Policy Announcement Design*

Anna Cieslak[†], Semyon Malamud[‡], and Andreas Schrimpf[§]

This version: December 16, 2019

Preliminary and incomplete.

Abstract

We study the general problem of information design for a policymaker—a central bank—that communicates its private information (the "state") to the public. We show that it is optimal for the policymaker to partition the state space into a finite number of "clusters" and to communicate to the public to which cluster the state belongs. Optimal communication is more precise when the policymaker's beliefs conform with prior public expectations, but is more vague in case of divergence. We characterize the policymaker's trade-offs via a novel object—the information relevance matrix—and label its eigenvectors as principal information components (PICs). PICs with the highest eigenvalues determine the dimensions of information with the highest welfare sensitivity and, hence, are the ones that the policymaker should be most precise about.

Keywords: Central Bank Announcements, Learning, Bayesian Persuasion, Information Design

JEL: D82, D83, E52, E58

*We thank seminar participants at Caltech, UBC, and SUFE for helpful comments. Parts of this paper were written when Malamud visited the Bank for International Settlements (BIS) as a research fellow. The views in this article are those of the authors and do not necessarily represent those of BIS.

[†]Duke University, Fuqua School of Business and CEPR, E-mail: anna.cieslak@duke.edu

[‡]Swiss Finance Institute, EPF Lausanne, and CEPR; E-mail: semyon.malamud@epfl.ch

[§]Bank of International Settlements (BIS) and CEPR; E-mail: andreas.schrimpf@bis.org

1 Introduction

The importance of macroeconomic announcements is hard to overstate. Announcements help market participants gauge the state of the economy, they serve as a coordination device for different types of agents, and influence decisions taken by investors and firms. To the extent that macroeconomic announcements simply release the latest data (e.g., on unemployment or inflation), their design is straightforward. In the case of monetary policy, however, announcements about the current policy and its future path are left at the discretion of a policy-making body, such as the Federal Open Market Committee (FOMC) in case of the US Federal Reserve. Viewed as a key tool in achieving monetary policy objectives, central bank communication has received particular attention in the years following the global financial crisis, but surveys of market participants suggest that the Fed does not always earn the highest marks on this front (see Figure 1). This naturally raises the question of how a policymaker should *optimally* design announcements so as to maximize public welfare. The goal of this paper is to investigate this question.

Communicating the stance of monetary policy is a multifaceted task. Prior to the announcement, the policymaker needs to process large amounts of data about economic and financial conditions. It is then up to the policymaker to decide how to map this complex information into a simpler message that informs the public about the policy rate decision and conveys any supporting information on the policy stance. Adding to the complexity is the fact that policy communication does not occur in a vacuum but involves a two-way information flow between the central bank and the public (e.g., Shin (2017), Morris and Shin (2018)). Market participants spend significant resources on forecasting and interpreting central bank announcements. Central banks, in turn, pay close attention to signals in the market prices, from which they try to infer the expectations of investors in anticipation of policy actions.

We accommodate those practical challenges of central bank communication into a theoretical model of policy announcement design. Our analysis follows the recent literature

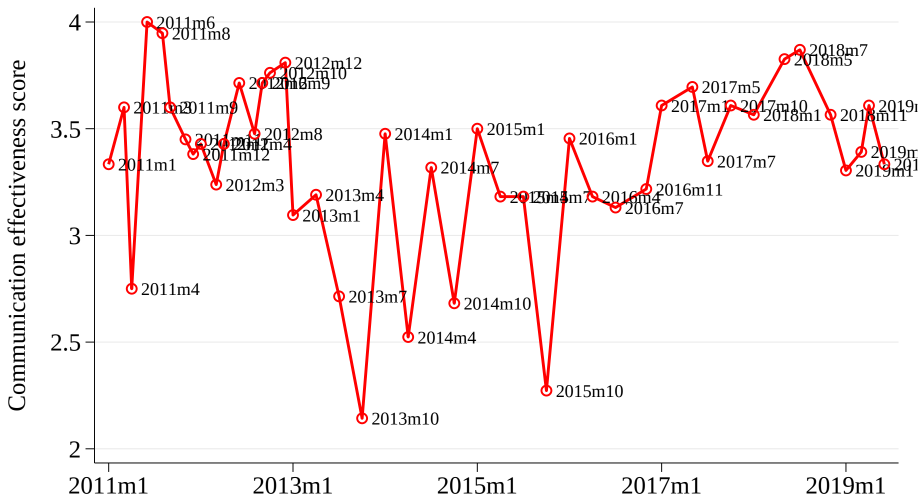


Figure 1: Fed communication effectiveness scores. The figure presents weighted average score from the Survey of Primary Dealers and the Survey of Market Participants compiled by the New York Fed. The survey question is: "How would you grade the Federal Reserve System's communication with the markets and with the public since the last policy survey? Please provide a rating between 1 and 5, with 1 indicating ineffectiveness and 5 indicating effectiveness."

on information design, initiated by Rayo and Segal (2010) and Kamenica and Gentzkow (2011). As in these papers, we assume that the policymaker can *commit* to the choice of the best information structure (from their perspective) that maximizes the social welfare ex-ante, before the state of the economy is realized. In general, the structure of the optimal design can be extremely complex and only existence results can be established (e.g., Bergemann and Morris, 2017). We focus on a special case of the general problem of "Bayesian persuasion" studied in Kamenica and Gentzkow (2011), with two modifications. We assume that (i) the private information of the policymaker (the state) is a continuous multidimensional random variable in $\mathbb{R}^L$, whereas (ii) the information structure chosen by the policymaker is finite, i.e., the signal sent to the public is always chosen from a finite set of possible announcements. Intuitively, having observed the realization of the state, the policymarker in our setting possesses a "dictionary" from which to select a small subset of messages that are best suited to enhance the public welfare. As such, the policymaker needs to decide about how best to design its communication reaction function, which should be a well-defined function of available data and opinions of the key decision makers.

Our first main result is that, under appropriate regularity conditions, it is optimal for the policymaker to partition the state space into a finite number of "clusters" (e.g., intervals when $L = 1$, rectangles when $L = 2$, etc.) and, then, to communicate to the public to which cluster the state belongs. This form of communication resembles common central bank practice. For example, central banks announce the interval to which the policy rate is expected to belong over certain horizon or they provide a confidence bands for the economic forecasts. Clearly, given the amount and dimensionality of information that policymakers face, it is not feasible to communicate all information in its raw form. Hence, the role of policy communication is to reduce the complexity of the multi-dimensional information bundle in a way that is accessible to the public. Our result formalizes this idea implying that, under optimal design, for each value of the state, the policymaker sends to the public a unique signal revealing the specific cluster to which state belongs. We expect that these clusters should have the property of "bunching" similar states together. We prove that this intuition is correct when the economy is close to the steady state: In the linearized problem, the partition consists of *convex polytops*.[1] Specifically, if two states belong to the same bin, then so does any convex combination of them, and therefore, mixing the two states does not change the nature of the announcement.[2] We use this general property to discuss some of the communication policies used by the Fed. For example, we argue that by violating the convexity property, the so-called "dot plots" are a suboptimal form of communication that could be improved upon.

We derive two more important and universal properties of linearized information designs. The first property corresponds to the one-dimensional case ($L = 1$), where the limiting partition is given by a set of intervals. We further assume that the prior distribution

[1]For example, intervals for $L = 1$, convex polygons for $L = 2$, and convex polyhedra for $L = 3$.

[2]The underlying linearized problem is analogous to the standard first-order approximation used in the macro literature; for example, it is this approximation that underlies the classical Taylor rule.

characterizing the beliefs of agents is unimodal. In this situation, optimal communication design implies that policymakers should be more precise when conforming with prior public expectations, but to be more vague in the case of a divergence of their beliefs from those of the public. This result lends credence to the idea that the central bank may want to avoid "spooking" financial markets, which in turn could manifest as gradualism in its decision making (see Stein and Sunderam (2018) for a recent discussion).

The second property corresponds to the multi-dimensional case ($L > 1$). We explicitly characterize the problem of the policymaker using a novel object—the information relevance matrix—which captures the trade-offs the policymaker faces given the public reaction function. The eigenvectors of that matrix correspond to what we call the *principal information components* (PIC) in analogy to the standard principal components known in the statistics literature. These eigenvectors define the directions of information that are most important for welfare under the equilibrium constraints. We show how to characterize the optimal partition in terms of these eigenvectors. Intuitively, we expect that the amount of information revealed about a given PIC is monotonically decreasing in its eigenvalue ranking. In line with this intuition, we derive *"only reveal the essential"* principle. When relevance of some PICs for welfare (as measured by the corresponding eigenvalues) is sufficiently low, it is optimal not to reveal any information about them at all.

These results point to the important role of central bank communication as an information filtering device, and explain why financial markets pay considerable attention to central bank announcements. In our model, the central bank's optimal communication is akin to a clustering algorithm, whereby the policymaker groups similar pieces of private information together into "clusters" and then assigns a common message to all pieces of information that fall in that cluster. In this regard, the actions of the central bank bear some resemblance to those of a data compression algorithm as is common in the machine learning domain.

An important consequence of our findings is that randomization is never optimal, despite being a common property of information design in the literature (Kamenica and Gentzkow (2011), Bergemann and Morris (2017)). Indeed, conditional on a particular realization of the state, the policymaker should always reveal the same type of information (the same partition) about the state. With randomization, instead, upon observing the state, the policymaker selects a signal value randomly, thus effectively injecting noise into their communication. Consequently, market participants confront different signals from the Fed in otherwise identical economic conditions.

Related literature. Our paper is related to the growing literature on central bank communication (see Blinder et al. (2008) for an excellent survey). The focus of investors and the media suggests that the Fed exerts significant power over financial markets and, by extension, the economy. The standard explanations for this impact comprise the conventional short-rate channel, the information channel, and the risk-premium/beliefs channel, but their relative importance remains debated. Yet, neither standard monetary shocks nor Fed's information advantage seem sufficient to justify the Fed's ability to affect markets. Measured monetary policy surprises are small and transitory (Kuttner, 2001). It is also unclear whether the Fed has superior information about the economy and better forecasting ability (Romer and Romer (2000), Faust et al. (2004)). Growing evidence suggests that the Fed directly affects risk premia in financial markets (Ai and Bansal (2018), Cieslak et al. (2019), Cieslak and Schrimpf (2019), Hanson and Stein (2015), Gertler and Karadi (2015), Kroencke et al. (2019)). This fact points to a crucial role of the Fed's communication for affecting the public beliefs which then translate into actions that can have real economic effects. In our model, we take the view of the Fed being a persuasive communicator able to convince the public to certain actions that affect welfare. Importantly, we do not require the Fed to possess any superior information about economic fundamentals, but merely require that it has some private information that the public would find valuable to know.

A related strand of research investigates the impact of announcements on belief formation.

