 and decreases in a_2, with $a_2 = \max A_2$ yielding completely uninformative experiments. Assuming $c_S(\alpha_1, a_2) = 0$ for all a_2 also makes the conditions on costs, as well as the bulletpoints, straightforward to satisfy; the conditions before the bulletpoint are satisfied if $c_S(\alpha_i, 0)$ are large enough so that the sender prefers α_1, but not so large that the sender would rather the experiment be uninformative; the first bulletpoint is satisfied whenever (1) is uniformly positive in a_2 at $a_1 = \alpha_1$; and the second bulletpoint becomes a condition on the marginal cost of a_2 for $a_1 \neq \alpha_1$.

3.5. Off-Path Beliefs[16]

The previous analysis makes use of the equilibrium choice of a_2 as a function of the observed choice of a_1. In particular, it takes as given that a_2 is correctly inferred by the receiver, and that this inference would be correct even if the sender were to instead choose some a_1 not part of the equilibrium profile. While Lemma 1 illustrates that an equilibrium exists where this is the case, this assumption matters for the conclusion tying the merits of transparency to complementarity.

Without this requirement, there may be other equilibria under limited transparency which outperform those outlined in Lemma 1. In fact, these other equilibria may outperform full transparency in cases where full transparency would otherwise be receiver-optimal. I provide an example of this in Appendix B.3, using an experimentation technology where more informative choices for a_1 yield experiments that are more susceptible to bias, yet limited transparency encourages the sender to increase a_1. This contradicts the conclusion of Proposition 2, highlighting the importance of correct off-path inferences.

I walk through the reason these improvements in receiver welfare may emerge. For simplicity, assume the marginal cost to higher a_2 is very large, so that the sender does not face any incentive to increase it if *forced* to choose any particular $a_1 \in A_1$. Suppose further that the sender picks $(a_1^{obs}, \underline{a}_2)$ under full transparency. However, consider a profile under limited transparency where instead the receiver conjectures $a_2 = \max A_2$ following any choice of a_1 other than $\tilde{a}_1$, where $\tilde{a}_1 > a_1^{obs}$. If the marginal cost to higher a_2 is large, these would not be equilibrium choices *given* a choice of $a_1 \neq \tilde{a}_1$. But if beliefs are not restricted off-path, then they would be valid if the receiver only observes $\tilde{a}_1$ on-path. Furthermore, if an inference of $a_2 = \max A_2$ makes the experiment sufficiently devalued, the sender may indeed be deterred from choosing $a_1 \neq \tilde{a}_1$, making this an equilibrium. If the receiver prefers experiments with higher a_1, then he is better off in this equilibrium (under limited transparency) than under full transparency.

This discussion illustrates that such off-path beliefs provide another channel through which limited transparency can improve payoffs. The use of an "incorrect" inference off-path is only possible when a_2 is unobserved.[17] Such beliefs can emerge in the general model—and in this case, studying whether limited transparency can benefit the receiver requires computing both the correct on-path inference, as well as the worst-case off-path inference. Interestingly, though the channel is different, limited transparency benefits the receiver for the same reason as in Section 2; a more informative a_1 avoids the perception of bias.

[16]I thank an anonymous referee who provided the key ideas in this Section.

[17]Note that, in a game where a_2 were chosen *after* a_1, this profile would not be an equilibrium in every subgame. This raises a well-known issue in games of incomplete information regarding off-path beliefs, which typically motivates a distinction between "weak PBE" and "PBE" (see, for instance, Mas-Colell, Whinston and Green (1995)).

4. POLICY IMPLICATIONS

This section studies the policy implications of the model. Proposition 1 and Lemma 3 summarize how observability of some dimension influences the equilibrium experiment choice. Here, I seek to calibrate each of these to the scientific research application. 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 may or may not be significant.

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. I elaborate on the preferences in Section 2. Suppose $\theta \in \{T, F\}$, and the receiver's utility function is (for $e \in [0, 1]$):

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

for some $b_R > 0$ and a continuously differentiable 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:

$$U^S(e, \theta, \hat{p}) = \lambda \overbrace{b_S \cdot e \cdot \mathbf{1}[\theta = T]}^{(*)} + (1 - \lambda) \overbrace{g(\hat{p})}^{(**)}. \tag{6}$$

I refer to $(*)$ as the *follow-on incentive* and $(**)$ as the *perception incentive*. The latter refers to benefits obtained in cases where they have a reputation for finding true hypotheses. Both are nested within the general model, and while the second term is reminiscent of the career-concerns literature, I am not aware of existing models that distinguish both incentives. Given that the relative importance of each may vary over a career, λ seeks to reflect differences in career stage.

