Information Design with Constraints*

Jenny S. Kim†

Abstract

Political attacks are the results of strategic decisions by actors who choose how much to learn about the issue at hand and whether to politicize it. How does such endogeneity of information collection affect the frequency of political attacks we observe? Would limiting a party's ability to learn information reduce or increase the attacks? I present a formal model in which two parties with opposite preferences about an outcome decide whether to investigate each other and reveal the truth or remain silent. They receive a public signal before making the decision, and one party—the designer—can strategically choose how accurate this public signal is. Importantly, I compare two versions of the model where (1) the designer can choose any signal and (2) the designer is constrained to choose a binary and symmetric signal. I show that imposing a constraint on the designer's choice of information leads to three differences. First, the constraint generates a strong "collusive obfuscation" where both parties stay silent in equilibrium. Second, it limits the upside to the designer and the downside to the receiver; in particular, the receiver may prefer to have a constrained information designer over receiving perfect information about the state of the world, while he never prefers an unconstrained designer over it. Lastly, parties are willing to give up their initial share of the pie only when the information designer is constrained.

*I thank John Patty, Maggie Penn, Jessica Sun, Teddy Kim, Carlo Prato, Ryan Hubert, Barton Lee, Peter Buisseret, Lauren Futter, Allison Stashko, Luis Martinez, Alex Hirsch, and Sangmin Lee, as well as the audience members at APSA 2023 Annual Meetings in LA, California, Aspen 2024 Conference on Political Economy, and Theory & Models Working Group at Emory University for very helpful comments on this project. All errors are my own.

†Departments of Political Science, Emory University. Email: *seoyeon.kim@emory.edu*.

1 Introduction

Political attacks are central to understanding many of the dynamics in today's politics. They shape how voters view a political event, and affect voters' attitudes toward the political party involved (Lee, 2018; von Sikorski, Heiss and Matthes, 2020; Wolsky, 2022), or more generally toward institutions and the political process (Bowler and Karp, 2004; Thompson, 2013). However, attacks do not occur automatically after a politician's misconduct or a bad policy; instead, they are the results of strategic decisions by the actors who choose how much to learn about the issue at hand and whether to politicize it (Tiffen, 1999; Entman, 2012; Sberna and Vannucci, 2013; Invernizzi, 2016; Sowers, Nelson et al., 2016; Nyhan, 2017; Dziuda and Howell, 2021).

Consider a party's decision problem of whether or not to gather information about a potentially controversial issue. The party would ideally want to perfectly and privately learn the truth so that she can attack the other party when the outcome is favorable to her and remain silent otherwise. However, a lot of the time two features of information render it difficult for a party to choose how much to learn. First, information may be public: something that a party learns can also be accessible to the members of the other party, which creates incentives for the party to not learn everything about the issue. Second, a party may have limited ability to control information.

The main contribution of this paper is to examine an information designer's choice under these two settings and show how constraints in the designer's ability to choose public information matter. My theoretical framework compares the optimal choice of an information designer with and without constraints. Without constraints, the designer is endowed with full flexibility to choose any public signal; with constraints, her choice of signal is restricted to be binary and symmetric. I find that constraints in information design explain some less intuitive phenomena in politics.

In November 2022, the cryptocurrency exchange FTX collapsed following a liquidity crisis and a failed bailout. The bankruptcy of one of the largest crypto exchanges has drawn attention to the lack of laws and regulations governing digital tokens and exchanges. While there have been several congressional hearings probing the failure of regulators to oversee the crypto companies and provide clear guidance, both parties have refrained from making a scandal out of FTX's heavy lobbying to the legislators. Such mutual silence was possible because the misconduct crossed ideology. The former CEO of FTX, Sam Bankman-Fried, has hosted meetings with members of both parties and donated huge sums of money to political candidates, including members of party leadership or members in committees with jurisdiction over cryptocurrency policy. Recent indictments further accuse Bankman-Fried of having directed at least 300 illegal campaign donations to both Democrats and Republicans made in the names of others. The fact that both parties shared responsibility for a potentially scandalous situation was the main reason behind such mutual silence. Attempting to learn more about this scandal could easily incriminate members of one's own party. Thus there was no partisan advantage to publicizing the scandal; each political party knew that its own side was vulnerable to scrutiny, so neither party probed too deeply and had a bipartisan interest in silence.

This paper uses a theoretical framework to explain how the above "mutual silence under shared responsibility" dynamic can only be observed when we *constrain* the information designer's choice of signal. In the model, one party is facing a scandal. Two parties want different outcome realizations for a party's scandal; the party facing the scandal wants her to be innocent, while the other party wants her to be guilty of the issue in question. First, parties observe their initial payoffs from the scandal, which may be represented by public perceptions or media portrayal of the scandal. Then, they receive a public signal before they decide whether to stay silent on the issue or attack the other party. Importantly, in my model, the party facing the scandal—the designer—can strategically choose how accurate this public signal is. I introduce and compare two versions of the model where (1) the designer has full flexibility in her choice of information as in the recent literature on Bayesian persuasion and information design and (2) she is constrained to choose a binary and symmetric signal. I further examine whether parties have incentives to bargain over their initial status quo division for each of these two cases.

Some of the central results of the article are as follows. First, I show that the constraint on the designer's design of information generates a strong "collusive obfuscation" where both parties stay silent and never attack each other in equilibrium. A designer with no constraint chooses a perfectly uninformative signal and induces mutual silence only when the circumstances are already highly in favor of the designer that she has no incentives to manipulate information. However, by reducing her room to maneuver, the designer frequently chooses a completely uninformative signal that leads to collusion. Now such equilibrium always occurs for any status quo division as long as the cost of attacking is non-zero, which has critical implications for democracy. Even in adversarial contexts where a party wins for the detriment of another, scandals may fail to materialize and parties may intentionally stay silent about the issue.

Second, the constraint limits the upside to the designer as well as the downside to the receiver. An unconstrained information design always maximizes the designer's payoff while it minimizes the receiver's payoff. With constraints, the designer is unable to fully extract the informational benefit, and thus the receiver is always weakly better off than with an unconstrained designer. As will be detailed later, despite this being a very adversarial context, the receiver may also prefer the constrained designer's choice of signal over receiving perfect information about the state of the world.

Third, parties may be willing to give up their initial pie only under constrained information design. Parties sometimes prefer to have less initial share of the division, as an unfair status quo division can lead to the designer optimally choosing some signal inaccuracy that leads to worse payoffs for the parties. Such preferences for a lower share of the pie allow parties to sometimes agree on a division, and the new division they agree on always leads to collusion in equilibrium.

This model is largely related to the literature that highlights how information provision can be endogenous to the actions of the decision-maker (Gailmard and Patty, 2017; Bils, Carroll and Rothenberg, 2020; Libgober, 2020; Prato and Turner, 2022), and more specifically how political actors strategically shape information available to another for policy influence (Austen-Smith, 1993; Potters and Van Winden, 1992; Schnakenberg, 2017; Schnakenberg and Turner, 2019, 2021, 2023; Awad, 2020; Ellis and Groll, 2020; Patty and Penn, 2022; Dellis, 2023; Patty and Penn, 2023). For instance, Patty and Penn (2022) find that the employer might optimally withhold sensitive information that can be used for discrimination. Similarly, in Patty and Penn (2023), the designer might want to incorporate noise into his decision in order to manipulate behavior. These results are equivalent to the designer in my model committing to receiving a noisy, inaccurate signal.

Studies have interpreted information accuracy as clarity of communication (Dewan and Myatt, 2008), transmission of information by an exogenous outside agency (Gal-Or, 1985), information technologies (Ellis and Groll, 2020), imperfect private monitoring of an investigation outcome (Beshkar and Park, 2016), or degree of transparency (Stasavage, 2003, 2007; Gavazza and Lizzeri, 2007; Gailmard and Patty, 2012). I adopt the last interpretation and study the strategic incentives behind the designer's optimal choice of signal structure.[1] In particular, my model builds on previous works that have identified possible costs of transparency, where more information may not necessarily benefit the political actors (Stasavage, 2003, 2007; Gavazza and Lizzeri, 2007; Meade and Stasavage, 2008; Meijer, 2013; Berliner, 2014; Turner, 2017, 2021; Gradwohl and Feddersen, 2018).

My theoretical framework contributes to this literature in two ways. First, I look at an extremely adversarial context where parties have completely opposite preferences about the outcome. I show that transparency can be harmful to both parties even in this setting, and more importantly that both may prefer to have one party choose the signal accuracy than to receive perfect information.[2] Second, transparency in my framework is not centered around the accountability problem - there is no *ex-post* monitoring of an agent's action. Instead, the designer strategically chooses the accuracy of a public signal that informs both herself and her opponent. She faces a direct trade-off where her attempt to learn the true outcome immediately leads to her opponent also acquiring more information; the designer optimally chooses a less informative signal despite this increasing her own chances of making a mistake.

Methodologically, I build on the foundations laid by a growing literature on Bayesian persuasion and information design. Since the seminal work of Kamenica and Gentzkow (2011) that set up the Bayesian persuasion problem that studies the game of strategic communication between a sender and a receiver, the framework has inspired an active line of research

[1] This project is different from the cheap talk literature in the sense that parties do not have private information; both parties receive a public signal before making a decision. The designer's role is therefore not to engage in a cheap talk but simply to decide how accurate that public signal would be.

[2] Note that parties in my model pay the cost of attacking c when they decide to engage in a conflict, but I show that for $c > 0$ that is arbitrarily close to 0 there still exists an equilibrium where both parties prefer some uncertainty over full transparency. This is not due to risk aversion, as parties in my model are risk-neutral with a linear payoff function.

(Gehlbach and Sonin, 2014; Alonso and Câmara, 2016; Dughmi, 2017; Cotton and Li, 2018; Kamenica, 2019; Bergemann and Morris, 2019; Luo and Rozenas, 2018, 2022; Gerardi, Grillo and Monzón, 2022; Tomasi, 2022). Applying this method, I solve for the designer's optimal experiment and compare it with the baseline case where her choice of signal structure is limited.

2 The Model

I consider a model where two parties (Party 1 and 2), given a status quo division $(x, 1-x)$ where $x \in [0,1]$, can decide whether or not to attack each other and reveal the truth about Party 1's scandal. Parties receive a public signal before they decide whether or not to attack. Importantly, Party 1 (she) has control over the signal accuracy.[3] We could expect two counteracting incentives to prevail: one where Party 1 wants an accurate signal so that she could make an informed decision, and another where she wants Party 2 (he) to be sufficiently unaware of the true state of the world so that he either 1) doesn't attack even when the state is in favor of him or 2) mistakenly attacks when it is disadvantageous for him. I model this trade-off to identify conditions under which some degree of inaccuracy can be strategically optimal. Each stage of the game is detailed below.