Evidence suggests that people react to economic announcements even if that information is public knowledge (Coibion et al. (2018), D'Acunto et al. (2019a), D'Acunto et al. (2019b)). A key insight from this literature is that it is possible to manage expectations of households and firms. Central bankers care a lot about expectations management and communication is arguably the main device serving that goal. Recent experience suggests for instance that forward guidance can be a powerful tool to steer market expectations and thereby affect decisions when interest rates are at the effective lower bound (Coeuré (2017), Haldane (2017)). However, as highlighted by Blinder et al. (2008), the large variation in communication strategies across central banks suggests that the optimal design of communication remains an open question.

The problem of optimal policy announcement design is directly related to the classical problem of optimal communication with commitment, also known as "Bayesian persuasion" after Kamenica and Gentzkow (2011).[3] In the context of monetary announcements, communication with commitment is closely related to the so-called Odyssean forward guidance (Campbell et al. (2012)), whereby a policymaker commits to a state-contingent policy that maps outcomes to announcements. Morris and Shin (2018) show how such announcements should be designed in the presence of the so-called reflection problem produced by the two-way flow between the market participants and the central bank.[4] They study a model where market participants' actions depend on monetary policy, whereas monetary policy reacts to market participants' actions. Our model also features a two-way interaction between market participants and the policymaker, with the policymaker "coordinating the beliefs" of the public in the spirit of Morris and Shin (2002), yet the mechanism is based on the Bayesian

[3]See also Aumann and Maschler (1995), Calzolari and Pavan (2006), Ostrovsky and Schwarz (2010), and Rayo and Segal (2010)

[4]The reflection problem had been discussed and investigated in a large literature, e.g., Goodhart (1975), Samuelson (1994), Woodford (1994), Svensson and Woodford (2004), Bond and Goldstein (2015), Goldstein and Yang (2017).

belief updating. The policymaker conditions its announcements on the (prior) beliefs of the public, whereas the (posterior) beliefs of the public are influenced by the announcements.[5]

A key technical assumption we make in our model is that the set of communication instruments available to a policymaker is discrete. Indeed, in practice, there is a large but finite number of words and sentences which the central bank can use to formulate announcements. Numerical values are also discrete; for example, policy rates are often adjusted by a quarter or half of a percentage point. The discreteness in central bank communication shares analogy with a rating agency that evaluates debtors' creditworthiness, and announces this assessment publicly by placing them into discrete rating buckets (e.g., above BBB-).

In a similar way, central bank communication in our model amounts to announcing a discrete "rating" for the state of the economy along with the appropriate policy stance associated with it. Intuitively, one expects that it should be optimal for the information designer to "pool" neighbouring states into clusters. This intuitive result has been established in several recent papers under the assumption that the state is one-dimensional (Kolotilin (2018), Hopenhayn and Saeedi (2019), Dworczak and Martini (2019)). Yet, little is known about the optimal communication design in the multi-dimensional case.[6] Remarkably, we show that, under mild technical conditions, the partition result continues to hold true. Additionally, we are able to characterize some universal properties of these partitions through the principal information components.

[5] In earlier work, Amato and Shin (2002), Morris and Shin (2002) study how releases of public information affect welfare in a static setting. Morris and Shin (2005) and Amador and Weill (2012) discuss dynamic effects of public information revelation.

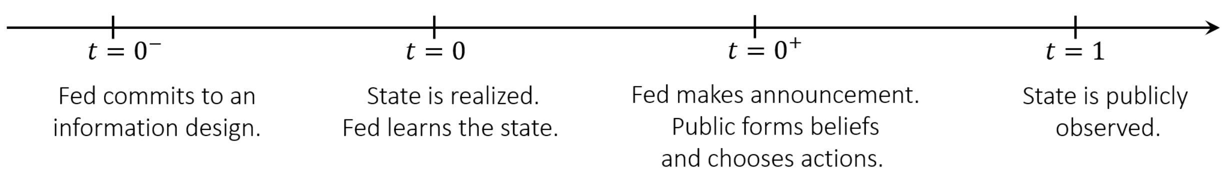
[6] The only result we are aware of is due to Dworczak and Martini (2019), who show that, under some technical conditions, with a two-dimensional state, four possible actions, and sender's utility that only depends on the first two moments of the distribution of state, the optimal communication design is given by a partition into four convex polygons.

2 The Model

There are four time periods, $t = 0-,\ 0,\ 0+,\ 1$. The policymaker believes that the state ω (the private information of the policymaker) is a random vector on $\Omega \subset \mathbb{R}^L$, a bounded, open and convex subset of $\mathbb{R}^L$ equipped with the Borel sigma-algebra $\mathcal{B}$, distributed with a density $\mu_0(\omega)$.

Following Kamenica and Gentzkow (2011), we assume that the policymaker is able to *commit* to an information design at time $t = 0-$, before the state ω is realized.[7] Full commitment means that the policymaker is completely transparent about the exact structure of the map from its private information to policy announcements. In particular, we abstract from issues related to imperfect commitment and reputation building by the policymaker. The policymaker learns the realization of the state ω at time $t = 0$, while the public only learns it at time $t = 1$. The vector ω represents the full set of private information of the policymaker. The policymaker's objective is then to decide how much, and what kind of, information about ω to reveal to the public through policy announcements.

The following graph illustrates the timing of the events in the model.



Formally, the policymaker needs to solve the problem of optimal announcement (information) design. The intuition is as follows. After the policymaker (for short, the Fed)

[7]It is important to distinguish the optimal information design problem from the problem of optimal signalling: In the latter, the policymaker cannot commit to an information structure, and the public needs to rationally guess it within a signalling game. By contrast, in the former, due to the assumed full commitment, guessing is not necessary. As a result, optimal information design is free from all the standard problems with signalling games (such as, for example, a huge multiplicity of equilibria). As Morris and Shin (2018) argue, such commitment policies are well suited for central bank communication.

receives the private signal ω, they choose how to optimally communicate the information to the public in a way that maximizes welfare. The menu of possible announcements that the Fed could send is finite. Given that the economy is populated by rational optimizing agents, when deciding on the optimal message to send, the Fed takes into account the action of the public that the announcement will induce. Conditional on the announcement, agents in the economy use the Bayes rule to form a posterior belief about the state, and then select the optimal action that maximizes their individual utility. The Fed's problem is therefore to design communication that is optimal from the welfare perspective while taking into account the rational reaction of the public that the Fed's announcement will induce. We now lay out the model details.

2.1 Policy announcements

We first define the information design that the policymaker commits to, and the basic structure of the policymaker's announcements.

Definition 1 (Information design) *An information design is a probability space $\mathcal{K}$ (henceforth, signal space) and a probability measure $\mathcal{P}$ on $\mathcal{K}\times\Omega$. An information design is K-finite if the signal space $\mathcal{K}$ has exactly K elements: $|\mathcal{K}| = K$. In this case, we can identify $\mathcal{K}$ with the set $\{1,\cdots,K\}$. An information design is finite if it is K-finite for some $K\in\mathbb{N}$. Without loss of generality, we assume that $\mathcal{K} = \{1,\cdots,K\}$.*

Once the public observes a policymaker announcement $k\in\{1,\cdots,K\}$, it updates the beliefs about the probability distribution of ω using the Bayes rule. To do this, the public just needs to know the probability $\pi_k(\omega)$ of ω given the announcement k :

$$\pi_k(\omega) \equiv \mathcal{P}(k|\omega)\,.$$

As such, a K-finite information design can be equivalently characterized by a set of measurable functions $\pi_k(\omega)$, $k = \{1, \cdots, K\}$ satisfying these conditions $\pi_k(\omega) \in [0,1]$ and $\sum_k \pi_k(\omega) = 1$ with probability one.

Intuitively, an announcement design is a map from the space Ω of possible states to a "dictionary" of K messages,[8] whereby the policymaker commits to a precise rule of selecting an announcement from the dictionary for every realization of ω. In principle, it is possible that this rule involves randomization, whereby, for given ω, the policymaker randomly picks an announcement from a non-singleton subset of messages in the dictionary. Clearly, an information design does not involve randomization if and only if it is a partition of the state space Ω. The policymaker divides Ω into K subsets and makes the announcement k if and only if ω belongs to the subset number k of the partition. The public knows that given state ω, the Fed would send the k-th signal with probability 1, and signal $m \neq k$ with probability zero. We formalize this discussion in the following definition.

Definition 2 (Randomization) *We say that information design involves randomization if $\pi_k(\omega) \notin \{0,1\}$ with positive probability for some k. We say that information design is a partition if $\pi_k(\omega) \in \{0,1\}$ with probability one for any k. In this case,*

$$\tilde{\Omega} \; = \; \cup_{k=1}^{K} \{\omega : \; \pi_k(\omega) = 1\} \tag{1}$$

is a Lebesgue-almost sure partition of Ω in the sense that $\Omega \setminus \tilde{\Omega}$ has Lebesgue measure zero, and the subsets of the partition (1) are Lebesgue-almost surely disjoint.

We use $\bar{\pi} \; = \; (\pi_k(\omega))_{k=1}^{K} \in [0,1]^K$ to denote the random K-dimensional vector represent-

[8]Note that, in practice, many of these messages are related to *future actions* of the policymaker. We do not assume that a message always implies a full commitment of the policymaker to implement the promised action. The only assumption we make is that the public uses the Bayes rule to update its probabilistic beliefs about the likelihood of the promised action.

ing the information design, and we use

$$\Pi \;=\; \{(\pi_k(\omega))_{k=1}^K :\; \pi_k(\omega) \geq 0,\; \sum_k \pi_k(\omega) = 1\}$$

to denote the set of all possible information designs, equipped with the metric:

$$\|\bar{\pi}^1 - \bar{\pi}^2\| = \max_k \int_\Omega |\pi_k^1(\omega) - \pi_k^2(\omega)| d\omega\,. \tag{2}$$

In this paper, we only deal with finite information designs. We believe that this is a realistic assumption as, even for central bank communication, the set of signals that the central bank can send to the public is discrete and finite, albeit it can be quite large. A key implication of this setting is that, with a continuous state space and under appropriate regularity conditions, randomization is never optimal and, hence, optimal information design is always given by a *partition*. Importantly, this implication is in contrast with the optimal information design that assumes a discrete state space and in which randomization is typically present (e.g., Kamenica and Gentzkow (2011) and Bergemann and Morris (2017)). To the extent that randomization produces artificial ambiguity in communication, it is a counter-intuitive property. Indeed, a randomized announcement design amounts to an injection of noise into the public information set: upon observing a realization of a state ω, the policymaker "tosses a coin" and then randomly picks a signal value. As such, market participants might repeatedly observe different messages under otherwise identical economic conditions.