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

Lemma 4. *Either of the following conditions imply the sender's follow-on incentive in (6) is convex in the receiver's belief:*

- $c_R''(e) > 0$ *and* $c_R'''(e) \leq 0$.
- $c_R(e) = e^n/k$ *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 4. But in the long-term, it is tempting add bias if a negative result decreases the influence of the researcher's future experiments.

I make two points distinguishing these two incentives. When $\lambda = 1$:

$$M_{a_1}(a_2) = b_S(e(\hat{p}(1)) - e(\hat{p}(0)))\frac{\partial \mathbb{P}[y = 1 \mid a_1, a_2, \theta = T]}{\partial a_2}\mathbb{P}[\theta],$$

but when $\lambda = 0$:

$$M_{a_1}(a_2) = (g(\hat{p}(1)) - g(\hat{p}(0)))\frac{\partial \mathbb{P}[y = 1 \mid a_1, a_2]}{\partial a_2}.$$

Note that 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_1 choices that are 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 true and false positives matters less.

Second, Lemma 4 shows that the incentive for follow-on research encourages information acquisition, whereas the perception incentive discourages information acquisition (provided $g''(\hat{p}) < 0$). This suggests the compensation effect highlighted is more relevant for more established researchers, as these researchers may be more insulated from skill perceptions. If perceptions themselves matter more, then the fact that experiments are per se perceived less informative does not lead to as much of a loss, 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 Aspects 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. The difference between the two examples is which aspect of the experiment this biasing concerns:

Example 3 (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 control for more factors 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 is $a_1 = h$.*

Example 4 (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 choice of a_1 corresponds to more informative experiments.*

I use Lemma 3 to articulate why a transparency requirement will have differing impacts for each of these kinds of research activities. For clarity, let us focus on the case where the *cost* of each activity does not differ depending on the underlying informativeness. This means it suffices to compare $M_{a_1}(a_2)$ across a_1.

Let us start with Example 3. 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 inclusion of more variables can lead to estimates that are less sensitive to ad-hoc discretionary choices.[18] Since the marginal impact on the results depends less on a_2 when h is chosen, it follows that $M_\ell(a_2) > M_h(a_2)$. Lemma 3 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 4. 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 more controls implies more possible associations that can be found, and any one interaction may yield a positive result.[19] While the experiment may be more informative by

[18] 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 non-core variables allows for easier and more precise hypothesis testing suggests that these designs endow researchers with greater commitment, consistent with the complementarity I impose in this example.

[19] That experiments with more explanatory controls may be more susceptible to "significance inflation" is also discussed in Vivalt (2018).

collecting more controls, it is also easier to bias. This contrasts with robustness tests, since the success of a particular robustness check may not be *responsible* for a positive result. 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 a rich assembly of controls is feasible, 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 (for instance, as in 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.

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.[20]

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.

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

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

scientific research. Arguments that biasing is solved endogenously are compelling if a substitution channel exists, provided consumers of research are sophisticated and update beliefs rationally. But the transparency requirements of Ioannidis et al. (2014) are attractive when these channels are absent, 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 directly to the relevant policy questions. Still, the model I have developed is fairly general, and thus it may shed light on other information acquisition settings as well. But to maintain focus, I have (until now) said much less on other applications.

I have studied an idealized setting where transparency requirements are immediate and costless to enforce. Although I view this idealized setting as an important benchmark, practically implementing them may introduce further complications not considered here. For instance, consider the specification in Section 1.1, Example 1, where a_2 is the number of times the experiment is repeated. One can imagine the transparency requirement as dictating how a relevant funding agency monitors how resources are used, making it possible to tell whether an experiment is repeated multiple times. In some instances, there may even be a legal requirement to register an experiment before undertaking it.[21] But in cases where funds come from other sources or are not significant, this might not work as a way of verifying the true number of times an experiment was run. In other cases, Institutional Review Board (IRB) approval may be required in order to conduct an experiment, with experimenters being bound to IRB statements. On the other hand, IRB is not required for every kind of study, and it is typically not public (see Ioannidis et al. (2014)). A policy to make all IRB proposals public may be one way of implementing transparency, but may have other consequences not considered here which make this undesirable.