Initial Stage. Prior to the start of the game, Nature draws the outcome of a political scandal $\omega \in \{0,1\}$. This state is not directly observed by any party. Parties have opposite preferences towards state ω; Party 1 prefers the outcome to be advantageous to her ($\omega = 1$), while Party 2 prefers otherwise ($\omega = 0$). Nature draws Party 1's favorable outcome $\omega = 1$ with probability $\pi \in (0,1)$. Probability π thus represents how likely Party 1 will survive the scandal, which can be measured by how many names from each party were involved with the political scandal, or the outcome of past political accusations. Parties then observe the status quo division $(x, 1-x)$ and the cost of attacking $c \in [0,1]$, which is common to both parties. This division can be interpreted as the parties' current payoff from the scandal, such

[3] More generally, and similar to the approach of Herrera, Reuben and Ting (2017), one can also envision that at some time in the future, the second party becomes the designer, and so forth in alteration for future periods. The one-shot asymmetric game equilibrium described here remains an outcome of this more complex repeated interaction.

as the public perception or media portrayal of the issue.

Signal Stage. Then, Party 1 chooses the structure of a public signal. Let $\mu \in [0,1]$ denote the parties' common posterior belief that $\omega = 1$. Formally, the signal structure is defined by a distribution F over posteriors μ subject to Bayesian plausibility, i.e., $\int \mu dF(\mu) = \pi$.

Conflict Stage. Lastly, parties simultaneously decide whether or not to learn the true state of the world based on the public signal they have received in the signal stage.[4] In this stage, they choose either to stay silent (S) or attack (A). If both parties choose to stay silent, they receive the initial status quo payoffs. If either of the parties chooses to attack, the true state of the world is revealed and the parties' realized payoffs are $(\omega, 1-\omega)$. Intuitively, Party 2 attacks if the outlook of the scandal after the signal is disadvantageous enough for Party 1; Party 1 may also attack if she is fairly confident about the outcome of the scandal and believes that she will be able to overturn the status quo. Further, a party choosing to attack pays a net cost of attacking c. This cost measures the general effort needed to reveal the new state of the world. Additionally [5] High cost implies that parties won't be willing to overturn the initial status quo division unless the additional benefit they expect to gain by attacking exceeds its cost. The sequence of the game is summarized below.

1. Nature draws a true state $\omega \in \{0,1\}$, where

$$\omega = \begin{cases} 0 & \text{w.p. } 1-\pi \\ 1 & \text{w.p. } \pi. \end{cases}$$

2. Parties observe the cost of attacking c and the initial division of dollar $(x, 1-x)$.

3. Party 1 chooses signal structure F.

4. Parties play a simultaneous game.

[4] Note that the results of the model are identical even if we assume that the parties move in sequential order. In such a setup, Party 2 moves first and decides whether to attack or stay silent. After Party 2's action, Party 1 then also chooses between the same actions. I assume throughout for simplicity that parties act simultaneously.

[5]

		Party 2: S	Party 2: A
Party 1	S	$(x, 1-x)$	$(\omega, 1-\omega-c)$
Party 1	A	$(\omega-c, 1-\omega)$	$(\omega-c, 1-\omega-c)$

Table 1: Payoff Structure

3 Equilibrium Analysis

I now proceed to characterize and describe the equilibrium behaviors of the parties. The solution concept for this game is Perfect Bayesian equilibrium. I analyze this game by backward induction.

3.1 Model 1: Information Design *without* Constraints

First, consider a case where Party 1 chooses any amount of signal. This is equivalent to Party 1 choosing any distribution F that satisfies Bayesian plausibility (Kamenica and Gentzkow, 2011). I thus refer to the distribution F as the signal (experiment) of Party 1.

Decision to attack. Given Party 1's choice of the optimal signal, Party 1 (2) attacks when the status quo is smaller (larger) than the expected utility of attacking. Let $U(\mu)$ denote Party 1's indirect utility when she induces posterior $\mu \in [0,1]$. Then,

$$U(\mu) = \begin{cases} \mu & \text{if } \mu < x - c \\ x & \text{if } \mu \in [x-c, x+c] \\ \mu - c & \text{if } \mu > x + c. \end{cases}$$

It follows that Party 2 attacks when μ is sufficiently small. Both parties remain silent for moderate values of μ. Lastly, Party 1 attacks when μ is sufficiently large.

Party 1's optimal signal. Given $U(\mu)$, if Party 1 chooses a distribution F, then her expected payoff is given by $\mathbb{E}[U(\mu)] = \int U(\mu) dF(\mu)$. This implies that Party 1's optimal persuasion problem can be written as

$$\max_{F \in \Delta([0,1])} \int U(\mu) dF(\mu) \text{ subject to } \int \mu dF(\mu) = \pi,$$

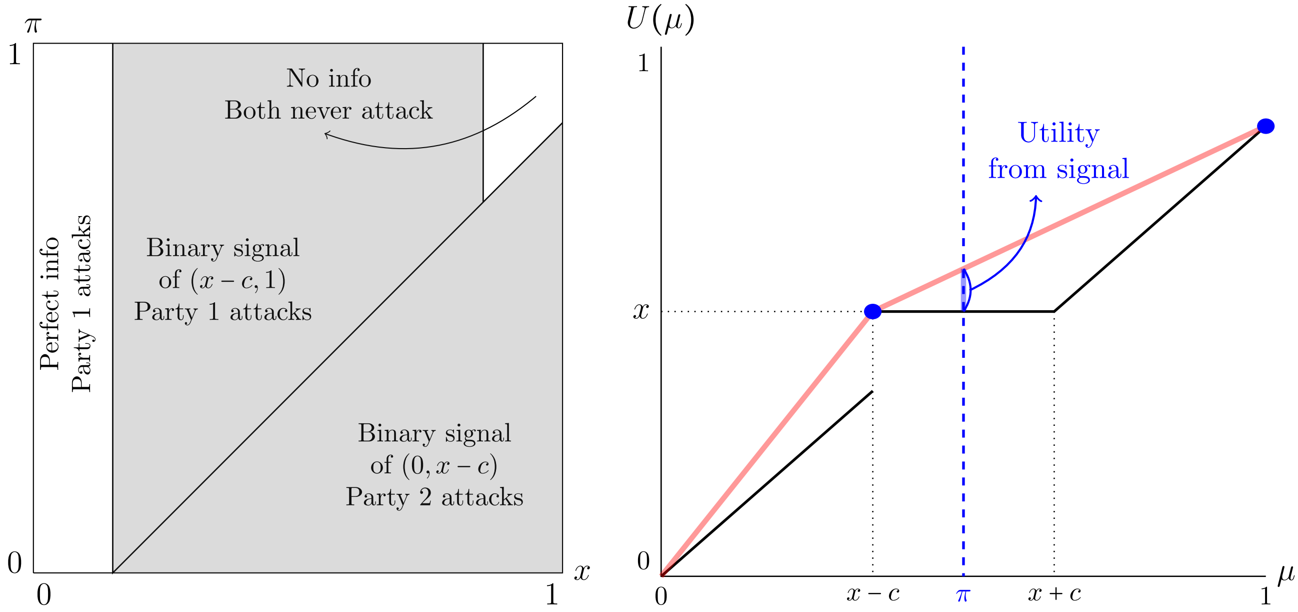
where $\Delta([0,1])$ denotes the set of all distributions over $[0,1]$. As observed by Kamenica and Gentzkow (2011), this problem can be solved by the method of concavification. Specifically, I look for the concave closure $\phi(\mu)$ which is equivalent to the smallest concave function above $U(\mu)$. Then, any distribution F is an optimal strategy if it satisfies Bayesian plausibility, the support of F is a subset of $\{\mu : U(\mu) = \phi(\mu)\}$, and $\int U(\mu)dF(\mu) = \phi(\pi))$. Applying this method to Party 1's problem leads to the following choice in equilibrium.

Proposition 1 *Consider Party 1's optimal choice of signal in equilibrium.*

- *Perfect information is optimal if* $x < \min\{c, 1-c\}$.
- *Any information is optimal if* $1 - c < x < c$.
- *No information is optimal if* $\max\{c, 1-c\} < x < c + \pi$.
- *A binary signal that leads to posterior* $x - c$ *or* 1 *is optimal if* $c < x < \min\{1-c, c+\pi\}$.
- *A binary signal that leads to posterior* 0 *or* $x - c$ *is optimal if* $x > c + \pi$.

In Figure 1a, I visualize the equilibrium outcomes given $c = 0.15$ and examine how they vary with respect to x and π. We can see that depending on these parameters, Party 1 optimally chooses varying levels of signal accuracy. When x is small, i.e., when the status quo is extremely unfavorable towards Party 1, Party 1 chooses perfect accuracy and conflict occurs in equilibrium only through Party 1's attack when the signal is 1. Her optimal choice is to perfectly learn about the state of the world despite the fact that Party 2 learns it as well. Here the additional benefit from accurately learning the state of the world is much larger than the risk of being attacked. Conversely, as x and π increase, Party 1 optimally chooses no accuracy or some binary experiment in equilibrium.

I examine the binary experiment in detail. Figure 1b illustrates Party 1's optimal choice of experiment given $x = 0.6$ and $\pi = 0.3$. Black solid lines are Party 1's indirect utility $U(\mu)$



(a) Region Plot

(b) Party 1's Indirect Utility ($x = 0.5, \pi = 0.5$)

Figure 1: Unconstrained Information
$c = 0.15$

as a function of the posterior μ. When $\mu < x - c$, the posterior probability that $\omega = 1$ is low enough that Party 2 attacks in equilibrium. With $\mu \in [x - c, x + c]$, the parties are uncertain enough about the state of the world that they stay silent; with high $\mu > x + c$, Party 1 attacks in equilibrium.

The red translucent lines represent the smallest concave function above Party 1's indirect utility function $U(\mu)$, which I denote $\phi(\mu)$. The lines imply that the optimal experiment is a binary experiment that mixes between posterior 0 and $x - c$. Note that $\mu = x - c$ is the smallest value among μ where Party 2 is unwilling to fight after he gets a signal of 1. Party 1 wants to maximize this probability while satisfying the Bayesian plausibility, which results in a binary experiment that mixes between 0 and $x - c$. If the induced posterior is 0, Party 2 attacks and is certain to win; if it is $x - c$ both players stay silent. Thus only Party 2 attacks after his favorable signal in equilibrium. Recall that the mixing probabilities of signals must satisfy Bayesian plausibility. Party 1's payoff from implementing the optimal experiment is thus $\phi(\pi)$. It follows that the distance between the red translucent line and the black solid line at point $\mu = \pi$, i.e. $\phi(\pi) - U(\pi)$, is Party 1's additional **utility from signal**.