2.2 Agents

We assume that the economy is populated by N classes of agents, indexed by $n = 1, \cdots, N$. Each class n may consist of a continuum of agents or a single, large agent. In the former case, we assume that all agents within each class are identical and take identical actions. Furthermore, we allow for the possibility that agents in each class have private information or simply differ in their prior beliefs. Namely, each class n has a class specific prior with a Radon-Nykodym density $\mu_0^n(\omega),\ n = 1, \cdots, N$ with respect to the central bank prior $\mu_0(\omega)$. At time $t = 0$, upon observing the public policy signal k, each agent of class n selects an action a_n from the action space $\mathcal{R} \subset \mathbb{R}^m$ to maximize the expected utility function

$$E^n[U^n(a_n, a, \omega)|k]\,,$$

where we use $a = \{a_n\}_{n=1}^N \in \mathbb{R}^{Nm}$ to denote the vector of actions of all agents' classes. Thus, we allow each agent's utility to depend on the actions of other agents in the economy. These actions could for instance represent consumption, investment, production, or price setting choices by market participants, consumers, or firms. Upon observing policymaker's signal k, agents of class n update their prior $\mu_0^n(\omega)\mu_0(\omega)$ using the Bayes rule to the class-specific posterior distribution

$$\frac{\pi_k(\omega)\mu_0^n(\omega)\mu_0(\omega)}{\int \pi_k(\omega)\mu_0^n(\omega)\mu_0(\omega)d\omega}\,.$$

As a result, their expected utility conditional on observing k is given by

$$E^n[U^n(a_n, a, \omega)|k] \;=\; \frac{\int \pi_k(\omega)\mu_0^n(\omega)\mu_0(\omega)U^n(a_n, a, \omega)d\omega}{\int \pi_k(\omega)\mu_0^n(\omega)\mu_0(\omega)d\omega}\,,$$

and hence, instead of assuming heterogeneous priors, we can redefine $\tilde{U}^n(a_n, a, \omega) \equiv \mu_0^n(\omega)U^n(a_n, a, \omega)$ and assume, without loss of generality, that all agents have a common prior $\mu_0(\omega)$, coinciding with that of the policymaker.

Expected utility depends on agents' own actions as well as on the vector a of actions of other agents and the aggregate state of world ω. Thus, an equilibrium $a(k) = \{a_n(k)\}_{n=1}^N$ is a solution to the fixed point system

$$a_n(k) \;=\; \arg\max_{a_n} E[U^n(a_n, a(k), \omega)|k]\,. \tag{3}$$

Given a bounded open set Ω, will will use $C^2(\bar{\Omega})$ to denote the set of functions that are continuously differentiable in Ω and whose first and second derivatives are continuous on the close $\bar{\Omega}$ of Ω. We equip $C^2(\bar{\Omega})$ with the standard norm $\|f\| = \max_{\bar{\Omega}}(|f(\omega)| + |\nabla f(\omega)| + |\nabla^2 f(\omega)|)$, where ∇ and ∇^2 denote the vectors of all first and second order derivatives of f, respectively.

Assumption 1 (Action space) *The action space $\mathcal{R}$ is an open subset of $\mathbb{R}^m$. The function $U^n(a_n, a, \omega) \in C^2(\bar{\mathcal{R}} \times \bar{\mathcal{R}}^N \times \bar{\Omega})$, where we have defined*[9]

$$\bar{\mathcal{R}}^N \;=\; \underbrace{\bar{\mathcal{R}} \times \cdots \times \bar{\mathcal{R}}}_{N \ times}\,.$$

While the system (3) may potentially have multiple solutions, we assume that the policymaker "believes" in a particular equilibrium structure[10] $a^* :\ \Pi \to \mathcal{R}^{Nm}$ that maps any given information design $\bar{\pi}$ into an equilibrium $a^*(\bar{\pi}) \;=\; (a^*(k, \bar{\pi}))_{k \in \{1, \cdots, K\}}\,.$

Assumption 2 *The equilibrium structure a^* is continuous.*

It will be helpful to rewrite the system (3) in form of agents' first-order conditions by applying several algebraic transformations. If there exist functions $G_n :\ \mathcal{R}^N \times \bar{\Omega} \to \mathbb{R}^m,\ n = 1, \cdots, N,$ such that the set of solutions to

$$E[G_n(a(k), \omega)|k] \;=\; 0,\ n \;=\; 1, \cdots, N \tag{4}$$

coincides with the set of equilibria, then we refer to such a system as an equivalent system. We often use the notation $G \;=\; (G_n)_{n=1}^N \;:\; \mathbb{R}^{Nm} \to \mathbb{R}^{Nm}$.

[9] $\bar{\mathcal{R}}$ denotes the closure of $\mathcal{R}$.

[10] It is straightforward to extend our analysis to the case when the policymaker believes in multiple equilibria occurring with some probabilities.

2.3 The policymaker's problem

In our perfect, rational model, for any information design $\{1, \cdots, K\}$, the policymaker exactly knows the vector of actions $a(k)$ of the public in response to any given signal. Thus, effectively, the policymaker induces an action $a(k)$ by sending the signal k. In particular, without loss of generality, we may assume that at the optimum we always have $a(k) \neq a(\tilde{k})$ for any $k \neq \tilde{k}$. That is, different signals always induce different actions, as otherwise, out of two signals k and $\tilde{k}$ that induce identical actions, one would be redundant. Formally, if $a(k) = a(\tilde{k})$, then equation (4) implies that $E[G_n(a(k), \omega)|k] = E[G_n(a(\tilde{k}), \omega)|\tilde{k}]$ and hence

$$E[G_n(a(k), \omega)|\{k \ or \ \tilde{k}\}] \ = \ P(k)E[G_n(a(k), \omega)|k] \ + \ P(\tilde{k})E[G_n(a(\tilde{k}), \omega)|\tilde{k}] \ = \ 0\,,$$

where we have defined

$$P(k) \ = \ Prob[k] \ = \ \int_\Omega \pi_k(\omega)\mu_0(\omega)d\omega$$

to be the unconditional probability of the policymaker sending signal k. Thus, a modified information structure where k is announced whenever either k or $\tilde{k}$ were to be announced in the old structure leads to the same equilibrium actions and, hence, also to the same expected social welfare.

Definition 3 *Given an information structure, a regular equilibrium is a solution to an equivalent system* (4) *such that the Jacobian* $\left(E\left[\frac{\partial}{\partial a_i} G_n(a(k), \omega)|k\right]\right)_{n,i=1}^{N}$ *is non-degenerate and* $a(k) \neq a(\tilde{k})$ *for any* $k \neq \tilde{k}$.

Given their beliefs in the equilibrium structure a^*, the policymaker selects an information design $\bar{\pi}^* = (\pi_k^*(\omega))$ to maximize the expected public welfare function $W(\omega, a)$:

$$\bar{\pi}^* \;=\; \arg\max_{\bar{\pi}} E[W(\omega, a^*(\bar{\pi}))] \;=\; \arg\max_{\bar{\pi}}\{E[W(\omega, a)] : \; a \;=\; \text{maximizes agents' utilities}\}$$
$$=\; \max_{\bar{\pi},\; a}\{E[W(\omega, a(k))] \;:\; E[G(a(k), \omega)|k] \;=\; 0\; \forall\; k\}\,. \tag{5}$$

By direct calculation, we can rewrite expected social welfare function as

$$E[W(\omega, a^*(\bar{\pi}))] \;=\; \sum_{k=1}^{K} \int W(\omega, a^*(k, \bar{\pi}))\, \pi_k(\omega)\mu_0(\omega) d\omega\,. \tag{6}$$

We say that $\bar{\pi}^*$ is regular with respect to an equivalent system (4) if the Jacobian of (4) with respect to a is non-degenerate at $a^*(\bar{\pi}^*)$. One class of regular equilibria corresponds to the case when the dependence of a on ω is explicit, i.e., when $a(k) = E[g(\omega)|k]$. Formally, this is equivalent to specifying $G(a, \omega) = a - g(\omega)$, so that $E[G(a(k), \omega)|k] = 0$ is equivalent to $a(k) = E[g(\omega)|k]$. The following result is just a restatement of this observation.

Lemma 4 *Suppose that $G(a, \omega) \;=\; a \;-\; g(\omega)$ for some real analytic function $g(\omega) : \;\Omega \;\to\; \mathcal{R}^{Nm}$. Then, there exists a unique equilibrium and this equilibrium is regular.*

To state the main result—the optimal information design—we need the following definition and technical conditions.

Definition 5 *We say that functions $\{f_1(\omega), \cdots, f_{L_1}(\omega)\},\; \omega \in \Omega,$ are linearly independent modulo $\{g_1(\omega), \cdots, g_{L_2}(\omega)\}$ if there exist no real vectors $h \in \mathbb{R}^{L_1},\; k \in \mathbb{R}^{L_2}$ with $\|h\| \neq 0$, such that*

$$\sum_i h_i f_i(\omega) \;=\; \sum_j k_j g_j(\omega) \qquad for\ all\ \omega \in \omega\,.$$

In particular, if $L_1 = 1$*, then* $f_1(\omega)$ *is linearly independent modulo* $\{g_1(\omega), \cdots, g_{L_2}(\omega)\}$ *if* $f_1(\omega)$ *cannot be expressed as a linear combination of* $\{g_1(\omega), \cdots, g_{L_2}(\omega)\}$.

We also need the following technical assumption.

Assumption 3 *The welfare function* $W(a, \omega)$ *is real analytic*[11] *in* $\omega \in \Omega$ *for any* $a \in \mathbb{R}^{Nm}$*. Furthermore, there exists an equivalent system* (4)*, such that* $G_n(a, \omega),\ n = 1, \cdots, N$*, is real analytic in* ω *for all* $a \in \mathcal{R}^N$ *and*

- *For any fixed* $a, \tilde{a} \in \mathcal{R}^N,\ a \neq \tilde{a}$*, the function* $W(\omega, a) - W(\omega, \tilde{a})$ *is linearly independent modulo* $\big\{\{G_n(a, \omega)\}_{n=1}^N, \{G_n(\tilde{a}, \omega)\}_{n=1}^N\big\}$*;*

Assumption 3 holds for generic real-analytic functions W and G. The real analyticity conditions are also not restrictive. In fact, almost all utility functions used in economic modelling are real analytic. Real analyticity is known to impose an important regularity structure on equilibria (e.g., Hugonnier, Malamud and Trubowitz (2012)). The key property of real analytic functions that we use in our analysis is stated in the following proposition.

Proposition 6 *If a real analytic function* $f(\omega)$ *is zero on a set of positive Lebesgue measure, then* f *is identically zero. Hence, if real analytic functions* $\{f_1(\omega), \cdots, f_{L_1}(\omega)\}$ *are linearly dependent modulo* $\{g_1(\omega), \cdots, g_{L_2}(\omega)\}$ *on some subset* $I \subset \Omega$ *of positive Lebesgue measure, then this linear dependence also holds on the whole* Ω *except, possibly, a set of Lebesgue measure zero.*

Using Proposition 6, it is possible to prove the main result of this paper:

Theorem 7 (Optimal information design) *Suppose that the equilibrium* $a^*(\pi^*)$ *is regular. Then, there exists an optimal information design* $\bar{\pi}^*$ *and it is a partition.*

[11] A function $f(\omega)$ is called real analytic in ω if it can be represented as a convergent power series in ω a small neighbourhood of any point in its domain Ω.