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 a_2 is continuous. Even in cases where it is not (such as Section 2), one can consider a version of the model where the

[21] Abrams, Libgober and List (2020) discuss issues related to registration requirements of AEA journals, comparing them to registration requirements in other disciplines as well. Medical experiments often are legally required to register prior to being performed. On the other hand, while AEA journals require RCTs to be registered in order to be published, this can be done during submission (when of course the experimental outcome is known), making it difficult to characterize this policy as a transparency requirement. Abrams, Libgober and List (2020) note that many researchers do indeed register late. Based on its current design, this registration requirement cannot rule out an experimenter only submitting the successful design having tried many others. Importantly, this point is not a criticism of the policy, which may be desirable for other reasons.

choice of the sender along this dimension consists of a *probability* of choosing some a_2. 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. Hence Assumption 3 does not seem out of line for my main application. In any case, one could still distinguish the direct and indirect effects with more general Y outcomes, provided pure strategy existence could be maintained through other means.

Proposition 2 shows that limited transparency can be receiver-optimal when receiver and sender share similar preferences over information. While this seems strong, it applies to many well-studied specifications in the communication literature. Under quadratic preferences a la Crawford-Sobel,[22] 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 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 inconsistent with others; for instance, an intrinsic preference for positive results by researchers.[23] 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 (as described in Olken (2015)) appear to be the simplest explanation regarding why they are not used more thoroughly. Note that these costs *would* need

[22] That is, taking $U^S(e, \theta) = -(e - (\theta + b))^2, U^R(e, \theta) = -(e - \theta)^2$, for $\theta \in \Theta \subset [0, 1]$.

[23] 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.

to be taken into account in order to describe *sender* welfare across transparency 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 be random. The forces in this paper would still be present, weighted by the probability that the dimension is observable. That said, stochastic verification might suggest of other modifications worth exploring further.

5.2. Relation to Prior Work

The limited transparency benchmark corresponds to a version of costly persuasion with partial verifiability. The sender commits to a dimension of the experiment that is observed, but can manipulate a dimension that is 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)).[24] Costs have been introduced to both Bayesian Persuasion and cheap talk elsewhere,[25] as have various forms of intermediate commitment.[26] 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 if costless to do so.

The combination of costs and limited commitment to an information structure emerges in the literature on selective disclosure; see, for instance, Henry (2009), Felgenhauer and Schulte (2014), and Felgenhauer and Loerke (2017). In these papers, a sender conducts a number of experiments, and can withhold results that are unfavorable. While these papers use particular sender preferences and feature an information acquisition process similar to Example 1, they allow a richer signal space (since the receiver observes a *number* of successes). In particular, observing some number of successes reveals part of the information structure, something ruled out by the full support assumption. On the other hand, since all outcomes are disclosed when information acquisition activity is observed (via unravelling logic), the comparison between observed and unobserved information acquisition is similar to the full and limited transparency comparison in this paper.

[24] 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.

[25] See Gentzkow and Kamenica (2014) for Bayesian Persuasion; Kartik (2009), Argenziano, Severinov, and Squintani (2014) and Pei (2014) for cheap talk.

[26] See Lipnowski, Ravid and Shishkin (2018), Nguyen and Tan (2018), Guo and Shmaya (2017) or Min (2017).

Henry (2009) and Felgenhauer and Loerke (2017) both show that the sender may be induced to choose more informative experiments in the latter case, to counteract the receiver's inference that disclosure is selective. By explicitly modelling distortion using a separate dimension, my results allow for complementarity between distortion and informativeness to play a more explicit role. This allows me to speak directly to different kinds of distortions as in Section 4.2. My paper thus extends the insights of these papers to an alternative modelling approach.

Outside of the communication literature, 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 $a \neq (1, 0)$.[27] 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 on one dimension for a loss of credibility on the other drives the result.

Multitasking models have also been studied in principal-agent settings with moral hazard, following Holmström and Milgrom (1991). It is important to note, however, that in these settings, observable effort implements the first-best due to the ability to condition transfers on choices (as is typical in moral hazard models with transfers, and along the same lines as the discussion in Section 4.3). Partial transparency therefore cannot outperform full transparency as a result. More generally, the presence of transfers gives the principal richer ability to influence the incentives of the agent. Holmström and Milgrom (1991) show it may be preferable to not condition on a performance measure if doing so distorts attention from harder to measure but more important actions. While similar to the distortion in this paper, unobservability does not change the contracting environment in this paper, unlike under moral hazard.

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 states 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

[27] 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.

marginal incentives as more information about *the state* becomes available. Instead, I compare incentives under different assumptions regarding observability of the agent's *actions*.[28] Their results thus only speak to the direct effect.