3.2 Model 2: Information Design *with* Constraints

Importantly, a designer may not always have complete discretion over information. Information design without constraint assumes that Party 1 has full authority to under-report unsuccessful outcomes that incriminate Party 1 and over-report successful ones that exculpate the party, differentiating the false-positive and false-negative rates induced by a given accuracy. In many cases, however, a party only has control over how accurate the public signal—whether it is incriminating or exculpatory—will be. Party 1 may have overall discretion over how accurately a third party agency produces signal overall, for example, but may be unable to set different levels of accuracy conditional on the true outcome, Below I consider the case of a constrained information designer who is constrained to choose a (1) binary and (2) symmetric signal.

Let $\sigma \in \{0,1\}$ denote the binary signal. Let $\mu(\sigma)$ denote the posterior belief of probability that $\omega = 1$ after observing signal σ. Without loss of generality, assume that $\mu(0) < \mu(1)$. Conceptually, we have

$$Pr(\omega = 1|\sigma = 0) = \mu(0)$$
$$Pr(\omega = 1|\sigma = 1) = \mu(1),$$

where $Pr(\sigma = \omega|\omega)$ is the signal **accuracy**. With the symmetry constraint, Party 1 has to choose a signal where $Pr(\sigma = \omega|\omega = 0) = Pr(\sigma = \omega|\omega = 1)$. Bayesian plausibility as well as Bayes' rule requires the following equality to be satisfied:

$$\frac{\pi}{1-\pi} = \frac{\mu(1)(\pi - \mu(0))}{(1-\mu(0))(\mu(1)-\pi)}.$$

Since π is exogenous, LHS is constant. RHS is decreasing in $\mu(0)$ and $\mu(1)$ respectively. Therefore, the constraint reduces the designer's power by making the choice of $\mu(0)$ depend on $\mu(1)$ and vice versa.

Remark 1 Constraining the signal structure to be binary and symmetric restricts the pairs of achievable posteriors in a way such that if one signal becomes more informative, the other one must be as well.

It is straightforward to show that perfect and no information satisfy Equation (1). Thus, the equilibrium regions where perfect information and no information were optimal in the model without constraints also follow through in this version. I solve equilibrium outcomes only for regions where a binary experiment was optimal in the unconstrained model; below are Party 1's optimal choices of information.

Proposition 2 *Consider Party 1's optimal choice of signal under constraints.*

- *Perfect information is optimal if* $x < \min\{c, 1-c\}$.
- *Any information is optimal if* $1 - c < x < c$.
- *No information is optimal if*

$$\max\{c, 1-c\} < x < c + \pi \quad \textit{or} \quad \left(c < x < \min\{1-c, c+\pi\} \quad \textit{and} \quad \frac{\pi^2(1-x+c)}{\pi^2 + (x-c)(1-2\pi)} < x + c \right).$$

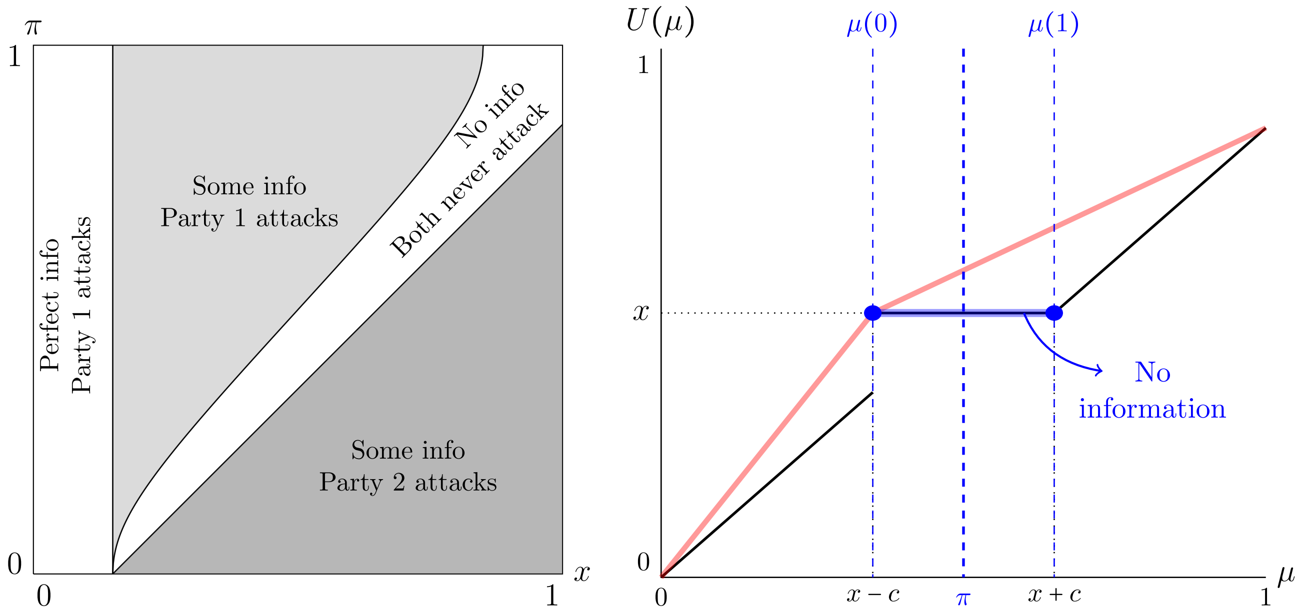
- *A binary signal that induces* $\mu^*(0) = x - c$ *is optimal if*

$$c < x < \min\{1-c, c+\pi\} \quad \textit{and} \quad \frac{\pi^2(1-x+c)}{\pi^2 + (x-c)(1-2\pi)} > x + c.$$

- *A binary signal that induces* $\mu^*(1) = x - c$ *is optimal if* $x > c + \pi$.

Note that Party 1 attacks conditional on receiving signal 1 after Party 1 chooses a binary signal that induces $\mu^*(0) = x - c$ and Party 2 attacks conditional on receiving signal 0 after she chooses a binary signal that induces $\mu^*(1) = x - c$. There are a total of three equilibrium strategies in the conflict stage: Party 1 attacks conditional on signal 1; Party 2 attacks conditional on signal 0; and both parties never attack.

In Figure 2a, I visualize the equilibrium outcomes given $c = 0.15$ and examine how they vary with respect to x and π. We can see that depending on these parameters, there are cases where Party 1 chooses perfect information, no information, and *some* information. When x is small, i.e., when the status quo is extremely unfavorable towards Party 1, Party 1 chooses perfect accuracy and conflict occurs in equilibrium only through Party 1's attack when the signal is 1. Her optimal choice is to perfectly learn about the state of the world



(a) Region Plot

(b) Party 1's Indirect Utility ($x = 0.5, \pi = 0.5$)

Figure 2: Unconstrained Information
$c = 0.15$

despite the fact that Party 2 learns it as well. Here the additional benefit from accurately learning the state of the world is much larger than the risk of being attacked. Conversely, as x and π increase, there exists an equilibrium region where Party 1 chooses no information. The probability that Party 1 gets out ahead of a scandal π is low enough that Party 2 is willing to attack given a favorable signal, but not too low such that Party 1's choice of a completely uninformative signal deters Party 2 from attacking. Both parties stay silent for all signals in this equilibrium. Finally, the two shaded regions are the outcomes where Party 1 strategically chooses some moderate level of accuracy, which I explain in detail below.[6]

Figure 2b illustrates the constraint that Party 1 is facing given $x = 0.5$ and $\pi = 0.5$. Black lines denote Party 1's indirect utility as a function of posterior μ, and the red translucent lines represent the smallest concave function above Party 1's indirect utility function $\phi(\pi)$. Note that an unconstrained designer can achieve her maximal payoff, so Party 1's equilibrium

[6]While not represented in Figure 2a, when x is moderate and c is sufficiently large, Party 1 is indifferent among any signal, and in equilibrium, both parties stay silent in the conflict stage. Moderate x means that the status quo is not largely in favor of either party. This, coupled with the high cost of attacking, discourages both parties from finding out the state of the world regardless of the public signal. Thus Party 1 chooses any information in this region.

payoff in Model 1 given $\pi = 0.5$ is simply $\phi(0.5)$. Such payoff cannot be induced with a symmetric signal; Party 1's optimal choice under constraints is instead to choose a completely uninformative signal that provides no additional information about the state of the world. Now Party 1's equilibrium payoff in this version is x because both parties remain silent in equilibrium. The conditions for such **collusion equilibrium** is in Proposition 2, which refers to the *ex-post* outcome where Party 1 chooses no information and neither party chooses to attack. Some degree of equality is necessary for such silence. Parties stay silent when the prior probability that Party 1 is innocent is not too unfair; they are satisfied enough that they do not wish to take risks and initiate a conflict. Note that for any π there always exists a range of x that results in this equilibrium outcome as long as c is positive, which is surprising given the adversarial nature of this model. It is important to think about the political consequences of this equilibrium; the collusion equilibrium tells us that in this region, both parties may decide to avoid revealing the truth and collude to hide a scandal.

4 How do constraints matter?

In this section, I examine how and why constraints matter. Comparing two models with and without informational constraints uncovers three main differences that I discuss below.

4.1 Collusive Obfuscation

The most crucial difference between the two models lies in the set of conditions that sustains the collusion equilibrium. Less power to design information creates a strong collusive obfuscation in equilibrium where Party 1 sets a completely uninformative signal and both parties never attack.