The proof of Theorem 7 is technical and is delegated to the appendix. However, the intuition is as follows. Suppose on the contrary that $\bar{\pi}$ is not a partition. Then, for some k, we have $\pi_k(\omega) \in (0,1)$ on some positive measure subset $I \subset \Omega$. At the global maximum, under arbitrary small perturbations social welfare should decrease. We show that this can only be true if Assumption 3 is violated for $\omega \in I$. However, since I has a positive Lebesgue measure, Proposition 6 implies that it has to be violated on the whole of Ω.

2.4 Distilling Information into Optimal Clusters

By Theorem 7, the optimal way for a policymaker to distill information that they communicate to the public is by partitioning the state space into clusters and then revealing the cluster to which the state belongs. Thus, it is crucial to understand the properties (shapes) of those clusters. The goal of this section is to provide a general characterization of the "optimal clusters."

We use $DG(a,\omega) \in \mathbb{R}^{(Nm)\times(Nm)}$ to denote the Jacobian of the map G, and, similarly, $DW(a,\omega) \in \mathbb{R}^{1\times(Nm)}$ the gradient of the welfare function $W(a,\omega)$ with respect to a. For any vectors $x_k \in \mathbb{R}^{Nm}$, $k = 1, \cdots, K$ and actions $\{a(k)\}_{k=1}^K$, let us define the partition

$$\begin{aligned} \Omega_k^*(\{x_\ell\}_{\ell=1}^K, \{a_\ell\}_{\ell=1}^K) \; = \; \Big\{ & \omega \;\in\; \Omega \;:\; W(a(k),\ \omega) - x_k^\top G(a(k),\ \omega) \\ & = \max_{1\le l\le K} \big(W(a(l),\ \omega) \;-\; x_l^\top G(a(l),\ \omega)\big) \Big\} \end{aligned} \tag{7}$$

Equation (7) is basically the first-order condition for the optimization problem (5), whereby x_k are the Lagrange multipliers of agents' participation constrains.

Theorem 8 *For any regular equilibrium $\{a(k)\}_{k=1}^K$ there exists a welfare-equivalent partition that is characterized as follows:*

- *local optimality holds:*

$$\Omega_k \;=\; \Omega_k^*(\{x_\ell\}_{\ell=1}^K, \{a_\ell\}_{\ell=1}^K)$$

with

$$x_k \;=\; \bar{D}W(k)(\bar{D}G(k))^{-1}\ ,\ k = 1, \cdots, K\,,$$

where we have defined

$$\begin{aligned}
\bar{D}W(k) \;&=\; \int_{\Omega_k} DW(a(k),\omega)\mu_0(\omega)d\omega\,,\\
\bar{D}G(k) \;&=\; \int_{\Omega_k} DG(a(k),\omega)\mu_0(\omega)d\omega\ ,\ k = 1, \cdots, K\,.
\end{aligned}$$

- *the actions $\{a(k)\}_{k=1}^K$ satisfy the fixed point system*

$$\int_{\Omega_k} G(a(k),\omega)\mu_0(\omega)d\omega \;=\; 0,\ k = 1,\ \cdots,\ K\,. \tag{8}$$

- *the boundaries of Ω_k are a subset of the real analytic variety*[12]

$$\cup_{k\neq l}\left\{\omega\in\mathbb{R}^m\ :\ W(a(k),\ \omega)\ -\ x_k^\top G(a(k),\ \omega)\ =\ W(a(l),\ \omega)\ -\ x_l^\top G(a(l),\ \omega)\right\}\,. \tag{9}$$

and hence have a Lebesgue measure of zero.

Equations (8) and (9) in Theorem 8 are re-statements of agents' first order conditions

[12] A real analytic variety in $\mathbb{R}^m$ is a subset of $\mathbb{R}^m$ defined by a set of identities $f_i(\omega) = 0,\ i = 1, \cdots, I$ where all functions f_i are real analytic. If at least one of functions $f_i(\omega)$ is non-zero, then a real analytic variety is always a union of smooth manifolds and hence has a Lebesgue measure of zero. By Assumption 3, the variety $\left\{\omega\in\mathbb{R}^m\ :\ W(a(k),\ \omega)\ -\ x_k^\top G(a(k),\ \omega)\ =\ W(a(l),\ \omega)\ -\ x_l^\top G(a(l),\ \omega)\right\}$ has Lebesgue measure zero for each $k\neq l$.

(4). The key insight of Theorem 8 comes from the characterization of the boundaries of the different clusters. The policymaker has to solve the problem of maximizing social welfare (6) by inducing the desired actions vector (a_n) for every realization of ω. Ideally, the policymaker would like to solve the problem

$$a^* \ = \ \arg\max_a W(\omega, a)\,.$$

However, the ability of the policymaker to induce desired actions is limited by the participation constraints of the public, that is, the map from the posterior beliefs induced by communication to the actions of the public. Indeed, while the policymaker can induce any Bayes-rational beliefs (i.e., any posteriors consistent with the Bayes rule, see, Kamenica and Gentzkow (2011)), she has no direct control over the actions of the public. The degree to which these constraints are binding is precisely captured by the Lagrange multipliers x_k, so that the policymaker is maximizing the Lagrangian $\max_a(W(\omega,a)-x_k^\top G(a,\omega))$. Formula (7) shows that, inside the cluster number k, the optimal action profile maximizes the respective Lagrangian. The boundaries of the clusters are then determined by the indifference conditions (9), ensuring that at the boundary between regions k and l the policymaker is indifferent between the respective action profiles a_k and a_l.

Several papers study the one dimensional case (that is, when $L = 1$ so that $\omega \in \mathbb{R}^1$) and derive conditions under which the optimal signal structure is a monotone partition into intervals. Such a monotonicity result is intuitive as one would expect that optimal information design only pools nearby states. The most general results currently available are due to Hopenhayn and Saeedi (2019) and Dworczak and Martini (2019),[13] but they cover the case when sender's utility (social welfare function in our setting) only depends on $E[\omega] \in \mathbb{R}^1$.[14] Under this assumption, Dworczak and Martini (2019) derive necessary and sufficient

[13]See also Mensch (2018).

[14]This is equivalent to $G(a,\omega) \ = \ a - g(\omega)$ in our setting, see Lemma 4.

conditions guaranteeing that the optimal signal structure is always a monotone partition of the interval Ω into a union of disjoint intervals. Theorem 8 shows that, under mild additional regularity conditions, optimal information design is in fact always a partition.[15]

2.5 The Information Relevance Matrix

The structure of the optimal partition (Theorem 8) can be complex and non-linear.[16] Therefore, it is useful to think about ways of simplifying. One may ask whether it is possible to "linearize" these partitions, just like one can linearize equilibria in complex, non-linear economic models, assuming the deviations from the steady state are small. As we show below, this is indeed possible.

Everywhere in the sequel, we make the following assumption.

Assumption 4 *There exists a small parameter ε such that the functions defining the equilibrium conditions, G, and the welfare function, W, are given by $G(a, \varepsilon\omega)$ and $W(a, \varepsilon\omega)$.*

Parameter ε has two interpretations. First, it could mean small deviations from a steady state (as is common in the literature on log-linear approximations). Second, ε could be interpreted as capturing the sensitivity of economic quantities to changes in ω. In the context of policy communication, one could think about the policymaker trying to stabilize the economy by steering public expectations towards its desired equilibrium.

In the limit when $\varepsilon = 0$, equilibrium does not depend on shocks to ω. We use a^0 to denote this "steady state" equilibrium. By definition, it is given by the unique solution to the system $G(a^0, 0) = 0$, and the corresponding social welfare is $W(a^0, 0)$. We assume that there exists a globally unique solution a^0 to $G(a^0, 0) = 0$ and $DG(a^0, 0)$ is non-degenerate.

[15] Of course, as Dworczak and Martini (2019) explain, even in the one-dimensional case the monotonicity cannot be ensured without additional technical conditions. No such conditions are known in the multi-dimensional case. Dworczak and Martini (2019) present an example with four possible actions ($K = 4$) and a two-dimensional state space ($L = 2$) for which they are able to show that the optimal information design is a partition into four convex polygons.

[16] In general, the boundaries of the sets Ω_k might be represented by complicated hyper-surfaces, and some of Ω_k might even feature multiple disconnected components.

The following matrices are important in our characterization of the optimal information design:

- $DG(a^0, 0) \in \mathbb{R}^{(Nm)\times(Nm)}$ is the Jacobian of the map $G: \mathbb{R}^{Nm} \to \mathbb{R}^{Nm}$ w.r.t. a;
- $G_\omega(a^0, 0) = (\partial G_n/\partial \omega_i) \in \mathbb{R}^{(Nm)\times L}$ is the gradient of the map G with respect to ω;
- $DW(a^0, 0) \in \mathbb{R}^{1\times(Nm)}$ is the gradient of the welfare function with respect to a;
- $DW_\omega(a^0, 0) \in \mathbb{R}^{(Nm)\times L}$ is the matrix of second order partial derivatives $(\partial^2 W/(\partial a_n \partial \omega_i))$;
- $D^2 W(a^0, 0) \in \mathbb{R}^{(Nm)\times(Nm)}$ is the matrix of second order partial derivatives $(\partial^2 W/(\partial a_n\, \partial a_m))$;
- $\widetilde{D}G_\omega(a^0, 0) \in \mathbb{R}^{(Nm)\times L}$ is given by

$$(\widetilde{D}G_\omega(a^0, 0))_{i,j} = \sum_n (DW(a^0, 0) DG(a^0, 0)^{-1})_n \frac{\partial^2 G_n(a^0, 0)}{\partial a_i \partial \omega_j};$$

- $\widetilde{D}^2 G(a^0, 0) \in \mathbb{R}^{(Nm)\times(Nm)}$ is given by

$$(\widetilde{D}^2 G(a^0, 0))_{i,j} = \sum_n (DW(a^0, 0) DG(a^0, 0)^{-1})_n \frac{\partial^2 G_n(a^0, 0)}{\partial a_i \partial a_j}.$$

When the policy-maker sends signal k, the public realizes that $\omega \in \Omega_k$. As a result, the public posterior estimate of the conditional mean of ω is then given by

$$M_1(\Omega_k) \equiv E[\omega|\omega \in \Omega_k] = \frac{\int_{\Omega_k} \omega \mu_0(\omega) d\omega}{P_k} \in \mathbb{R}^m,$$

where

$$P_k = \text{Prob}[k] = \int_{\Omega_k} \mu_0(\omega) d\omega.$$

In our linearized approximation, the policymaker only cares about the first moments of the random variables in ω. When deviations from the steady state are small, it is enough for the policymaker to send signal k that induces the public to update their beliefs about $E[\omega|k]$ towards the "optimal" level. Higher order moments, such as $E[\omega^2|k]$ become negligible. As a result, similar to the standard Taylor rule that is derived using a log-linear approximation, policy can be formulated exclusively in terms of the first moments of the random variables in question. However, interestingly enough, the shape of optimal linearized clusters (see Theorem 8) depends on the interactions between different $\omega_i,\ i = 1, \cdots, m$ in a non-trivial way.