Lastly, several papers study incentives in academic publication, cautioning against associating false positives with problems in scientific conduct. Glaeser (2006) studies the incentives behind false positives, arguing that eliminating them may be socially harmful. His focus is on the incentives to choose novel hypotheses with exogenous value. Lacetera and Zirulia (2008) develop a model of scientific research which involves both the choice of hypothesis as well as the possibility of fraud, and study how the incidence of fraud interacts with researcher incentives. 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.

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.

[28] 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.

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 some of the same directions that work following Kamenica and Gentzkow (2011) has proceeded. One could allow for richer private information from the sender or receiver. And allowing for multiple senders also seem important and realistic modifications the machinery of this paper could be adapted to. Second, one could analyze 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. This way of modelling limited verifiability may be relevant in other contracting settings where information acquisition is endogenous.

References

Eliot Abrams, Jonathan Libgober, and John List. How Can Research Registries Be Improved? An Examination of the AEA RCT Registry. Working paper, February 2020.

Ricardo Alonso and Odilon Câmara. Bayesian Persuasion with Heterogeneous Priors. Journal of Economic Theory, 165: 672-706, September 2016.

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

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

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

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

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

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.

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

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

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

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

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.

Mike Felgenhauer and Petra Loerke. Bayesian Persuasion with Private Experimentation. *International Economic Review*, 58(3):829–855, August 2017.

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

Andreu Mas-Colell, Michael Whinston and Jerry Green. Microeconomic Theory. Oxford University Press, New York, NY.

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

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.

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

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

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

Bengt Holmström and Paul Milgrom. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics and Organization*, 7:24–52, January 1991.

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.

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

Navin Kartik. Strategic Communication with Lying Costs. *Review of Economic Studies*, 76(4):1359–1395, October 2009.

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

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

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

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

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

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

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

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

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

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

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 (noting that the requirement in the Lemma holds immediately whenever $a = \tilde{a}$).

Lemma 5. *Suppose the receiver infers the sender's choice as $\tilde{a}$. Then if the sender chooses experiment a such that $\mathbb{P}[y \mid \tilde{a}, \theta] = 0$ implies $\mathbb{P}[y \mid a, \theta] = 0$, sender's payoffs can be written:*

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

Proof of Lemma 5. 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. Note that the distribution over y is a function of the true sender choice (i.e., a), and not the inferred sender choice (i.e., $\tilde{a}$). Further noting that we can restrict to y which occur under a with positive probability, and that $\mathbb{P}[y \mid a, \theta] > 0$ implies $\mathbb{P}[y \mid \tilde{a}, \theta] > 0$. We thus rewrite the expectation as over realizations of y and θ:

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

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

Essentially, the argument follows from noting that under the full support condition, it is possible to divide and multiply every term in the sum by $\mathbb{P}[y \mid \tilde{a}, \theta]$ and $\mathbb{P}_{\tilde{a}}[y]$, in order to move between the objective distribution over states and the distribution perceived by the receiver after observing 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. □

An alternative way of expressing sender's payoffs in the case where $\tilde{a} \neq a$ is:

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

which follows from noting that $\hat{p}_a(y)[\theta] \cdot \mathbb{P}[y \mid a] = \mathbb{P}[\theta, y \mid a] = \mathbb{P}[y \mid a, \theta]\mathbb{P}[\theta]$. In other words, to use the belief ratio to switch between "actual" and the "inferred." In the case where the sender and receiver disagree over the *prior* (as opposed to the experiment), Alonso and Câmara (2016) use belief ratios to rewrite the receiver's induced belief as a function of the sender's (their Proposition 1), and subsequently apply the belief based approach to the sender's value function. Note that their Proposition 1 is true regardless whether the experiment is sender-optimal, although their primary applications of this result relate to this case.

The next proof is stated assuming no restrictions on the number of indices. In this case, Assumptions 2 and 3 should be understood as applying to the vector of indices which may potentially be unobservable to the receiver.

Proof of Lemma 1. Suppose M is the index set of observable indices, and partition the sender's action into $a = (a_M, a_{-M})$. This proof shows 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. With these preliminaries in mind, the proof applies Kakutani's theorem to the sender's best reply correspondence. Define the function $\phi(a)$ as follows:

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

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 in a_{-M}, noting that the variables inside U^S do not respond to $\tilde{a}$. Take $a_n \to a$, and $b_n \in \phi(a_n)$ with $b_n \to b$. Write $a_n = (a_M, a^n_{-M})$.

Recall 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 now show that if $\hat{p}_n \to \hat{p}^*$, then $e(\hat{p}_n) \to e(\hat{p}^*)$ (that is, $e(\hat{p})$ is continuous in $\hat{p}$); since effort is chosen from 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, contradicting the definition of $e(\hat{p}_n)$.