We can clearly observe such a distinction in Figure 3, which compares Figure 1a and Figure 2a. Importantly, the region where the equilibrium probability of attack is non-zero in Model 1 but is zero in Model 2 is shaded in red. Collusion in the unconstrained model (Model 1) is only possible when both x and π are significantly large. In other words, when Party 1 has full flexibility over the choice of information, she chooses to completely obfuscate

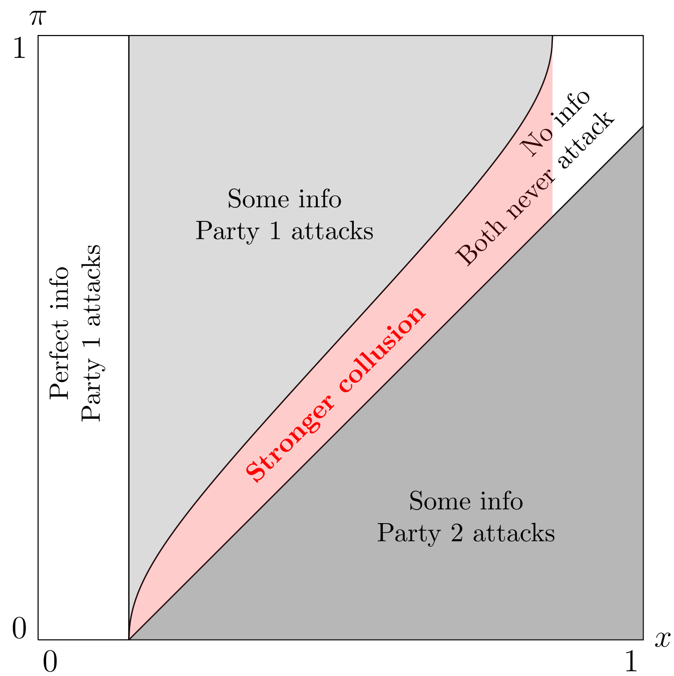


Figure 3: Preferences over the Status Quo Division
$\pi = 0.3$

the signal when the circumstances are already favorable to Party 1: her status quo payoff x is high, and the probability that she is innocent π is also high that she has no incentives to manipulate information. Otherwise, an unconstrained Party 1 can carefully tailor the signal so that she can learn enough to attack while still being able to deter the other party from attacking, and silence disappears.

A constrained Party 1, on the other hand, is faced with a dilemma where making one signal more informative means that the other one must be as well, and vice versa. This means that an attempt to minimize the probability of Party 2 attacking conditional on signal 1 might lead to Party 1 providing *too much* information to Party 2, which induces him to attack conditional on 0. In this case, her best choice among the restricted pairs of achievable posteriors is to provide no information about the state of the world. From an inferential standpoint, when the values of x and π are sufficiently small or moderate, we are able to observe the collusion equilibrium when the designer has limited power over information.

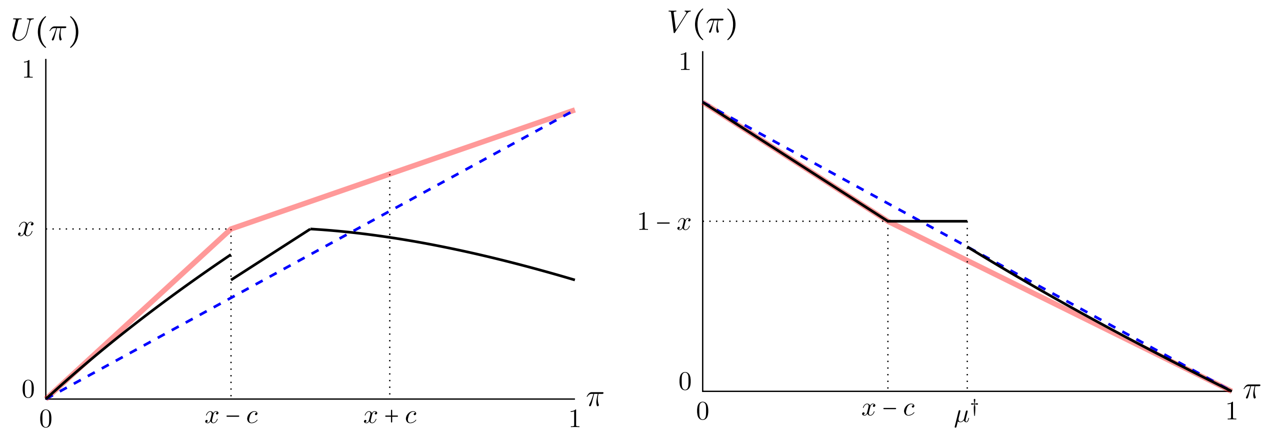
Proposition 3 *Suppose $c < x < 1-c$. There exists a region where collusion occurs only when*

Party 1 is restricted in her choice of information. This is when

$$x - c < \pi < \frac{c^2 - x^2 + \sqrt{(c-x)(c+x)(c-1+x)(c+1-x)}}{1-2x}.$$

4.2 Welfare Consequences of Parties

Informational constraints also meaningfully affect parties' equilibrium payoffs. Below I compare each party's equilibrium payoffs in Model 1 and Model 2, as well as the case where parties exogenously receive perfect information about the state of the world.



(a) Party 1 (b) Party 2

Figure 4: Comparison of Payoffs
$c = 0.15$

Each party's equilibrium payoff in the unconstrained and the constrained models are represented by black and red translucent lines in Figure 4, respectively. The blue dashed line additionally illustrates the parties' payoffs from perfectly learning about the state of the world. With no constraints on Party 1's information provision power, Party 1 (2) achieves her best (worst) possible payoff; Party 1 is able to extract all the potential surplus that can be gained via information about the state of the world. A comparison with constrained information design is as expected. Party 1 obviously prefers to have more power and thus prefers to be unconstrained, while Party 2 is weakly better off in the model where Party 1 is constrained.

Another notable property in Party 2's payoff under constrained information design is the presence of discontinuity at $\mu^{\dagger}$.[7] When π is sufficiently small, i.e., when Party 1 is less likely to be innocent, Party 1 prefers Party 2 to be unwilling to fight. As π increases, however, Party 1 is more willing to attack Party 2 since the state of the world is more likely to be favorable to her. This increases Party 1's incentive to design herself into the willingness to fight. Such a change in the choice of information creates a discontinuity in Party 2's payoff.

Lastly, I find that parties often strictly prefer to let Party 1 choose information with constraints over exogenously receiving perfect information about the state of the world. Insomuch as the actors have conflicting preferences over the state of the world, and conditional on c being sufficiently low, with perfect information about the state of the world parties always attack with a good signal and stay silent with a bad one. Note that with an exogenously given perfect information, there is no mutual silence in equilibrium for the given parameters. We can immediately observe that Party 1 is always weakly better off being able to endogenously choose how much information to provide. Party 2, on the other hand, has mixed preferences toward perfect information. He prefers to receive a fully informative signal when π is sufficiently low, but for larger values of π he benefits from the information that Party 1 designs. This means that it is sometimes *mutually beneficial* to let one party choose a preferred level of information provision rather than to receive a perfectly accurate public signal.

Proposition 4 *Suppose $c < x < 1 - c$. Both parties may prefer "less information." Party 1 always prefers to choose how much information to provide; Party 2 prefers Party 1's choice of information with constraints in the following two cases:*

1. ***Value of inaccuracy.*** *Party 2 prefers less information because it induces Party 1 to make mistakes. This is when*

$$\frac{c^2 - x^2 + \sqrt{(c-x)(c+x)(c-1+x)(c+1-x)}}{1-2x} < \pi < \frac{1}{2}$$

[7] $\mu^{\dagger}$ is the value of π that satisfies

$$\frac{\pi^2(1-x+c)}{\pi^2 + (x-c)(1-2\pi)} = x + c.$$

2. ***Value of silence.*** *Party 2 prefers less information because it enables collusion between the parties. This is when*

$$1 - \frac{1-x}{1-c} < \pi < \frac{c^2 - x^2 + \sqrt{(c-x)(c+x)(c-1+x)(c+1-x)}}{1-2x}$$

Proof: Presented in the appendix. ∎

Proposition 4 tells us that Party 2 prefers to let Party 1 be a constrained information designer under two mechanisms. First, when π is not too large, Party 2's additional benefit from letting Party 1 choose the public signal comes from her choice to partially obfuscate the information. Perfect information allows Party 1 to attack with a perfect success rate given signal 1. Endogenous information, however, significantly increases the false positive rate for Party 1's attack - Party 1 thus makes mistakes more often, which conversely increases Party 2's equilibrium payoff (**value of inaccuracy**). Note that π is the likelihood that Party 1 comes out of the signal unharmed; when the likelihood is low, this means that Party 1's choice of inaccuracy will further decrease her chances of a successful attack. This strategy, however, is optimal to Party 1. She voluntarily risks making more mistakes by setting lower levels of accuracy in order to deter Party 2 from attacking. When π is slightly larger, Party 2 prefers Party 1's choice of no information (**value of silence**). Here mutual silence benefits both parties, and Party 2 thus prefers that his opponent, Party 1, choose total noise over receiving perfect information about the state of the world.

4.3 Bargaining over the Initial Division

I further examine whether the presence of constraints changes parties' preferences over the initial division. Suppose that Party 1 knows a scandal is on the horizon and wants to grant concessions to Party 2. If parties could bargain over the status quo division *prior to* playing the game, would they agree on a different division? How will the parties' payoffs change? The sequence of the pre-game bargaining is as follows:

1. Nature draws a true state $\omega \in \{0,1\}$, where

$$\omega = \begin{cases} 0 & \text{w.p. } 1-\pi \\ 1 & \text{w.p. } \pi. \end{cases}$$

2. Parties observe the cost of attacking c and the status quo division $(x, 1-x)$.

3. Party 1 decides whether to keep the current status quo division $(x, 1-x)$ or propose a new division of dollar $(x', 1-x')$. If Party 1 keeps the current division, the parties move on to play the baseline game.

4. If Party 1 proposes a new division, Party 2 decides whether or not to accept the proposal. If Party 2 accepts, the parties move on to play the baseline game with the new division. If Party 2 rejects, the parties move on to play the baseline game with the initial division.

I look for Party 1's optimal offer $(x', 1-x')$ given status quo division $(x, 1-x)$, cost c, and probability π. Note that in equilibrium, the new division Pareto dominates the status quo division. Otherwise, Party 1 keeps the current status quo division.

Importantly, I find that parties can prefer to give up their initial share of the pie only when the information designer is constrained. When the designer can choose any signal, Party 1 can manipulate the information to always achieve the supremum of the convex hull (of her induced utility), and thus always prefers to have a larger share of the initial pie. Party 2 never benefits from information in this case and prefers to have a higher status quo payoff. With a constrained designer, however, both parties may have incentives to concede some of their initial pie in exchange for mutual silence. This is stated in Proposition 5.