Define

$$\mathcal{G} \;\equiv\; (DG(a^0, 0))^{-1} G_\omega(a^0, 0) \;\in \mathbb{R}^{(Nm)\times L}\,.$$

The following lemma follows by direct calculation.

Lemma 9 *For any sequence $\varepsilon_\nu \to 0,\ \nu \in \mathbb{Z}_+$, there exists a sub-sequence $\varepsilon_{\nu_j},\ j > 0$, such that the optimal partitions $\{\Omega_k(\varepsilon)\}_{k=1}^K$ converge to a limiting partition $\{\Omega_k(0)\}_{k=1}^K$. In this limit,*

$$a_k(\varepsilon_{\nu_j}) \;=\; a_k^0 \;-\; \varepsilon_{\nu_j}\mathcal{G}\, M_1(\Omega_k(0)) \;+\; o(\varepsilon_{\nu_j})\,.$$

Lemma 9 provides an intuitive explanation for the role of the matrix $\mathcal{G}$. Namely, in the linear approximation, the public action is given by a linear transformation of $E[\omega|k]$, the first moment of ω given the policy announcement: $a_k \;\approx\; a_k^0 \;-\; \mathcal{G}E[\omega|k]$. Thus, $-\mathcal{G}$ captures the sensitivity of actions to beliefs.

Assumption 5 (The information relevance matrix) *We assume that the $L \times L$ sym-*

metric matrix $\mathcal{D}$ given by

$$\mathcal{D} \;\equiv\; \mathcal{G}^\top \left(D^2W(a^0,0) \;-\; \widetilde{D}^2G(a^0,0)\right)\mathcal{G}$$
$$+\; \mathcal{G}^\top(\widetilde{D}G_\omega(a^0,0) - DW_\omega(a^0,0)) \;+\; (\widetilde{D}G_\omega(a^0,0) - DW_\omega(a^0,0))^\top\mathcal{G}\,.$$

is non-degenerate. We refer to $\mathcal{D}$ as the information relevance matrix.

The information relevance matrix plays a key role in our subsequent analysis. It captures the relevance of different parts of the private information vector, ω, for social welfare. The first term, $\mathcal{G}^\top D^2W(a^0,0)\,\mathcal{G}$ is the second order derivative (the Hessian) of social welfare to the action profile a, multiplied by the sensitivity $-\mathcal{G}$ of actions to beliefs. Thus, $\mathcal{G}^\top D^2W(a^0,0)\,\mathcal{G}$ captures the second order sensitivity of welfare to beliefs. Similarly, the terms $-\mathcal{G}^\top DW_\omega(a^0,0)$ and $-DW_\omega(a^0,0)\mathcal{G}$ are the mixed derivatives with respect to $E[\omega]$ and ω, capturing the sensitivity of welfare to simultaneous changes in ω and public expectations about ω. Finally, each of these terms is "compensated" by the counter-acting term coming from the function G. The latter determines how public posterior expectations affect agents' actual decisions. Section 2.7 explains that the definiteness of matrix $\mathcal{D}$ is closely tied to the amount and type of information being optimally revealed.

We are now ready to state the main result of this section, showing how the optimal linearized partition can be characterized explicitly in terms of the information relevance matrix $\mathcal{D}$.

Theorem 10 (Linearized partition) *Under the hypothesis of Theorem 7 and Assumptions 4 and 5, let $\{\Omega_k(\varepsilon)\}_{k=1}^K$ be the corresponding optimal partition. Then, for any sequence $\varepsilon_k \to 0,\ k > 0,$ there exists a sub-sequence $\varepsilon_{k_j},\ j > 0,$ such that the optimal partition $\{\Omega_k(\varepsilon_{k_j})\}_{k=1}^K$ converges to an almost sure partition $\{\tilde{\Omega}_k^*\}$ satisfying*

$$\tilde{\Omega}_k^* \;=\; \{\omega\in\Omega:\ (M_1(k)-M_1(l))^\top\mathcal{D}\omega \;>\; 0.5(M_1(k)^\top\,\mathcal{D}\,M_1(k)-M_1(l)^\top\,\mathcal{D}\,M_1(l))\;\forall\, k\neq l\}\,,$$

where we have defined $M_1(k) \equiv M_1(\tilde{\Omega}_k^*)$ *. In particular, for this limiting partition each set* $\tilde{\Omega}_k^*$ *is convex. If the matrix* $\mathcal{D}$ *from Assumption 5 is negative semi-definite, then all sets* Ω_k^* *are empty except for one: That is, it is optimal to reveal no information.*

To understand the result of Theorem 10, it is important to remember that we are designing an announcement mechanism under full commitment, that is, the policymaker commits to the announcement policy before ω is realized. The following two examples illustrate the mechanism.

Example 1 (Concealing all information) Suppose that the policymaker commits to an announcement policy about the interest rate, so that $\omega = r$. For simplicity, assume that public action is proportional to public expectations about ω. Mathematically, this equivalent to $G(a,\omega) = a-\omega$. This public action in turn affects important aggregate outcome variables (such as the output gap and/or inflation). We assume that the policymaker's welfare function is negative quadratic in the public action: $W(a,\omega) = -a^2$, so that the objective of the policymaker is to choose the information design $\{1,\cdots,K\}$ to maximize

$$E[W] = -E[E[a|k]^2] = -\sum_k P_k E[\omega|k]^2 \,,$$

where $P_k = E^{pm}[k]$ is given by the policymaker's expectations about the probability of signal k. Suppose first that the beliefs of the policymaker coincide with those of the public. Then, a direct application of Jensen's inequality implies that

$$\sum_k P_k E[\omega|k]^2 \geq \left(\sum_k P_k E[\omega|k]\right)^2 = E[\omega]^2 \,. \tag{10}$$

That is, it is ex-ante optimal for the policymaker *to commit not to reveal any information at all.* This result depends crucially on the commitment assumption (much like a commitment to the Odyssean forward guidance). Ex post, the policymaker would find it optimal to reveal good news and conceal bad news, but this would require breaking the commitment. The consequences of breaking commitment could be very severe. Once the public loses the belief in policymakers' commitment power, the economy ends up is a signalling game regime, whereby the public attempts to interpret every unclear communication to understand, what is it that the policymaker is trying to conceal. This makes policymaker's job incredibly difficult because it is impossible to directly influence the way the public interprets communication. This example shows that clear, transparent policy announcement design with full commitment is crucial for the ability the policymaker to communicate with the public in an efficient manner.

Example 2 (Communication with differences in beliefs) The above result is sensitive to the assumption of identical expectations. Indeed, suppose that policymaker over- or under-weights some states so that $W = w(a)f(\omega)$ for some function $f > 0$ capturing the weights and some w. Then, by direct calculation,

$$\mathcal{D} \;=\; w''(a_0) \;+\; 2w'(a_0)f'(\omega),$$

with a_0 given by the ex-ante expectations, $a_0 = E[\omega]$. In this setting, "bad" states correspond to high (low) values of ω if function w is decreasing (increasing) in a. By Theorem 10, it may be optimal to reveal some information only if $\mathcal{D} > 0$. If the policymaker significantly over-weights "bad" states (i.e., $f'(\omega)$ and $w'(a_0)$ have the same sign), this pushes $\mathcal{D}$ up and the policymaker may find it optimal to reveal some information. By contrast, if the policymaker under-weights bad states relative to the public (for example, because they views those states as being less likely), this pushes $\mathcal{D}$ down and may potentially make information revelation sub-optimal. It is interesting to draw some parallels between this example and the behaviour of the Fed during the 2007-2009 financial crisis. Anecdotal evidence suggests that the Fed might have had inside information about the actual state of the financial system, but decided not to reveal it.[17] The above theoretical result suggests that this might have been optimal if the public was in panic and significantly over-estimated the probabilities of "bad" states.

Example 3 (Communicating two-dimensional information) Consider now a two-dimensional example, where the public tries to learn about two policy-relevant variables $\omega_1,\ \omega_2$. For example, ω_1 might contain information about policy rate and its future path, while ω_2 might contain information about the size and composition of the balance sheet. Again, we assume for simplicity that the public action is a linear function of expectations: $G(a,\omega) \ = \ a-(c_1\omega_1+c_2\omega_2)$, so that $a(k)=E[c_1\omega_1+c_2\omega_2|k]$. The policymaker has $W \ = \ w(a)f(\omega_1,\omega_2)$. In this case, $\mathcal{G} \ = \ -(c_1,c_2)\in\mathbb{R}^{1\times 2}$, whereas

$$\mathcal{D} \ = \ \begin{pmatrix} c_1^2 & c_1c_2 \\ c_1c_2 & c_2^2 \end{pmatrix} w''(a_0) \ + \ w'(a_0)\begin{pmatrix} 2c_1f_{\omega_1} & c_1f_{\omega_2}+c_2f_{\omega_1} \\ c_1f_{\omega_2}+c_2f_{\omega_1} & 2c_2f_{\omega_2} \end{pmatrix}$$

If w is concave, $w''(a_0)<0$ and this discourages information revelation. Yet, if the disagreement between the public and the policymaker is sufficiently large, the policymaker may find it optimal to reveal some information. Here, interesting new effects may occur. If both $f_{\omega_1}c_1$ and $f_{\omega_2}c_2$ are sufficiently large and positive, this may make $\mathcal{D}$ positive definite just like in the previous example: Over-weighting "bad" states encourages information revelation. However, here, a non-trivial interaction effect may occur. Even if both $f_{\omega_1}c_1$ and $f_{\omega_2}c_2$ are negative, a sufficiently large $|c_1f_{\omega_2}+c_2f_{\omega_1}|$ will make $\mathcal{D}$ have at east one positive eigenvalue, making it optimal to reveal some information. The nature of this information will be different: It will depend on the relative behaviour of ω_1 and ω_2. For example, it might be optimal not to reveal any information about $\omega_1,\ \omega_2$, but only about $\omega_1-\omega_2$. We formalize this intuition below in Proposition 15.

Theorem 10 shows that the standard linearization intuition is correct when applied to optimal information design: Non-linear partitions become *piecewise-linear* in the limit of

[17] Note however that one a crisis example might correspond to a large value of ε and hence might not satisfy Assumption 5.

small shock size. Boundaries of different regions become unions of hyperplanes, and are thus much easier to understand. Another key result of Theorem 10 is convexity: Since Ω_k is defined by a set of linear inequalities, it is always convex. Thus, optimal partition consists of *convex polytopes* (intervals for $L = 1$, convex polygons for $L = 2$, and convex polyhedra for $L = 3$, etc). While convexity of the sets $\Omega_k,\ k = \{1, \cdots, K\}$, can by no means be guaranteed in Theorem 8 when shocks are large, optimal partitions are always convex for small shocks. In particular, since convex sets are always connected,[18] we obtain the result that optimal partitions always consist of connected subsets when shocks are sufficiently small. Thus, for small shocks, the intuition behind "pooling nearby states" is correct.