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

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

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

$$\sum_{\theta}\left(\sum_{y} U^S(e_a(y),\hat{p}_a(y),\theta)(\mathbb{P}[y \mid \theta,(a_M,\delta)]-\mathbb{P}[y \mid \theta,(a_M,b_{-M})])\right)\mathbb{P}[\theta] > c(a_M,\delta)-c(a_M,b_{-M}).$$

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 b_n and a with a_n—that is, sufficiently close to the limit—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. To see this, note that A_{-M} is compact, and as we have argued, the objective is continuous in $\tilde{a}$. Since the set of maximizers of a continuous function on a compact set is itself a compact nonempty set, we have that $\phi(a)$ is compact.

To show that $\phi(a)$ is convex, we show that the objective is concave in $\tilde{a}$. Note that we can write each term inside the sum over θ in (††) as:

$$\left(U^S(e_a(0),\hat{p}_a(0),\theta)+(U^S(e_a(1),\hat{p}_a(1),\theta)-U^S(e_a(0),\hat{p}_a(0),\theta))\mathbb{P}[1 \mid \theta,a_M,\tilde{a}]\right)\mathbb{P}[\theta].$$

Note that $U^S(e_a(0),\hat{p}_a(0),\theta)$ is a constant in $\tilde{a}$, and by Assumption 3, since $U^S(e_a(1),\hat{p}_a(1),\theta) - U^S(e_a(0),\hat{p}_a(0),\theta) > 0$ and $\mathbb{P}[1 \mid \theta,a_M,\tilde{a}]$ is concave, we have this expression is concave in $\tilde{a}$ (after summing over states θ). Furthermore, by Assumption 2, c_S is convex in $\tilde{a}$, meaning that the objective in the definition of ϕ is concave in $\tilde{a}$.

So suppose that a', a'' are both in $\phi(a)$ (noting that these must differ only in the coordinates A_{-M}; so in other words, supposing both a' and a'' are maximizing choices of $\tilde{a}$). Since $\phi(a)$ is concave $\tilde{a}$, it follows 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.

To conclude that a PBE exists, note that the above result shows that we can write the inferred choice of a_{-M} as a function of a_M, i.e. $a_{-M}(a_M)$, given some selection. (When there are two dimensions, this selection is written $a_2(a_1)$) Since A_M (or A_1) is finite, it follows that the image of A_1 under the sender's utility is finite as well. Its maximizer is achieved at an element of A_1, and so sender maximizes payoff by choosing this element of a_1, yielding a PBE.

I now show the claim on mixed strategy equilibria. If there were multiple equilibria, 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 expression for the sender's benefit in terms of

$U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)$, we have:

$$\nabla_{\tilde{a}} c(a_M, \tilde{a}^i) = \sum_\theta \left(U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)\right) \mathbb{P}[\theta] \nabla_{\tilde{a}} \mathbb{P}[y = 1 \mid (a_M, \tilde{a}^i), \theta],$$

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

$$\alpha \cdot \nabla_{\tilde{a}}(c(a_M, \tilde{a}^2) - c(a_M, \tilde{a}^1)) =$$
$$\sum_\theta (U^S(e_a(1), \hat{p}_a(1), \theta) - U^S(e_a(0), \hat{p}_a(0), \theta)) \mathbb{P}[\theta] (\alpha \cdot \nabla_{\tilde{a}}(\mathbb{P}[y \mid (a_M, \tilde{a}^2), \theta] - \mathbb{P}[y \mid (a_M, \tilde{a}^1), \theta])).$$

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

$$\alpha \cdot (\nabla^2_{\tilde{a}} c(a_M, a_c) \cdot (\tilde{a}^2 - \tilde{a}^1)) =$$
$$\sum_\theta (U^S(e_{(a_M,a)}(1), \hat{p}_{(a_M,a)}(1), \theta) - U^S(e_{(a_M,a)}(0), \hat{p}_{(a_M,a)}(0), \theta)) \mathbb{P}[\theta] \alpha \cdot \nabla^2_{\tilde{a}} \mathbb{P}[y = 1 \mid (a_M, a_\theta), \theta] \cdot (\tilde{a}^2 - \tilde{a}^1).$$

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. I provide a proof showing that (1) being positive implies higher choices of a_2; showing that lower choices are implied when this expression is negative follows identical reasoning. First, note that the sender's payoff (in equilibrium) can be written:

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

The proof considers the first-order condition of this expression. Note that the first-order conditions (as an equality for interior a_2 and an inequality for $a_2 \in \{\min A_2, \max A_2\}$) must hold both under full transparency, as well as under partial transparency (as per the proof of Lemma 1). The difference in these expression is that when a_2 is unobservable, $\hat{p}_{\tilde{a}}$ does not not change as the sender changes a_2.