Proposition 5 *Consider the model with a constrained designer where parties can bargain over their initial status quo division. In equilibrium,*

1. *Party 1 offers* $x' = 1-(1-\pi)(1-x)$ *and* $pS \to SS$ *when*
   - $c > 1/2$ *and* $\left\{(\pi < 1/2 \text{ and } 1-\frac{c}{1-\pi} < x < 1-c) \text{ or } (\pi > 1/2 \text{ and } x < 1-c)\right\}$

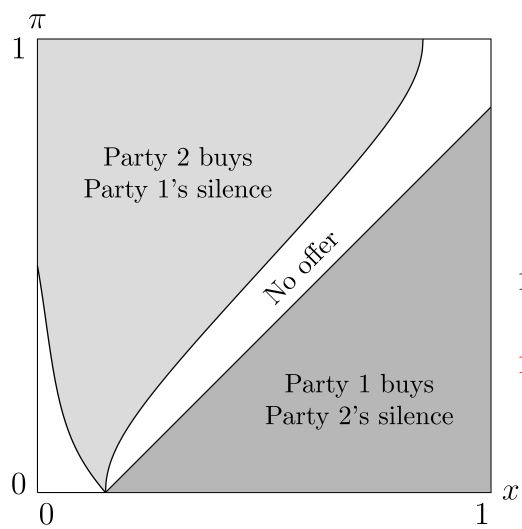
- *$c < 1/2$ and $\left\{ (\pi < 1/2 \text{ and } 1 - \frac{1-x^\dagger}{1-\pi} < x < c) \text{ or } (\pi > 1/2 \text{ and } x < c) \right\}$*

2. *Party 1 offers $x' = c + \pi$ and $Sq \to SS$ when ($c < 1/2$ or ($c > 1/2$ and $\pi < 1/2$)) and $x > c + \pi$.*

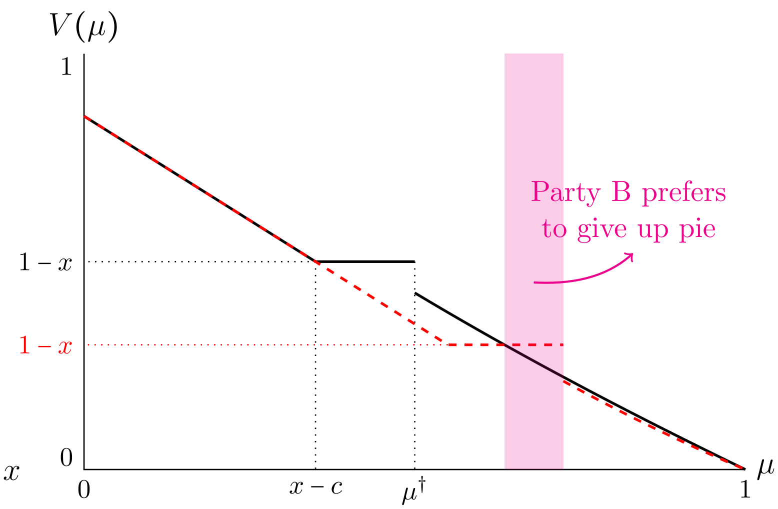
3. *Party 1 offers $x' = \frac{\pi(\pi(1+c-2x)+x)}{\pi+c(-1+2\pi)+x-2\pi x}$ and $pS \to SS$ when $c < 1/2$ and $c < x < x^\dagger$.*

The new division always leads to both parties staying silent in equilibrium.

Proof: Presented in the appendix. ∎



(a) Equilibrium Outcomes



(b) Party 2's Induced Utility

Figure 5: Pre-game Settlement in Model 2
$c = 0.15$

The shaded regions in Figure 5a represent values of x and π where such Pareto optimal division exists. More specifically, Figure 5b illustrates the intuition behind the results. Suppose the initial division is $1 - x = 0.5$ (black line), and Party 1's new offer is $1 - x = 0.3$. For all values of π in the pink-shaded region, Party 2 is strictly better off with the new offer. In equilibrium, Party 1 proposes to take more share of the pie ($x' > x$) and Party 2 accepts the offer. By "offering" some portion of his initial share to Party 1, Party 2 gains an additional payoff by reducing Party 1's chances of attacking. This is represented by the lightly shaded region in Figure 5a; the logic is identical for the darkly shaded region. Essentially, with a new Pareto optimal division $(x', 1 - x')$, both parties become sufficiently satisfied with the new division that they are disincentivized to attack each other. Proposition 5 also tells us that the new division in equilibrium always leads to mutual silence. Parties may therefore prefer to compromise some of their pie to the other player to essentially "buy each other's silence."

Further note that when both parties stay silent in equilibrium with the initial division, Party 1 never proposes a new division, i.e. there is no offer that Pareto dominates the original division.[8] This means that bargaining over the status quo division strictly increases the size of the collusion equilibrium. The pre-game settlement, in this sense, further allows parties to collude and keep the scandal in the dark.

5 Discussion

Silence in politics. When do parties leave some stones unturned? The model finds that imposing an informational constraint on the designer generates a large region of an additional collusive equilibrium where parties remain silent in equilibrium. Without any restriction on information, collusion occurs only when x and π are sufficiently high, i.e., when Party 1 is already highly likely to win, and so she has *no need to* manipulate information. Conversely, with constraints, we observe collusion much more frequently - parties remain silent when the status quo x and the outlook of the scandal π are similar in size. Here, collusion equilibrium always exists given a positive cost of attacking ($c > 0$), meaning that silence can be sustained even when the cost of attacking is extremely low. Parties also have preferences for this outcome. In a model with constraints, when parties can bargain over the status quo division, parties sometimes willingly concede some of their pie to their opponent; and the new division they agree on *always* leads to more collusion between parties. Although outside of this model, if we assume voters' payoffs as a function of the information they learn about the scandal, this result tells us that collusion can be both a stable phenomenon and

[8]Party 1 in equilibrium may not propose a new division when Party 1 attacks in the original equilibrium. However, she always does when Party 2 attacks and never does when both never attack in the baseline model.

an outcome that can be mutually beneficial for both parties.

Political investigations. Then when do political investigations occur? The endogeneity of signal accuracy is crucial to understanding the conditions under which parties investigate each other. Without any constraints on information, Party 1 is able to successfully manipulate the information such that makes Party 2 barely unwilling to attack after signal 1 (when π is low) and after signal 0 (when π is high). With constraints, however, we may observe a less intuitive dynamic where Party 1 sometimes voluntarily chooses to learn less information even when she knows that doing so increases her chances of making a mistake; she may also learn more information even when she has more to lose from the status quo or has a dim outlook on the scandal. This also comes from the designer's incentives to deter the other party from attacking and learning the truth, but the limited degree to which she can do so forces her to forego some of her mistakes in exchange for a successful deterrence. This further shapes parties' preferences toward the status quo division. In particular, parties may prefer to have less share of the pie since the reduction in Party 2's share of the status quo division induces Party 1 to set a completely uninformative signal that results in collusion.

References

Alonso, Ricardo and Odilon Câmara. 2016. "Persuading voters." *American Economic Review* 106(11):3590–3605.

Austen-Smith, David. 1993. "Information and influence: Lobbying for agendas and votes." *American Journal of Political Science* 37(3):799–833.

Awad, Emiel. 2020. "Persuasive lobbying with allied legislators." *American Journal of Political Science* 64(4):938–951.

Bergemann, Dirk and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57(1):44–95.

Berliner, Daniel. 2014. "The political origins of transparency." *The journal of Politics* 76(2):479–491.

Beshkar, Mostafa and Jee Hyeong Park. 2016. Pre-trial settlement with imperfect private monitoring. Technical report Mimeo.

Bils, Peter, Robert J Carroll and Lawrence S Rothenberg. 2020. Strategic Avoidance and Rulemaking Procedures. Technical report Working Paper. URL: https://www. wallis. rochester. edu/assets/pdf . . . .

Bowler, Shaun and Jeffrey A Karp. 2004. "Politicians, scandals, and trust in government." *Political behavior* 26:271–287.

Cotton, Christopher S and Cheng Li. 2018. "Clueless politicians: on policymaker incentives for information acquisition in a model of lobbying." *The Journal of Law, Economics, and Organization* 34(3):425–456.

Dellis, Arnaud. 2023. "Legislative informational lobbying." *Journal of Economic Theory* 208:105595.

Dewan, Torun and David P Myatt. 2008. "The qualities of leadership: Direction, communication, and obfuscation." *American Political science review* 102(3):351–368.

Dughmi, Shaddin. 2017. "Algorithmic information structure design: a survey." *ACM SIGecom Exchanges* 15(2):2–24.

Dziuda, Wioletta and William G Howell. 2021. "Political scandal: a theory." *American Journal of Political Science* 65(1):197–209.

Ellis, Christopher J and Thomas Groll. 2020. "Strategic legislative subsidies: Informational lobbying and the cost of policy." *American Political Science Review* 114(1):179–205.

Entman, Robert M. 2012. *Scandal and silence: Media responses to presidential misconduct.* Polity.

Gailmard, Sean and John W Patty. 2012. "Formal models of bureaucracy." *Annual Review of Political Science* 15:353–377.

Gailmard, Sean and John W Patty. 2017. "Participation, process and policy: the informational value of politicised judicial review." *Journal of Public Policy* 37(3):233–260.

Gal-Or, Esther. 1985. "Information sharing in oligopoly." *Econometrica: Journal of the Econometric Society* pp. 329–343.

Gavazza, Alessandro and Alessandro Lizzeri. 2007. "The perils of transparency in bureaucracies." *American Economic Review* 97(2):300–305.

Gehlbach, Scott and Konstantin Sonin. 2014. "Government control of the media." *Journal of public Economics* 118:163–171.

Gerardi, Dino, Edoardo Grillo and Ignacio Monzón. 2022. "The perils of friendly oversight." *Journal of Economic Theory* 204:105500.

Gradwohl, Ronen and Timothy Feddersen. 2018. "Persuasion and transparency." *The Journal of Politics* 80(3):903–915.

Herrera, Helios, Ernesto Reuben and Michael M Ting. 2017. "Turf wars." *Journal of Public Economics* 152:143–153.

Invernizzi, Giovanna. 2016. "Political Scandals.".

Kamenica, Emir. 2019. “Bayesian persuasion and information design.” *Annual Review of Economics* 11:249–272.

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

Lee, Francis LF. 2018. “The spillover effects of political scandals: The moderating role of cynicism and social media communications.” *Journalism & Mass Communication Quarterly* 95(3):714–733.

Libgober, Brian D. 2020. “Strategic proposals, endogenous comments, and bias in rulemaking.” *The Journal of Politics* 82(2):642–656.

Luo, Zhaotian and Arturas Rozenas. 2018. “Strategies of election rigging: trade-offs, determinants, and consequences.” *Quarterly Journal of Political Science* 13(1):1–28.