2.6 Optimal Information Precision

Suppose that the state space Ω is a subset of $\mathbb{R}^1$ (that is, $L = 1$). Then, since convex subsets of R^1 are intervals, we get that each Ω_k is an interval. The following proposition is a direct consequence of Theorem 10.

Proposition 11 *Under the hypotheses of Theorem 10, let $L = 1$. Then:*

- *if $\mathcal{D} < 0$, then, in the limit as $\varepsilon \to 0$, the optimal partition converges to no information revelation.*
- *if $\mathcal{D} > 0$, then, in the limit as $\varepsilon \to 0$, the optimal partition converges to a partition of $[a, b]$ by*

$$a \le \frac{x_0 + x_1}{2} \le \cdots \le \frac{x_{K-1} + x_K}{2} \ \le \ b$$

[18] Recall that a set is connected if any two points in the set can be linked through a continuous curve belonging to the set.

where $x_k,\ k = 0, \cdots, K$ satisfy the system of equations

$$\begin{aligned}
x_0 &= \frac{\int_a^{0.5(x_0+x_1)} \omega \mu_0(\omega) d\omega}{\int_a^{0.5(x_0+x_1)} \mu_0(\omega) d\omega} \\
x_k &= \frac{\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \omega \mu_0(\omega) d\omega}{\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \mu_0(\omega) d\omega}\ ,\ k = 1, \cdots, K \\
x_{K+1} &= \frac{\int_{0.5(x_K+x_{K+1})}^{b} \omega \mu_0(\omega) d\omega}{\int_{0.5(x_K+x_{K+1})}^{b} \mu_0(\omega) d\omega}
\end{aligned}$$

In general, this system may have multiple solutions, each of these solutions being a candidate optimal partition. As we now show, the nature of these solutions depends crucially on the behaviour of the prior distribution $\mu_0(\omega)$. Suppose first that $\mu_0(\omega)$ is uniform. This corresponds to a 100% diffuse prior with a constant likelihood.

Corollary 12 (Uniform Distribution) *In the case when $\mu_0(\omega)$ is uniform (fully uniformed, diffused) on $[a, b]$, the optimal partition is also uniform:*

$$x_{k+1} - x_k = x_k - x_{k-1} \quad \textit{for all} \quad k = 1, \cdots, K\,.$$

By Corollary 12, with a uniform prior, optimal information design has a constant precision: for every value of ω, the signal revealed by the policymaker always has exactly the same precision. By contrast, when the prior is not uniform, it is natural to expect that the optimal information design features a precision that depends on the behaviour of the public prior, $\mu_0(\omega)$. The following proposition explicitly characterizes the link between information precision and the prior distribution.

Proposition 13 *Suppose that $\mu_0(\omega) \in C^1[a, b]$ and $|\mu_0'(\omega)|$ is sufficiently small. Then,* *optimal information precision*[19] *is increasing whenever $\mu_0(\omega)$ is increasing, and decreasing whenever $\mu_0(\omega)$ is decreasing. Namely,*

- $x_{k+1} - x_k \ \leq \ x_k - x_{k-1}$ *when $\mu_0(\omega)$ in increasing on $\Omega_k \cup \Omega_{k+1}$;*
- $x_{k+1} - x_k \ \geq \ x_k - x_{k-1}$ *when $\mu_0(\omega)$ in decreasing on $\Omega_k \cup \Omega_{k+1}$.*

The following proposition follows from Proposition 13 and is the main result of this section.

Proposition 14 (Confirmations are precise, surprises are vague) *Suppose that $\mu_0(\omega) \in C^1[a,b]$ and $|\mu_0'(\omega)|$ is sufficiently small. If $\mu_0(\omega)$ is unimodal, with a mode at ω_*, then precision is increasing on $[a, \omega_*]$ and then decreasing on $[\omega_*, b]$, being highest when ω is close to ω_*.*

Proposition 14 shows that optimal information precision depends in an intuitive way on the prior distribution of public beliefs. If the actual value of ω observed by the policymaker is close to where the public prior beliefs are concentrated (i.e., the mode ω_* of μ_0), then it is optimal for the policymarket to provide a more precise signal about where exactly ω is. By contrast, when the actual value of ω is different from the public expectations, then it is optimal not to surprise the public too much. In this case, the policymaker sends a vague signal of the type: "maybe we are close to the old prior ω_*, but one cannot know for sure."

2.7 Principal Information Components and "Only Reveal The Essential" Principle

Theorem 10 shows that the matrix $\mathcal{D}$ completely determines the nature of optimal information design. In this subsection, we aim to gain a deeper understanding of the link between

[19] Note that we use the term "precision" to denote the length of the interval Ω_k. However, since the measures $\mu_0(\omega)$ are, by assumption, close to uniform (this is what the assumption of a sufficiently small $|\mu_0'(\omega)|$ guarantees), the length of the interval is roughly proportional to the variance of ω conditional on the event that ω takes values in that interval.

the properties of this matrix and the exact nature of "optimal clusters." Since matrix $\mathcal{D}$ is symmetric, it can be diagonalized in the basis of eigenvectors:

$$\mathcal{D} \;=\; \mathcal{V}\,\mathrm{diag}(\Lambda)\,\mathcal{V}^\top\,,$$

where V is the $L \times L$ matrix with columns being the eigenvectors of $\mathcal{D}$, and with $\Lambda = (\lambda_1, \cdots, \lambda_L)$ being the eigenvalues of $\mathcal{D}$ in the decreasing order,

$$\lambda_1 \;\geq\; \cdots \;\geq\; \lambda_L\,.$$

As we explain above, when $\mathcal{D}$ is negative semi-definite (that is, all eigenvalues $\lambda_i,\ i = 1, \cdots, L$ are non-positive), it is optimal not to reveal any information. Intuitively, eigenvectors with negative eigenvalues correspond to the directions such that the information about projections on them should not be revealed. By contrast, directions corresponding to large, positive eigenvalues are the most important ones: These are the directions of information that are optimal to be revealed. Thus, in analogy with principal components used in statistics, eigenvectors of the matrix $\mathcal{D}$ represent a decomposition of the information space into directions, ordered with respect to their importance in terms of impact on welfare. We call these eigenvectors *Principal Information Components* (PICs). Since eigenvectors (the columns of $\mathcal{V}$) form an L-dimensional orthonormal basis of $\mathbb{R}^L$, we can write any state as

$$\omega \;=\; \sum_{i=1}^{L} (\omega \cdot \mathcal{V}_i)\mathcal{V}_i\,,$$

where

$$\omega \cdot \mathcal{V}_i \;=\; \sum_{j=1}^{L} \omega_j \mathcal{V}_{j,i}$$

is the (linear) component of the state along the i-th PIC. Thus, revealing information about the L-dimensional vector $\omega = (\omega_1, \cdots, \omega_L)$ is equivalent to revealing information about the projections $\omega^{\mathcal{V}} = (\omega_i^{\mathcal{V}})_{i=1}^L \equiv (\omega \cdot \mathcal{V}_i)_{i=1}^L$. Intuitively, one might expect that the optimal precision of information revealed about $\omega_i^{\mathcal{V}}$ corresponding to large λ_i (high PICs) will be higher, while little (or, equivalently, very imprecise) information will be revealed about lower PICs, with the amount of information going to zero as λ_i converges to zero.

Defining $x_k^{\mathcal{V}} \equiv \mathcal{V}M_1(k)$, we can rewrite the system of Theorem 10 as

$$\begin{aligned} \Omega_k^{\mathcal{V}} &= \{\omega^{\mathcal{V}} \in \Omega^{\mathcal{V}} : \ (x_k^{\mathcal{V}} - x_l^{\mathcal{V}})^\top \Lambda \omega^{\mathcal{V}} \ \geq \ (x_k^{\mathcal{V}})^\top \Lambda x_k^{\mathcal{V}} - (x_l^{\mathcal{V}})^\top \Lambda x_l^{\mathcal{V}} \ \forall\, l\} \\ x_k^{\mathcal{V}} &= \int_{\Omega_k^{\mathcal{V}}} \omega^{\mathcal{V}} \mu^{\mathcal{V}}(\omega^{\mathcal{V}}) d\omega / M^{\mathcal{V}}(\Omega_k^{\mathcal{V}})\,. \end{aligned} \tag{11}$$

Let us now split the eigenvalues into two groups: $\Lambda_1 = (\lambda_1, \cdots .\lambda_{i_1})$ and $\Lambda_2 = (\lambda_{i_1+1}, \cdots, \lambda_L)$. Let also $\omega = \binom{\omega_1^{\mathcal{V}}}{\omega_2^{\mathcal{V}}}$ and $x_k = \binom{x_k^1}{x_k^2}$ be the respective splits of the $\omega^{\mathcal{V}}$ and $x_k^{\mathcal{V}}$ variables. Then, we have

$$\Omega_k^{\mathcal{V}} = \left\{ \omega^{\mathcal{V}} \in \Omega^{\mathcal{V}} : \ \sum_{i=1}^2 (x_k^i - x_l^i)^\top \Lambda_i \omega_i^{\mathcal{V}} \ \geq \ \sum_{i=1}^2 ((x_k^i)^\top \Lambda_i x_k^i - (x_l^i)^\top \Lambda_i x_l^i) \ \forall l \right\}, \tag{12}$$

Suppose that ω_1 contains no information about the expectation of ω_2, so that $E[\omega_2|\omega_1] = E[\omega_2]$ is independent of ω_1. Then, while there might be multiple solutions to the system (11), there exists a solution that reveals no information about ω_2 : Indeed, such a solution has $x_k^2 = E[\omega_2]$ independent of k; in this case, by (12), the system (11) takes the form

$$\begin{aligned} \Omega_k^{\mathcal{V}} &= \left\{ \omega^{\mathcal{V}} \in \Omega^{\mathcal{V}} : \ (x_k^1 - x_l^1)^\top \Lambda_1 \omega_1^{\mathcal{V}} \ \geq \ ((x_k^1)^\top \Lambda_1 x_k^1 - (x_l^1)^\top \Lambda_1 x_l^1) \ \forall l \right\} \\ x_k^1 &= \int_{\Omega_k^{\mathcal{V}}} \omega_1^{\mathcal{V}} \mu^{\mathcal{V}}(\omega^{\mathcal{V}}) d\omega / M^{\mathcal{V}}(\Omega_k^{\mathcal{V}})\,, \end{aligned} \tag{13}$$

where $M^{\mathcal{V}}$ is the "rotated" probability measure, $M^{\mathcal{V}}(A) = M(\mathcal{V}^{-1}A)$. As the following proposition shows, the partition (13) is indeed optimal if the relevance of information in $\omega_2^{\mathcal{Y}}$ is sufficiently small.