First consider limited transparency. Given a pure strategy PBE, the first order condition following a correct inference by the receiver is:

$$\frac{dc_S(a)}{da_2} = \sum_y \sum_\theta U^S(\hat{p}_a(y), \theta) \frac{d\mathbb{P}[y \mid a, \theta]}{da_2} \mathbb{P}[\theta].$$

When a_2 is observable, a term is added to the right hand side, which corresponds to the change in the receiver's belief about the state. The action is higher whenever this term is positive. Using that the beliefs are differentiable as a function of the action, chain rule gives us that the added term is:

$$\sum_y \sum_\theta \nabla_{\hat{p}} U^S(\hat{p}_a(y), \theta) \cdot \frac{d\hat{p}_a(y)[\cdot]}{da_2} \mathbb{P}[y \mid a, \theta] \mathbb{P}[\theta],$$

which is (1) (noting that the brackets reflect that $\hat{p}$ is a belief over many states).

Now, suppose the first order condition holds at an interior value of a_2, say a^*_{obs}. If (1) is positive, then:

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

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_2. It follows that for the first order conditions to hold, the resulting a_2 must be lower. Hence if (1) is positive, then keeping a_2 hidden lowers it, so that the choice of a_2 is higher under observability, as claimed. Similar reasoning reveals the same conclusion holds at boundary cases, although the change need not be strict since the first order conditions only hold as inequalities in these cases (with the direction of the inequality depending on which boundary is considered). □

Proof of Lemma 3. Denote by $a^*_2(a_1)$ the equilibrium response of a_2, fixing the choice of a_1. By Lemma 1, this is characterized by the first order condition:

$$\frac{\partial}{\partial a_2} c_S(a_1, a^*_2(a_1)) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_2} \mathbb{P}[y \mid (a_1, a^*_2(a_1)), \theta] \mathbb{P}[\theta], \tag{7}$$

with equality holding whenever a_2 is interior. Using that Y is binary and evaluating at $a_1 = \tilde{a}_1$, we rewrite this as:

$$\frac{\partial}{\partial a_2} c_S(\tilde{a}_1, a^*_2(\tilde{a}_1)) \leq \sum_\theta (U^S(\hat{p}(1), \theta) - U^S(\hat{p}(0), \theta)) \frac{\partial}{\partial a_2} \mathbb{P}[y = 1 \mid (\tilde{a}_1, a^*_2(\tilde{a}_1)), \theta] \mathbb{P}[\theta]. \tag{8}$$

First suppose that the first order condition defining $a^*_2(\tilde{a}_1)$ holds with equality. Then adding $\frac{\partial c_S(a^*_1, a_2)}{\partial a_2} - \frac{\partial c_S(\tilde{a}_1, a_2)}{\partial a_2} \leq M_{a^*_1}(a_2) - M_{\tilde{a}_1}(a_2)$ to both sides of (8) when it holds with equality yields:

$$\frac{\partial}{\partial a_2} c_S(a_1^*, a_2^*(a_1^*)) \leq \sum_\theta \sum_y U^S(\hat{p}(y), \theta) \frac{\partial}{\partial a_2} \mathbb{P}[y \mid (a_1^*, a_2^*(a_1^*)), \theta] \mathbb{P}[\theta], \tag{9}$$

Using the fact that $c(a_1, a_2)$ is convex in a_2, and that $\mathbb{P}[y = 1 \mid (a_1, a_2), \theta]$ is concave in a_2, it follows that given an inferred choice of a_2 must be higher when a_1^* is chosen, and strictly so when a_2 is interior and the inequality is strict. On the other hand, if the first order condition holds as a strict inequality, then a_2 is chosen as an edge case, and the same reasoning implies that a_2 could only increase as well. □

Proof of Proposition 2. The key observation is that, even though transparency may increase a_2 uniformly over all choices of a_1, the losses from this increase in a_2 is small relative to the benefit (to the receiver) from inducing a higher a_1. Therefore, I first argue that it suffices to show the following, given the conditions of the Proposition:

- The losses (to the sender) from keeping $a_1 = a_1^{obs}$ are large, and
- The losses (to both the sender and the receiver) due to higher a_2 are small when $a_1 > a_1^{obs}$.

To see that this suffices, 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, given any (realized or conjectured) a_2.