Luo, Zhaotian and Arturas Rozenas. 2022. “Ruling the ruling coalition: Information control and authoritarian power-sharing.” *Quarterly Journal of Political Science, forthcoming* .

Meade, Ellen E and David Stasavage. 2008. “Publicity of debate and the incentive to dissent: Evidence from the US Federal Reserve.” *The Economic Journal* 118(528):695–717.

Meijer, Albert. 2013. “Understanding the complex dynamics of transparency.” *Public administration review* 73(3):429–439.

Nyhan, Brendan. 2017. “Media scandals are political events: How contextual factors affect public controversies over alleged misconduct by US governors.” *Political Research Quarterly* 70(1):223–236.

Patty, John W and Elizabeth Maggie Penn. 2022. “Ban The Box? Information, Incentives, and Statistical Discrimination.” *arXiv preprint arXiv:2208.08348* .

Patty, John W and Elizabeth Maggie Penn. 2023. “Algorithmic fairness and statistical discrimination.” *Philosophy Compass* 18(1):e12891.

Potters, Jan and Frans Van Winden. 1992. “Lobbying and asymmetric information.” *Public Choice* 74:269–292.

Prato, Carlo and Ian R Turner. 2022. "Institutional foundations of the power to persuade." *Center for Open Science SocArXiv* (4w9af).

Sberna, Salvatore and Alberto Vannucci. 2013. ""It's the politics, stupid!". The politicization of anti-corruption in Italy." *Crime, law and social change* 60:565–593.

Schnakenberg, Keith. 2017. "Informational lobbying and legislative voting." *American Journal of Political Science* 61(1):129–145.

Schnakenberg, Keith and Ian Turner. 2019. "Signaling with reform: How the threat of corruption prevents informed policy-making." *American Political Science Review* 113(3):762–777.

Schnakenberg, Keith and Ian Turner. 2021. "Helping friends or influencing foes: Electoral and policy effects of campaign finance contributions." *American Journal of Political Science* 65(1):88–100.

Schnakenberg, Keith and Ian Turner. 2023. "Formal Theories of Special Interest Influence." *Working Paper* .

Sowers, Thomas E, James P Nelson et al. 2016. "The Timing of Presidential Scandals: The Role of Economics, Divided Government and the Media." *Open Journal of Political Science* 6(01):83.

Stasavage, David. 2003. "Transparency, democratic accountability, and the economic consequences of monetary institutions." *American Journal of Political Science* 47(3):389–402.

Stasavage, David. 2007. "Polarization and publicity: rethinking the benefits of deliberative democracy." *The Journal of Politics* 69(1):59–72.

Thompson, John B. 2013. *Political scandal: Power and visability in the media age*. John Wiley & Sons.

Tiffen, Rodney. 1999. *Scandals: Media, politics & corruption in contemporary Australia*. Thomas Telford.

Tomasi, Arduino. 2022. "Merchants of Reputation Privatization Under Elites' Manipulation of Information." *Available at SSRN 4112044* .

Turner, Ian. 2021. "Reviewing Procedure vs. Judging Substance: How Increasing Bureaucratic Oversight Can Reduce Bureaucratic Accountability.".

Turner, Ian R. 2017. "Working smart and hard? Agency effort, judicial review, and policy precision." *Journal of Theoretical Politics* 29(1):69–96.

von Sikorski, Christian, Raffael Heiss and Jörg Matthes. 2020. "How political scandals affect the electorate. Tracing the eroding and spillover effects of scandals with a panel study." *Political Psychology* 41(3):549–568.

Wolsky, Adam D. 2022. "Scandal, hypocrisy, and resignation: How partisanship shapes evaluations of politicians' transgressions." *Journal of Experimental Political Science* 9(1):74–87.

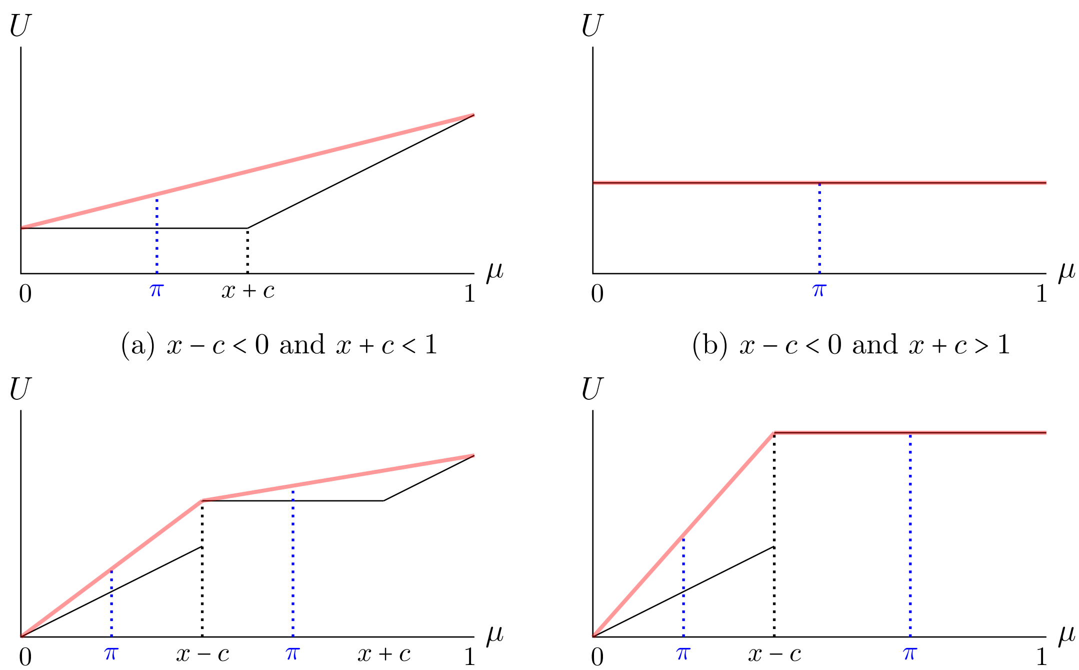
A Proofs

A.1 Proposition 1

Proposition 1 examines Party 1's optimal choice of signal in a game where Party 1 has full flexibility in her choice of signal structure. I view Party 1's *ex-post* payoff as a function of the posterior belief.

$$U(\mu) = \begin{cases} \mu & \text{if } \mu < x - c \\ x & \text{if } \mu \in [x - c, x + c] \\ \mu - c & \text{if } \mu > x + c. \end{cases}$$

The indirect utility U is a piecewise linear function of μ. I apply the concavification method.



(a) $x - c < 0$ and $x + c < 1$

(b) $x - c < 0$ and $x + c > 1$

(c) $x - c > 0$ and $x + c < 1$

(d) $x - c > 0$ and $x + c > 1$

Figure A.1: Different Shapes of Indirect Utility $U(\mu)$

Since μ is between 0 and 1, the shape of the function depends on whether $x - c$ is larger than 0 and whether $x + c$ is larger than 1. There are four different cases.

1. In Figure A.1a, the smallest concave function is a linear function that connects $U(0)$

to $U(1)$ when $x - c < 0$ and $x + c < 1$. In other words, the optimal signal structure is a binary experiment that perfectly informs the parties about the state of the world.

2. In Figure A.1b, the indirectly utility $U(\mu)$ is constant when $x - c < 0$ and $x + c > 1$. Therefore, any signal structure will lead to the same ex-ante and ex-post outcome and payoff.

3. In Figure A.1c, the smallest concave function is a piecewise linear function that connects $U(0)$ to $U(x-c)$ then $U(x-c)$ to $U(1)$ when $x-c > 0$ and $x+c < 1$. The optimal signal structure that satisfies Bayesian plausibility is a binary experiment that mixes between 0 and $x - c$ if the prior is smaller than $x - c$ (i.e., π on the left), and a binary experiment that mixes between $x - c$ and 1 if the prior is larger than $x - c$ (i.e., π on the right).

4. In Figure A.1d, the smallest concave function is a piecewise linear function that connects $U(0)$ to $U(x-c)$ then $U(x-c)$ to $U(1)$ when $x-c > 0$ and $x+c > 1$. The optimal signal structure that satisfies Bayesian plausibility is a binary experiment that mixes between 0 and $x - c$ if the prior is smaller than $x - c$ (i.e., π on the left). The optimal signal structure is to provide no information if the prior is larger than $x - c$ (i.e., π on the right).[9]

I summarize Party 1's optimal choice of signal structure below.

1. It is optimal to provide perfect information when

$$x < \min\{c, 1 - c\}.$$

2. It is optimal to provide any information when

$$1 - c < x < c.$$

[9] Technically, any signal structure of which support is a subset of interval $[x - c, 1]$ can achieve the same outcome as long as Bayesian plausibility is satisfied. However, I focus on no information provision for a more sensible comparison with the baseline model.

3. It is optimal to provide no information when

$$\max\{c, 1-c\} < x < c + \pi.$$

4. It is optimal to mix between $1, x-c$ when

$$c < x < \min\{1-c, c+\pi\}.$$

5. It is optimal to mix between $0, x-c$ when

$$x > c + \pi.$$

A.2 Proposition 2

Now consider imposing some constraints on the designer's power to choose the signal structure. In this version, the designer is constrained to only choose a binary and symmetric signal. Let $\sigma \in \{0,1\}$ denote the binary signal, and let $\mu(\sigma)$ denote the posterior belief of probability that $\omega = 1$ after observing signal σ. Without loss of generality, assume that $\mu(0) < \mu(1)$. Conceptually, we have

$$Pr(\omega = 1|\sigma = 0) = \mu(0),$$
$$Pr(\omega = 1|\sigma = 1) = \mu(1).$$

Bayes plausibility implies

$$Pr(\sigma = 0) = \frac{\mu(1) - \pi}{\mu(1) - \mu(0)},$$
$$Pr(\sigma = 1) = \frac{\pi - \mu(0)}{\mu(1) - \mu(0)}.$$

Bayes rule implies

$$Pr(\sigma = 0|\omega = 0) = \frac{(1 - \mu(0))(\mu(1) - \pi)}{(1 - \pi)(\mu(1) - \mu(0))},$$
$$Pr(\sigma = 1|\omega = 1) = \frac{\mu(1)(\pi - \mu(0))}{\pi(\mu(1) - \mu(0))}.$$

Define *accuracy* as $Pr(\sigma = \omega|\omega)$; an accuracy is *symmetric* if $Pr(\sigma = \omega|\omega = 0) = Pr(\sigma = \omega|\omega = 1)$. I impose that the designer can only choose a binary signal structure with symmetric accuracy in the constrained problem: i.e.,

$$\frac{\pi}{1 - \pi} = \frac{\mu(1)(\pi - \mu(0))}{(1 - \mu(0))(\mu(1) - \pi)}.$$

Since π is exogenous, LHS is constant. RHS is decreasing in $\mu(0)$ and $\mu(1)$ respectively. Therefore, the constraint reduces the designer's power by making the choice of $\mu(0)$ depend on $\mu(1)$ and vice versa. For example, if the designer wishes to decrease $\mu(0)$, RHS will increase. Therefore, the designer must increase $\mu(1)$ accordingly to adjust the value back to

LHS.