Proposition 15 ("Only reveal the essential" principle) *Under the hypotheses of Theorem 10, suppose that $\omega_1^{\mathcal{Y}}$ contains no information about the expectation of $\omega_2^{\mathcal{Y}}$ and vice-versa, that is,*

$$E[\omega_2^{\mathcal{Y}}|\omega_1^{\mathcal{Y}}] \;=\; E[\omega_2^{\mathcal{Y}}]\,,\; E[\omega_1^{\mathcal{Y}}|\omega_2^{\mathcal{Y}}] \;=\; E[\omega_1^{\mathcal{Y}}]\,. \tag{14}$$

Then, for a generic $\mu_0(\omega) \in \{\mu_0 \in C(\bar{\Omega}) :$ (14) holds$\}$, the gain from revealing information about $\omega_2^{\mathcal{Y}}$ is bounded from above by a constant times $\max(\Lambda_2)$. In particular, these gains vanish when $\min(\Lambda_1)/|\max(\Lambda_2)|$ is sufficiently small. In this case, the best partition solving (13) is approximately optimal.

It is important to note that the result of Proposition 15 depends in a non-trivial way on the joint distribution of $\omega_1^{\mathcal{Y}}$ and $\omega_2^{\mathcal{Y}}$ (e.g., when inflation and output are correlated): Even if revealing information about $\omega_2^{\mathcal{Y}}$ is suboptimal in isolation, the policymaker may still decide to reveal some information if it reveals important information about $\omega_1^{\mathcal{Y}}$.

The result of Proposition 15 implies an "Only Reveal The Essential" Principle in the optimal information design, whereby the relevance of information is determined by the size of the respective eigenvalues.

As an illustration of the underlying mechanisms, let us rotate coordinates to the PIC basis and assume for simplicity that $G(a,\ \omega) \;=\; a-\omega$. Then,

$$E[W] \;\approx\; \sum_k P_k \sum_i \lambda_i E[\omega_i|k]^2 \;=\; \sum_i \lambda_i \sum_k P_k E[\omega_i|k]^2, \tag{15}$$

and the Jensen inequality argument of (10) suggests that it is only optimal to reveal information about those ω_i with $\lambda_i > 0$. The following proposition formalizes this intuition.

Proposition 16 *Suppose that* $E[\omega_2^{\mathcal{Y}}|\omega_1^{\mathcal{Y}}] = E[\omega_2^{\mathcal{Y}}]$ *and* $E[\omega_1^{\mathcal{Y}}|\omega_2^{\mathcal{Y}}] = E[\omega_1^{\mathcal{Y}}]$*, so that* $\omega_1^{\mathcal{Y}}$ *and* $\omega_2^{\mathcal{Y}}$ *are mean independent. If* $\max \Lambda_2 < 0$*, then it is optimal not to reveal any information about* $\omega_2^{\mathcal{Y}}$*.*

3 Discussion

In a recent article, Cecchetti and Schoenholtz (2019) call for improvement of the Fed's communication strategies by ways of "... simplifying public statements, clarifying how policy will react to changing conditions, and highlighting uncertainty and risks." Our model describes from a normative standpoint how central banks should communicate under commitment. It is therefore informative to examine some of the actual communication choices of the Fed through the lens of this theory.

The implication of our setting that randomization is suboptimal provides a formal backing of the view expressed by many current and former policymakers. Indeed, Cecchetti and Schoenholtz (2019) write:

> (...) in conducting policy, there is one uncertainty that policymakers can and should reduce: the uncertainty they themselves create. Everyone agrees that monetary policymakers should do their best to minimize the noise that their actions add to the environment. When policy is transparent and effective, people in the economy and financial markets respond to the data, not to the policymakers.

Historically, however, the Fed's communication has not followed this prescription. The discretionary communication style of Chair Greenspan (Blinder and Reis, 2005) can be viewed as deliberately injecting noise into Fed statements, with Greenspan publicly admitting that "Since I've become a central banker, I've learned to mumble with great coherence... If I seem unduly clear to you, you must have misunderstood what I said." While in recent years the Fed has increased emphasis on clarity and consistency of communication, one could argue that randomization (albeit, most likely, unintentional) continues to characterize some of its choices. A case in point is the decision to hold a press conference after *every other*

FOMC meeting, thus potentially inducing changes in communication that are independent of macroeconomic conditions.[20] Such a design has been criticized by the Fed watchers, as documented in the recent Brookings survey by Olson and Wessel (2016), with the former president of the Minneapolis Fed, Narayana Kocherlakota, calling the Fed to hold a press conference after every FOMC meeting (Kocherlakota, 2016)—a change that the Fed ultimately implemented in June 2019.

The implication of Theorem 7, that policymakers communicate a partition of the state space, has analogy in the practice of central bank communication. For example, FOMC members' projections of the future path of the policy rate are communicated in form of ranges as opposed to point estimates. Similarly, projections for key macroeconomic variables such as output and inflation are published with confidence intervals. Our theoretical result that "confirmations are precise, surprises are vague" implies that the amount of uncertainty conveyed by the Fed, intuitively described by the "size" of the partition, should be related to the differences in the beliefs between the Fed and the public. From a perspective of communication efficacy, it may therefore be beneficial for the Fed to disclose the amount of uncertainty that they face as an integral part of the statement about the policy decision itself. In its current practice, however, the Fed releases uncertainty indicators to its quarterly forecasts as part of the FOMC minutes about three weeks following the policy decision announcements.

The convexity property of the optimal partition established in Theorem 10 has implications for whether it is advisable for the Fed to rely on the so-called "dot plots"—one of the more controversial devices used in recent years. Effectively, the dot plots serve to communicate a 19-dimensional state vector ω to describe individual forecasts for the federal funds rate target by the FOMC members and regional Fed presidents. In doing so, the Fed partitions the real line into intervals with a step of 25 basis points, and then, for each interval

[20] Between 2011 and 2018, the Fed held a press conference after every other meeting. It is well-known that the meetings followed by the press conference were characterized by a disproportionate response in financial markets (Boguth et al., 2018).

and forecast horizon, communicates the number of FOMC members whose forecast belongs to a given interval, but without revealing the members' identities. Such a design leads to a partition into bins that are not convex. Hence, under the linear approximation, our result implies that the dot plots are a suboptimal form of communication. It could be improved, for instance, by revealing to the public an interval to which an average of individual forecasts belongs.

Finally, by characterizing the properties of the information relevance matrix and deriving the "only reveal the essential" principle, we formalize the notion that cacophony in communication can be counterproductive, especially when it disseminates information with little effect for welfare. As argued by Kliesen et al. (2019), the Fed's communication has become more complex on multiple dimensions, such as the length and linguistic sophistication of the FOMC statements, frequency of speeches by the Fed officials, and usage of different media (e.g., interviews with the press). Part of this development is justified by the transparency goals, the attainment of the Fed's mandate, and the inherent uncertainty that the Fed faces. However, our results suggest that "the more is better" does not necessarily apply in the context of optimal communication. Instead, "less can be more," in the sense that the Fed could benefit from a targeted choice of topics to emphasize.

4 Conclusion

Central banks around the world use a variety of strategies to communicate to financial markets information about the state of the economy and monetary policy. The observed heterogeneity of approaches raises the question which communication strategies are appropriate and which should be discarded. Building on the Bayesian persuasion literature, in this paper, we study the optimal design of communication by a policymaker under commitment. We extend the literature by introducing features necessary to capture the practical challenges faced by central banks. Most importantly, we allow the information set of the policymaker to be multidimensional, reflecting the fact that central banks process vast amounts of information which they need to condense into a concise message to be shared with the public. We show that optimal communication is given by a partition of the state space, whereby the policymakers reveals to the public the "cluster" to which the state belongs. This result provides a micro-foundation for the preference for rules versus discretion, with the key implication that randomization in central bank communication—whereby the policymaker injects noise into the public information set—is never optimal. Optimal communication is more precise when the beliefs of the central bank align with those of the public and is vague in case of divergence. To characterize the properties of the optimal communication in the multidimensional case, we introduce a novel object, the information relevance matrix, along with the principal information components (PICs) associated with it. We use the PICs to describe the information vectors that the central bank should focus on to maximize welfare and which they should discard.

References

Ai, Hengjie and Ravi Bansal, "Risk Preferences and the Macroeconomic Announcement Premium," *Econometrica*, 2018.

Amador, Manuel and Pierre-Olivier Weill, "Learning from Private and Public Observations of Others' Actions," *Journal of Economic Theory*, 2012, *147*, 910–940.

Aumann, Robert J and Michael Maschler, *Repeated games with incomplete information*, MIT Press, 1995.

Bergemann, Dirk and Stephen Morris, "Information Design: A Unified Perspective," Technical Report, Yale University 2017.

Blinder, Alan and Ricardo Reis, "Understanding the Greenspan standard," in "Federal Reserve Bank of Kansas Jackson Hole Symposium on the Global Dimensions of Unconventional Monetary Policy, Jackson Hole, August" 2005, pp. 11–96.

Blinder, Alan S, Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen, "Central bank communication and monetary policy: A survey of theory and evidence," *Journal of Economic Literature*, 2008, *46* (4), 910–45.

Boguth, Oliver, Vincent Grégoire, and Charles Martineau, "Shaping expectations and coordinating attention: The unintended consequences of FOMC press conferences," *Journal of Financial and Quantitative Analysis*, 2018, pp. 1–27.

Bond, Philip and Itay Goldstein, "Government Intervention and Information Aggregation by Prices," *Journal of Finance*, 2015, *70*, 2777–2812.

Calzolari, Giacomo and Alessandro Pavan, "On the optimality of privacy in sequential contracting," *Journal of Economic Theory*, 2006, *130* (1).

Campbell, Jeffrey, Charles Evans, Jonas Fisher, and Alejandro Justiano, "Macroeconomic Effects of FOMC Forward Guidance," *Brookings Papers on Economic Activity*, 2012.

Cecchetti, Stephen G. and Kermit L. Schoenholtz, "Improving U.S. Monetary Policy Communications," Technical Report, NYU Stern 2019. Working paper.

Cieslak, Anna, Adair Morse, and Annette Vissing-Jorgensen, "Stock Returns over the FOMC Cycle," *Journal of Finance*, 2019, *74*, 2201–2248.

— **and Andreas Schrimpf**, "Non-Monetary News in Central Bank Communication," *Journal of International Economics*, 2019.

Coeuré, Benoit, "Central Bank Communication in a Low Interest Rate Environment," *speech at Bruegel event, Brussels, 31 March 2017*, 2017.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar, "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, 2018, *108*, 2671–2713.

D'Acunto, Francesco, Daniel Hoang, and Michael Weber, "Managing Households' Expectations with Simple Economic Policies," *Working paper*, 2019.

— **,** — **, Maritta Paloviita, and Michael Weber**, "Cognitive Abilities and Inflation Expectations," *American Economic Review: Papers and Proceedings*, 2019.

Dworczak, Piotr and Giorgio Martini, "The Simple Economics of Optimal Persuasion," *Journal of Polytical Economy*, 2019.