Since the receiver's payoff increases in a_1, holding *fixed* the choice of a_2, there exists a discrete increase in the receiver's payoffs, say α, when the sender chooses a higher a_1. Therefore, by continuity of the receiver's payoff function, we can find some ε such that an increase from a_1 to $\alpha_i > a_1^{obs}$ also delivers a higher payoff, whenever the increase in a_2 is no more than ε (since as $\varepsilon \to 0$ corresponds to the case where a_2 does not increase).

Using these arguments, the proposition follows from the following observations: First, when (1) is sufficiently large, the loss to the sender is large as well. And second, the change in a_2 is minimal when a_1 increases, provided $M_{\alpha_i}(a_2) - M_{a_1^{obs}}(a_2) - \left(\frac{\partial c_S(\alpha_i, a_2)}{\partial a_2} - \frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}\right)$ is large relative to (1).

Both of these claims follow immediately from the same arguments as in Proposition 1 and Lemma 3. Denoting the equilibrium response of a_2 given an (observable) choice of a_1 by $a_2^*(a_1)$, given (1) sufficiently positive relative to $\frac{\partial c_S(a_1^{obs}, a_2)}{\partial a_2}$, Proposition 1 shows that $a_2^*(a_1^{obs})$ approaches $\max A_2$. By assumption, this lowers sender's payoff relative to any other action, provided a_2 is chosen sufficiently close to a_2^{obs}. Indeed, Lemma 3 shows that a_2^* decreases when a_1 increases. But

in fact, inspecting the first order conditions, we see that if the second bulletpoint holds, then the first order condition will be satisfied at a value of a_2 no more than ε larger than a_2^{obs}.

Hence as discussed above, the conditions ensure that the change in payoffs due to the (potentially) higher a_2 is small relative to the increase in payoffs due to higher a_1. This proves the proposition. The reasoning for the converse case is identical and hence omitted. □

Proof of Lemma 4. I first use Lemma 2 to write the sender's payoff from $(*)$ as a function of the receiver's ex-post beliefs. Applying this Lemma shows that the Sender's payoff is proportional to $\hat{p}e(\hat{p})$. Thus, to prove Lemma 4, it suffices to show that $\hat{p}e(\hat{p})$ is convex. This is immediate in the case of polynomial effort costs, since then $e(\hat{p})$ is proportional to $\hat{p}^{1/(n-1)}$, so that $\hat{p} \cdot \hat{p}^{1/(n-1)}$ is convex. More generally, take the second derivative of $pe(p)$ and observe that it is equal to:

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

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

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

differentiating with respect to p to obtain:

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

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:

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

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

B. MISCELLANEOUS

B.1. Counterexample to Lemma 1 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 positives and negatives. The failure arises due to a failure of concavity in the objective stated in Lemma 1. This can be avoided to a certain extent by taking a transformation of the index; hence the point of this example is to show that the real technical issue arises when the "positive" and "negative" label depends on the experiment choice, which is not fixed by taking a monotone transformation of a_2. Note that in this example, the feasible experiment set is convex.

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). By symmetry, 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 the objective $-\frac{a}{4} + a^2(1 - 2\tilde{a}^2)$. 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. Thus, there is no pure strategy equilibrium where the sender chooses $\tilde{a} \geq \frac{1}{\sqrt{2}}$. On the other hand, if $\tilde{a}^2 < \frac{1}{2}$, then the second derivative of the objective is positive. Since this is a quadratic

function, the optimum is on the boundary of the choice set, i.e., either 0 or 1, for any choice of $\tilde{a}$. Hence the only choice of a less than $\frac{1}{\sqrt{2}}$ that could possibly be part of a pure strategy equilibrium would $a = 0$. However, $a = 1$ is a best response to $\tilde{a} = 0$, showing that there is no pure strategy equilibrium.

B.2. Preferences over y

I comment on a modification to the model where I allow for the sender to have preferences over y itself. For simplicity, I consider the case where the payoffs are separable, and the sender obtains an added benefit of $\lambda_y \cdot y$. from a positive result. In principle, 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 4). 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 $U^S(e, \hat{p}, y, \theta)$ 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, only on the realized experiment choice. Hence neither do payoffs if all that matters for the scientist is the probability of a positive result.