It is straightforward to show that perfect information satisfies this equation. Full disclosure of the true ω is equivalent to binary signal that induces $\mu(0) = 0$ and $\mu(1) = 1$. Then,

$$\frac{1 \cdot (\pi - 0)}{(1-0)(1-\pi)} = \frac{\pi}{1-\pi}.$$

Similarly, we can show that no information satisfies the equation. An uninformative signal structure is equivalent to a binary signal that induces $\mu(0) = \mu(1) = \pi$. Then,

$$\frac{\pi \cdot (\pi - \pi)}{(1-\pi)(\pi-\pi)} = \frac{\pi}{1-\pi}.$$

Consider on the other hand a binary signal that induces $\mu(0) = 0$ and $\mu(1) = x - c$:

$$\frac{(x-c)(\pi - 0)}{(1-0)(x-c) - \pi} = \frac{\pi(x-c)}{x-c+\pi} \neq \frac{\pi}{1-\pi}.$$

The equation is not satisfied when $0 < \pi < x - c$. Similarly, consider a binary signal that induces $\mu(0) = x - c$ and $\mu(1) = 1$:

$$\frac{1 \cdot (\pi - x + c)}{(1-x+c)(1-\pi)} = \frac{\pi - x + c}{(1-x+c)(1-\pi)} \neq \frac{\pi}{1-\pi}.$$

The equation is not satisfied when $0 < x - c < \pi$ and $x + c < 1$.

Two things for the ease of proof:

1. Total utility is a convex combination of $(\mu(0), U(\mu(0)))$ and $(\mu(1), U(\mu(1)))$ with weight $\alpha \in (0,1)$ such that

   $$\alpha\mu(0) + (1-\alpha)\mu(1) = \pi \quad \Leftrightarrow \quad \alpha = \frac{\mu(1) - \pi}{\mu(1) - \mu(0)}.$$

   Note that α is equivalent to the probability that $\sigma = 0$.

2. Symmetry of accuracy condition allows us to write $\mu(0)$ in terms of $\mu(1)$ and vice

versa:

$$\mu(0) = \frac{\pi^2(1-\mu(1))}{\pi^2 - 2\pi\mu(1) + \mu(1)},$$
$$\mu(1) = \frac{\pi^2(1-\mu(0))}{\pi^2 - 2\pi\mu(0) + \mu(0)}.$$

We confirmed that these two forms of optimal signal structure are infeasible under the constraints. I show that the solutions to the constrained problem are not too deviating from the unconstrained one. Formally,

1. The optimal signal structure induces $\mu^*(1) = x - c$ when $0 < \pi < x - c$.

   (a) Consider an alternative such that induces $\mu'(0) > \mu^*(0)$. Then, the utility of this strategy is π since $\mu'(\sigma) < x - c$ for all σ. Therefore, these strategies are dominated by μ^*.

   (b) Consider an alternative such that induces $\mu'(0) < \mu^*(0)$ and $\mu^*(1) < \mu'(1) < x + c$. The utility of this strategy is

   $$\alpha(\mu'(0), \mu'(0)) + (1-\alpha)(\mu'(1), x),$$

   which is increasing in $\mu'(0)$. Then, the local maximum is at $\mu'(0) \to \mu^*(0)$, which is dominated by μ^*.

   (c) Consider an alternative such that induces $\mu'(0) < \mu^*(0)$ and $\mu^*(1) > x + c$. The utility of this strategy is

   $$\alpha(\mu'(0), \mu'(0)) + (1-\alpha)(\mu'(1), \mu'(1) - c).$$

      i. The utility is increasing in $\mu'(0)$ if $\pi > 1/2$. Then, the local maximum is at $\mu'(1) = x + c$, which is dominated by μ^*.

      ii. The utility is decreasing in $\mu'(0)$ if $\pi < 1/2$. Then, the local maximum is at $\mu'(0) = 0$, which is dominated by μ^*.

2. The optimal signal structure induces $\mu^*(0) = x - c$ when $0 < x - c < \pi$ and $x + c < 1$.

(a) Suppose that $\mu^*(1) < x + c$. Then, the designer is indifferent between the choice $\mu(1) \in [\pi, \mu^*(1)]$. Essentially, the designer is choosing an uninformative signal structure with the payoff of x.

i. Consider an alternative such that induces $\mu'(0) < \mu^*(0)$ and $\mu^*(1) < \mu'(1) < x + c$. The utility of this strategy is

$$\alpha(\mu'(0), \mu'(0)) + (1 - \alpha)(\mu'(1), x),$$

which cannot be larger than x as a convex combination.

ii. Consider an alternative such that induces $\mu'(0) < \mu^*$ and $\mu'(1) > x + c$. The utility of this strategy is

$$\alpha(\mu'(0), \mu'(0)) + (1 - \alpha)(\mu'(1), \mu'(1) - c).$$

A. The utility is increasing in $\mu'(0)$ if $\pi > 1/2$. Then, the local maximum is at $\mu'(1) = x + c$, which is dominated by μ^*.

B. The utility is decreasing in $\mu'(0)$ if $\pi < 1/2$. Then, the local maximum is at $\mu'(0) = 0$, which is dominated by μ^*.

(b) Suppose that $\mu^*(1) > x + c$. Then, the optimal signal meaningfully differentiates the outcome.

i. Consider an alternative such that induces $\mu'(0) > \mu^*(0)$ and $\mu'(1) < x + c$. Then, the utility of this strategy is x since $x - c < \mu'(\sigma) < x + c$ for all σ. Therefore, these strategies are dominated by μ^*.

ii. Consider an alternative such that induces $\mu^*(0) < \mu'(0) < x + c$ and $x + c < \mu'(1) < \mu^*(1)$. Then, the utility of this strategy is

$$\alpha(\mu'(0), x) + (1 - \alpha)(\mu'(1), \mu'(1) - c),$$

which is decreasing in $\mu'(0)$. Then, the local maximum is at $\mu'(0) \to \mu^*(0)$, which is dominated by μ^*.

iii. Consider an alternative such that induces $\mu'(0) > x + c$ and $\mu'(1) < \mu^*(1)$. Then, the utility of this strategy is $\pi - c$ since $\mu'(\sigma) > x + c$ for all σ. Therefore, these strategies are dominated by μ^*.

iv. Consider an alternative such that induces $\mu'(0) < \mu^*(0)$ and $\mu'(1) > \mu^*(1)$. Then, the utility of this strategy is

$$\alpha(\mu'(0), \mu'(0)) + (1 - \alpha)(\mu'(1), \mu'(1) - c). \tag{1}$$

A. The utility is increasing in $\mu'(0)$ if $\pi > 1/2$. Then, the local maximum is at $\mu'(1) \to \mu^*(1)$, which is dominated by μ^*.

B. The utility is decreasing in $\mu'(0)$ if $\pi < 1/2$. Then, the local maximum is at $\mu'(0) = 0$, which is dominated by μ^*.

A.3 Proposition 3

To prove Proposition 3, I compare the equilibrium outcomes of the two models below.

1. Consider $x-c<0$ and $x+c>1$. The signal structure does not matter because equilibrium outcomes are always (S,S) for all μ. This is true for both the unconstrained and the constrained problem.

2. Consider $x-c<0$ and $x+c<1$. Then, the optimal choice of signal structure is perfect information in the unconstrained problem. Specifically, $\mu(0)=0$ induces (S,S) and $\mu(1)=1$ induces (A,S) in equilibrium. This is also true for the constrained problem since perfect information can be achieved with symmetric accuracy.

3. Consider $0<x-c<\pi$ and $x+c>1$. Then, the optimal choice of signal structure is no information in the unconstrained problem. Specifically, $\mu(0)=\mu(1)=\pi$ induces (S,S) in equilibrium. Similarly, this is also true for the constrained problem since no information can be achieved with symmetric accuracy.

4. Consider $0<x-c<\pi$ and $x+c<1$. Then, the optimal choice of signal structure is a binary signal such that $\mu(0)=x-c$ induces (S,S) and $\mu(1)=1$ induces (A,S). This cannot be achieved in the constrained problem. The designer will instead choose a binary signal such that generates $\mu(0)=x-c$ and $\mu(1)=\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)}$.

    (a) If $\mu(1)>x+c$ then $\mu(0)$ induces (S,S) and $\mu(1)$ induces (A,S).

    (b) If $\mu(1)<x+c$ then it is essentially no information that induces (S,S) for all σ.

5. Consider $0<\pi<x-c$. Then, the optimal choice of signal structure is a binary signal such that $\mu(0)=0$ induces (S,A) and $\mu(1)=x-c$ induces (S,S). Similarly, this cannot be achieved in the constrained problem. The designer will instead choose a binary signal such that $\mu(0)=\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)}$ induces (A,S) and $\mu(1)=x-c$ induces (S,S).

A.4 Proposition 4

Consider the setting where the parties are provided perfect information by an exogenous signal structure that induces $\mu(0) = 0$ and $\mu(1) = 1$. Since perfect information is in the designer's feasible choice set under both unconstrained and constrained problems, the designer is always weakly better off with the power to endogenously choose the signal structure. We explore whether the receiver prefers to endow the designer with such power as well.

In the unconstrained problem, the set of signal structures in the supremum of a convex hull constructed by the designer's indirect utility is equal to the set of signal structures in the infimum of a convex hull constructed by the receiver's indirect utility. This implies that the designer's endogenous choice of signal structure always leads to the worst scenario for the receiver without constraint. In other words, their interests are completely conflicting. Therefore, we compare the exogenous information only with the endogenous information *with* constraint.

1. Consider $x - c < 0$ and $x + c > 1$. Then, the choice of signal structure does not matter. The receiver is indifferent.

2. Consider $x - c < 0$ and $x + c < 1$. Then, endogenous design leads to perfect information. The receiver is indifferent. Interestingly, the utility from perfect information is in the infimum of a convex hull constructed by the receiver's indirect utility. In other words, the designer chooses perfect information when it is the worst scenario for the receiver.

3. Consider $x - c > 0$ and $x + c > 1$. Then, the utility from perfect information is in the supremum of a convex hull constructed by the receiver's indirect utility. However, endogenous design never leads to perfect information. In other words, the designer does not choose perfect information when it is the best scenario for the receiver. Therefore, the receiver does not want to endow the designer with the power to endogenously choose the signal structure.

4. Consider $x - c > 0$ and $x + c < 1$. It is not obvious whether the receiver prefers exogenous or endogenous information. The receiver gets $(1 - \pi)(1 - c)$ from exogenous perfect information.

(a) If $\pi < x - c$, then the receiver receives $1 - \pi - c$ from endogenous information, which is in the infimum of a convex hull constructed by the receiver's indirect utility. Therefore, the receiver does not want to endow the designer with the power to endogenously choose the signal structure.

(b) If $\pi > x - c$ and $\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)} < x + c$ then the receiver receives $1 - x$ from endogenous information. Perfect information always leads to an attack either by the designer or the receiver. This can be harmful to the receiver, especially when π is too large. On the other hand, endogenous information leads to collusion in equilibrium: *value of silence*. Therefore, the receiver wants to endow the designer with the power to endogenously choose the signal structure when

$$1 - \frac{1-x}{1-c} < \pi < \frac{c^2 - x^2 + \sqrt{(c-x)(c+x)(c-1+x)(c+1-x)}}{1-2x}. \tag{2}$$

(c) If $\pi > x - c$ and $\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)} > x + c$, then the receiver receives $1 - x$ from $\sigma = 0$ and $1 - \mu(1)$ from $\sigma = 1$. Although perfect information allows the receiver to attack without error when $\sigma = 0$, it also allows the designer to attack without error when $\sigma = 1$. On the other hand, endogenous information significantly increases the false positive rate for the designer's attack. If π is sufficiently small, it may be more profitable for the receiver to not attack and just anticipate the designer's error: *value of inaccuracy*. Therefore, the receiver wants to endow the designer with the power to endogenously choose the signal structure when

$$\frac{c^2 - x^2 + \sqrt{(c-x)(c+x)(c-1+x)(c+1-x)}}{1-2x} < \pi < \frac{1}{2}. \tag{3}$$

Similarly, consider the setting where the parties are provided no information by the exogenous absence of the signal structure. It turns out that the receiver never strictly prefers to endow the designer with the power to endogenously choose the signal structure.

We already established that the designer without constraint optimally chooses the infimum of a convex hull constructed by the receiver's indirect utility. Obviously, this cannot be better than no information. This does not change much with constraint.

1. In regions where the designer chooses $\mu(0) = 0$ and $\mu(1) = x - c$ in the unconstrained problem, we observe *strategic clarity* equilibrium with constraint. In fact, the receiver's equilibrium payoff is identical with or without constraint: i.e., in the infimum.

2. In regions where the designer chooses $\mu(0) = x - c$ and $\mu(1) = 1$ in the constrained problem,

   (a) We observe *strategic obfuscation* when $\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)} > x + c$. While this is better than the unconstrained problem, the convex combination is still below the receiver's indirect utility. Therefore, no information is preferred.

   (b) We observe *collusion* when $\frac{\pi^2(1-x+c)}{\pi^2+(x-c)(1-2\pi)} < x + c$. In other words, the designer is endogenously choosing no information in the constrained problem, so the comparison with the exogenous absence of signal structure is trivial.

A.5 Proposition 5

Consider the setting where the parties can bargain over x prior to playing the constrained design problem. Specifically, both the designer (Party 1) and receiver (Party 2) observe the initial $x \in [0, 1]$. Then, the designer offers new $x' \in [0, 1]$. The receiver chooses whether to accept the new x' or proceed with the original x. Then, this extension is essentially looking for Pareto improving x such that maximizes the designer's utility if it exists.

1. In the *collusion* equilibrium, the parties receive

$$D = x,$$
$$R = 1 - x.$$

   There does not exist Pareto improving x' within this outcome.

   (a) Suppose that the designer chooses $x' > x$ such that can induce the *strategic clarity* equilibrium. Then, the parties receive

$$D' = \frac{\pi(\pi(1 + c - 2x') + x')}{\pi + (x - c)(1 - 2\pi)},$$
$$R' = 1 - \pi - c.$$

   This cannot be a Pareto improvement because it is always the case that $R' < R$.

   (b) Suppose that the designer chooses $x' < x$ such that can induce he t*strategic obfuscation* equilibrium. Then, the parties receive

$$D' = \pi - c + \frac{2\pi(1 - \pi)c}{\pi + (x' - c)(1 - 2\pi)},$$
$$R' = \frac{(1 - \pi)(\pi(1 - x') + (1 - \pi)(x' - c))}{\pi + (x' - c)(1 - 2\pi)}.$$

   This cannot be a Pareto improvement because conditions for $D' > D$ and $R' > R$ are not compatible.

   (c) Suppose that the designer chooses $x' < x$ such that can induce the *perfect infor-*

mation equilibrium. Then, the parties receive

$$D' = (1-\pi)x' + \pi(1-c),$$
$$R' = (1-\pi)(1-x').$$

This cannot be a Pareto improvement because conditions for $D' > D$ and $R' > R$ are not compatible.

Therefore, there does not exist a Pareto improvement from this equilibrium.

2. In the *perfect information* equilibrium, the parties receive

$$D = (1-\pi)x + \pi(1-c),$$
$$R = (1-\pi)(1-x).$$

There does not exist Pareto improving x' within this outcome.

(a) Suppose that the designer chooses $x' > x$ such that can induce the *strategic clarity* equilibrium. Then, the parties receive

$$D' = \frac{\pi(\pi(1+c-2x')+x')}{\pi+(x-c)(1-2\pi)}, \tag{4}$$
$$R' = 1-\pi-c. \tag{5}$$

This cannot be a Pareto improvement because it is always that $R' < R$.

(b) Suppose that the designer chooses $x' > x$ such that can induce the *strategic obfuscation* equilibrium. Then, the parties receive

$$D' = \pi - c + \frac{2\pi(1-\pi)c}{\pi+(x'-c)(1-2\pi)}, \tag{6}$$
$$R' = \frac{(1-\pi)(\pi(1-x')+(1-\pi)(x'-c))}{\pi+(x'-c)(1-2\pi)}. \tag{7}$$

This cannot be a Pareto improvement because conditions for $D' > D$ and $R' > R$ are not compatible.

(c) Suppose that the designer chooses $x' > x$ such that can induce the *collusion* equilibrium. Then, the parties receive

$$D' = x', \tag{8}$$
$$R' = 1 - x'. \tag{9}$$

This can be a Pareto improvement if

$$(1-\pi)x + \pi(1-c) < x' < 1 - (1-\pi)(1-x). \tag{10}$$

Therefore, the designer will choose $x' = 1 - (1-\pi)(1-x) > x$ such that induces the *collusion* equilibrium as Pareto improving division if feasible. This is when

$$\max\left\{1 - \frac{c}{1-\pi}, \frac{\sqrt{c^2(1-2\pi)^2 + (1-\pi)^2\pi^2} - (1-\pi)\pi}{(1-\pi)(1-2\pi)}\right\} < x < \min\{c, 1-c\}. \tag{11}$$

3. In the *strategic clarity* equilibrium, the parties receive

$$D = \frac{\pi(\pi(1+c-2x)+x)}{\pi + (x-c)(1-2\pi)}, \tag{12}$$
$$R = 1 - \pi - c. \tag{13}$$

The receiver's payoff R is not only constant with respect to x but is also the worst he could get for exogenous π and c. Therefore, we can just search for the designer's optimal alternative x'.

(a) Suppose that the designer chooses $x' < x$ such that can induce the *perfect information* equilibrium. The designer receives

$$D' = (1-\pi)x' + \pi(1-c). \tag{14}$$

(b) Suppose that the designer chooses $x' < x$ such that can induce the *strategic obfus-*

cation equilibrium. The designer receives

$$D' = \pi - c + \frac{2\pi(1-\pi)c}{\pi + (x'-c)(1-2\pi)}. \tag{15}$$

(c) Suppose that the designer chooses $x' < x$ such that can induce the *collusion* equilibrium. The designer receives

$$D' = x', \tag{16}$$

which is increasing in x'.

Therefore, the designer will choose $x' = c + \pi < x$ such that induces the *collusion* equilibrium as a Pareto improving division.

4. In the *strategic obfuscation* equilibrium, the parties receive

$$D = \pi - c + \frac{2\pi(1-\pi)c}{\pi + (x-c)(1-2\pi)}, \tag{17}$$

$$R = \frac{(1-\pi)(\pi(1-x) + (1-\pi)(x-c))}{\pi + (x-c)(1-2\pi)}. \tag{18}$$

There does not exist a Pareto improving x' within this outcome.

(a) Suppose that the designer chooses $x' < x$ such that can induce the *perfect information* equilibrium. Then, the parties receive

$$D' = (1-\pi)x' + \pi(1-c), \tag{19}$$

$$R' = (1-\pi)(1-x'). \tag{20}$$

This can be a Pareto improvement if

$$\frac{2c\pi}{\pi + (x-c)(1-2\pi)} - c < x' < \frac{c\pi}{\pi + (x-c)(1-2\pi)}. \tag{21}$$

(b) Suppose that the designer chooses $x' > x$ such that can induce the *collusion* equilibrium. Then, the parties receive

$$D' = x', \tag{22}$$

$$R' = 1 - x'. \tag{23}$$

This can be a Pareto improvement if

$$\pi - c + \frac{2c(1-\pi)\pi}{\pi + (x-c)(1-2\pi)} < x' < \frac{\pi(\pi(1+c-2x)+x)}{\pi + (x-c)(1-2\pi)}. \tag{24}$$

(c) Suppose that the designer chooses $x' > x$ such that can induce the *strategic clarity* equilibrium. Then, the parties receive

$$D' = \frac{\pi(\pi(1+c-2x')+x')}{\pi + (x-c)(1-2\pi)}, \tag{25}$$

$$R' = 1 - \pi - c. \tag{26}$$

This cannot be a Pareto improvement because it is always that $R' < R$.

Therefore, the designer will choose $x' = \frac{\pi(\pi(1+c-2x)+x)}{\pi+(x-c)(1-2\pi)} > x$ such that induces the *collusion* equilibrium as a Pareto improving division.