While this is theoretically immediate, it may seem surprising in the context of the application—the preference for positive results appears so widespread that it is tempting to think it is intrinsic. This paper takes the view that the benefit from a "positive" or "negative" result is endogenous and depends on the belief movement. If a preference for positive results emerges entirely because negative results are harder to publish, then this would suggest an *interaction* between having a positive result and the other payoff terms. While this would add a non-convexity in the sender's payoff as a function of the receiver's belief (due to the jump at the prior), provided it is small, the main conclusions of the paper should not change drastically (albeit with some additional

notation). On the other hand, it would make it more unwieldy 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 identify the source and nature of any intrinsic preference for positive results, this is left for future work.

B.3. Example Illustrating the Role of Off-path beliefs

This section presents an example illustrating the role of the "correct inference assumption." This example is a modified version of Section 2. Specifically, consider payoffs exactly as in Section 2, but suppose instead the signal distribution is as follows:

$$\begin{aligned}
&\mathbb{P}[y=1 \mid \theta=T, a_1=0, a_2=0] = 2/5, && \mathbb{P}[y=1 \mid \theta=F, a_1=0, a_2=0] = 0,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=0, a_2=1] = 1/2, && \mathbb{P}[y=1 \mid \theta=F, a_1=0, a_2=1] = 1/6,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=1, a_2=0] = 3/4, && \mathbb{P}[y=1 \mid \theta=F, a_1=1, a_2=0] = 0,\\
&\mathbb{P}[y=1 \mid \theta=T, a_1=1, a_2=1] = 7/8, && \mathbb{P}[y=1 \mid \theta=F, a_1=1, a_2=1] = 2/3.
\end{aligned}$$

Intuitively, relative to the previous, we now let the $a_1 = 1$ experiment be (severely) susceptible to bias. Take $c(a_1, a_2) = c \cdot a_1 + k \cdot a_2$. Note that $\pi_R(a_1, a_2)$ is exactly as before for $a \neq (1, 1)$. Recall that in Section 2, if $c \in (\pi_R(1,0) - \pi_R(0,0), \pi_R(1,0) - \pi_R(0,1))$, then the sender would choose $a = (0, 0)$ under full transparency, but would choose $a = (1, 0)$ in order to credibly show that $a_2 \neq 1$.

However, whereas in Section 2 the $a_1 = 1$ experiment is *resistant* to bias, here the $a_1 = 1$ experiment is *highly susceptible* to bias. This suggests complementarity goes in the opposite direction, and thus that limited transparency would favor the $a_1 = 0$ experiment instead.[29] Now, $c > \pi_R(1,0) - \pi_R(0,0)$ implies the sender would rather choose $a_1 = 0$ if a_2 were inferred equal to 0. Furthermore, if k is very large, then indeed $a_2(1) = a_2(0) = 0$, since the marginal benefit to a higher probability of $y = 1$ is fixed and finite. I conclude that the limited transparency and full transparency experiments coincide if $c > \pi_R(1,0) - \pi_R(0,0)$ and k is sufficiently large, under the belief refinement in Lemma 1.

[29] To see this, first note that $\pi_R(1,1) < \pi_R(0,1)$, and $\pi_R(1,0) - \pi_R(1,1) > \pi_R(0,0) - \pi_R(0,1)$. In particular, one can calculate that $\pi_R(1,1) = \frac{29}{414} \approx .07 < \frac{1}{12}$. We have $\pi_R(1,0) - \pi_R(1,1) = \frac{5}{26} - \frac{29}{414} \approx 0.12$, and $\pi_R(0,0) - \pi_R(0,1) = \frac{1}{8} - \frac{1}{12} \approx 0.04$. As a result, for $k = 0$ and $c \in (\pi_R(1,1) - \pi_R(0,1), \pi_R(1,0) - \pi_R(0,0))$, the sender chooses $a_1 = 1$ under full transparency but $a_1 = 0$ under partial transparency.

Now take k large and $c \in (\pi_R(1,0) - \pi_R(0,0), \pi_R(1,0) - \pi_R(0,1))$. With different off-path beliefs, limited transparency can favor $a_1 = 1$, despite the opposite complementarity. Consider the following profile under limited transparency:

- Sender chooses $(1,0)$
- Following an observation of $a_1 = 1$, the receiver infers $a_2 = 0$. Following an observation of $a_1 = 0$, the receiver infers $a_2 = 1$.

If beliefs are not restricted off-path, then the sender would not have any incentive to deviate from $(1,0)$ given the receiver's inference following an observation of $a_1 = 0$. And while this inference is inconsistent with an action profile where $a_1 = 0$ is exogenously imposed on the sender, the stated belief profile is valid if off-path beliefs are not restricted. I thus conclude that, despite the opposite complementarity from Section 2, again the more informative experiment is chosen in this equilibrium under limited transparency.