Faust, Jon, Eric T Swanson, and Jonathan H Wright, "Do Federal Reserve policy surprises reveal superior information about the economy?," *Contributions in Macroeconomics*, 2004, *4* (1).

Gertler, Mark and Peter Karadi, "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 2015, *7*, 44–76.

Goldstein, Itay and Liyan Yang, "Information Disclosure in Financial Markets," *Annual Review of Financial Economics*, 2017, *9*, 101–125.

Goodhart, Charles A. E., "Problems of Monetary Management: The U.K. Experience," *Papers in Monetary Economics. Reserve Bank of Australia*, 1975.

Haldane, Andrew, "A little more conversation. A little less action," *speech at the Federal Reserve Bank of San Francisco Macroeconomics and Monetary Policy Conference, 31 March 2017*, 2017.

Hanson, Samuel G and Jeremy C Stein, "Monetary policy and long-term real rates," *Journal of Financial Economics*, 2015, *115* (3), 429–448.

Hopenhayn, Hugo and Maryam Saeedi, "Optimal Ratings and Market Outcomes," Technical Report, UCLA 2019.

Hugonnier, Julien, Semyon Malamud, and Eugene Trubowitz, "Endogenous Completeness of Diffusion Driven Equilibrium Markets," *Econometrica*, 2012, (80), 1249–1270.

Jeffery, Morris Stephen Amato and Hyun Song Shin, "Communication and Monetary Policy," *Oxford Review of Economic Policy*, 2002, *18*, 495–503.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *101*, 2590–2615.

Kliesen, Kevin L., Brian Levine, and Christopher J. Waller, "Gauging Market Response to Monetary Policy Communication," *Federal Reserve Bank of St. Louis Review*, 2019.

Kocherlakota, Narayana, "Bring the Fed's Dead Meetings to Life: Markets dismiss non-press-conference meetings. The Fed should bow to reality.," *Bloomberg Opinion, October 12, 2016*, 2016.

Kolotilin, Anton, "Optimal information disclosure: a linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–636.

Kroencke, Tim, Mike Schmeling, and Andreas Schrimpf, "The FOMC risk shift," Technical Report, BIS 2019.

Kuttner, Kenneth N., "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 2001, *47*, 523–544.

Mensch, Jeffrey, "Monotone Persuasion," *Manuscript*, 2018.

Morris, Stephen and Hyun Song Shin, "The Social Value of Public Information," *American Economic Review*, 2002, *92*, 1521–1534.

___ **and** ___ , "Central Bank Transparency and the Signal Value of Prices," *Brookings Papers on Economic Activity*, 2005, *2*, 1–66.

___ **and** ___ , "Central Bank Forward Guidance and the Signal Value of Market Prices," *BIS Working Papers*, 2018, *692.*

Olson, Peter and David Wessel, "Federal Reserve Communications: Survey Results," Technical Report, Hutchins Center on Fiscal & Monetary Policy, Brookings Institution 2016.

Ostrovsky, Michael and Michael Schwarz, "Information disclosure and unraveling in matching markets," *American Economic Journal: Microeconomics*, 2010, *2* (2).

Rayo, Luis and Ilya Segal, "Optimal Information Disclosure," *Journal of Political Economy*, 2010, *118*, 949–987.

Romer, Christina D. and David H. Romer, "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 2000, *90*, 430–457.

Samuelson, Paul, *panel remarks, Federal Reserve Bank of Boston conference, Conference Series*, 1994, *38.*

Shin, Hyun Song, "Can Central Banks Talk Too Much?," *speech at the ECB conference on Policy Effectiveness, Accountability and Reputation*, 2017.

Stein, Jeremy and Adi Sunderam, "The Fed, the Bond Market, and Gradualism in Monetary Policy," *Journal of Finance*, 2018, *73*, 1015–1060.

Svensson, Lars and Michael Woodford, "Indicator Variables for Optimal Policy under Asymmetric Information," *Journal of Economic Dynamics and Control*, 2004, *28*, 661–690.

Woodford, Michael, "Nonstandard Indicators for Monetary Policy: Can Their Usefulness Be Judged from Forecasting Regressions?," *Monetary Policy, edited by N. G. Mankiw. Chicago: Univ. of Chicago Press for NBER*, 1994.

A Proof of Theorem 7

We start the proof with a simple technical compactness result.

Lemma 17 *The set of information designs is compact in the topology induced by the metric* (2)*, and so is the set of partition in the corresponding induced metric.*

Proof. Denote by $\bar{\Omega}^K$ the set of K copies of $\bar{\Omega}$. Clearly, any signal structure π corresponds to a unique probability measure m^π on $\bar{\Omega}^K$: for any set $A = (A_1, \cdots, A_K) \subset \bar{\Omega}^K$ (with $A_k \subset \bar{\Omega}$), we can define

$$m^\pi(A) \;=\; \sum_k \int_{A_k} \pi_k(\omega) d\omega.$$

Since $\bar{\Omega}$ is compact, Prokhorov's theorem implies that the set of probability measures on $\bar{\Omega}^K$ is compact in the topology of weak convergence. Since $\pi_k(\omega) \in (0,1)$, it is straightforward to show that weak convergence also implies convergence in the metric (2). Furthermore, if an information design is a partition, so that $\pi_k \in \{0,1\}$ Lebesgue-almost surely, then in the limit it also has to converge to a partition. Q.E.D.

The equilibrium conditions can be rewritten as

$$E_{\mu_s}[G(a(s),\omega)|s] \;=\; 0\,.$$

Here,

$$\mu_s(\omega) \;=\; \frac{\pi(s|\omega)\mu_0(\omega)}{\int \pi(s|\omega)\mu_0(\omega)d\omega}$$

and hence

$$E_{\mu_k}[G(a(s),\omega)] \;=\; \frac{\int \pi(k|\omega)\mu_0(\omega)G(a(k),\; \omega)d\omega}{\int \pi(k|\omega)\mu_0(\omega)d\omega}\,.$$

By assumption, equilibrium a depends continuously on $\{\pi_k\}$. Since the map

$$(\{\pi_k\},\ \{a_k\}) \ \rightarrow\ \left\{\int \pi_k(\omega)\mu_0(\omega)G(a(k,\varepsilon),\ \omega)d\omega\right\}$$

is real analytic, and has a non-degenerate Jacobian with respect to a, the assumed continuity of a and the implicit function theorem implies that a is in fact real analytic in $\{\pi_k\}$. To compute the Frechet differentials of $a(s)$, we take a small perturbation $\eta(\omega)$ of $\pi_k(\omega)$. By the regularity assumption and the Implicit Function Theorem,

$$a(k,\varepsilon) \ = \ a(k) + \varepsilon a^{(1)}(k) + 0.5\varepsilon^2 a^{(2)}(k) \ + \ o(\varepsilon^2)$$

for some $a^{(1)}(k),\ a^{(2)}(k)$. Let us rewrite

$$\begin{aligned}
0 \ &= \ \int (\pi_k(\omega) + \varepsilon\eta(\omega))\mu_0(\omega)G(a(k,\varepsilon),\ \omega)d\omega \\
&= \ \int (\pi_k(\omega) + \varepsilon\eta(\omega))\mu_0(\omega)G(a(k) + \varepsilon a^{(1)}(k) + 0.5\varepsilon^2 a^{(2)}(k),\ \omega)d\omega \\
&\approx \ \Bigg(\int \pi_k(\omega)\mu_0(\omega)\Bigg(G(a(k),\ \omega) + G_a(\varepsilon a^{(1)}(k) + 0.5\varepsilon^2 a^{(2)}(k)) \\
&+ 0.5G_{aa}(\varepsilon a^{(1)}(k), \varepsilon a^{(1)}(k))\Bigg)d\omega \\
&+ \varepsilon \int \eta(\omega)\mu_0(\omega)\Bigg(G(a(k)) + G_a\varepsilon a^{(1)}(k)\Bigg)d\omega\Bigg)\Bigg) \\
&= \ \Bigg(\varepsilon\Bigg(\int \pi_k(\omega)\mu_0(\omega)G_a a^{(1)}(k)d\omega + \int \eta(\omega)\mu_0(\omega)G(a(k))d\omega\Bigg) \\
&+ 0.5\varepsilon^2\Bigg(\int \pi_k(\omega)\mu_0(\omega)[G_a a^{(2)}(k) + G_{aa}(a(k),\omega)(a^{(1)}(k), a^{(1)}(k))]d\omega \\
&+ 2\int \eta(\omega)\mu_0(\omega)G_a(a(k),\omega)a^{(1)}(k)d\omega\Bigg)\Bigg)\Bigg)
\end{aligned}$$

As a result, we get

$$a^{(1)}(k) \;=\; -\bar{G}_a(k)^{-1}\int \eta(\omega)\mu_0(\omega)G(a(k),\omega)d\omega,\;\bar{G}_a(k) \;=\; \int \pi_k(\omega)\mu_0(\omega)G_a d\omega\,,$$

while

$$a^{(2)}(k) \;=\; -\bar{G}_a(k)^{-1}\Bigg(\int \pi_k(\omega)\mu_0(\omega)G_{aa}(a(k),\omega)(a^{(1)}(k),a^{(1)}(k))d\omega$$
$$+\;2\int \eta(\omega)\mu_0(\omega)G_a(a(k),\omega)a^{(1)}(k)d\omega\Bigg)\,.$$

Consider the social welfare function

$$\bar{W}(\pi) \;=\; E[W(\omega,a(s))] \;=\; \sum_k \int_\Omega W(\omega,a(k))\pi_k(\omega)\mu_0(\omega)d\omega\,.$$

Suppose that the optimal information structure is not a partition. Then, there exists a subset $I\subset\Omega$ of positive μ_0-measure and an index k such that $\pi_k(\omega)\in(0,1)$ for μ_0-almost all $\omega\in I$. Since $\sum_i \pi_i(\omega)=1$ and $\pi_i(\omega)\in[0,1]$, there must be an index $k_1\neq k$ and a subset $I_1\subset I$ such that $\pi_{k_1}(\omega)\in(0,1)$ for μ_0-almost all $\omega\in I_1$. Consider a small perturbation $\{\tilde{\pi}(\varepsilon)\}_i$ of the information design, keeping $\pi_i,\; i\neq k,k_1$ fixed and changing $\pi_k(\omega)\to\pi_k(\omega)+\varepsilon\eta(\omega),\;\pi_{k_1}(\omega)\to\pi_{k_1}(\omega)-\varepsilon(\omega)$ where $\eta(\omega)$ in an arbitrary bounded function with $\eta(\omega)=0$ for all $\omega\notin I_1$. Define $\eta_k(\omega)=\eta(\omega),\;\eta_{k_1}(\omega)=-\eta(\omega)$, and $\eta_i(\omega)=0$ for all $i\neq k,k_1$. A second-order Taylor expansion in ε gives

$$
\begin{aligned}
&\sum_i \int_\Omega W(\omega, a(i,\varepsilon))(\pi_i(\omega)+\varepsilon\eta_i(\omega))\mu_0(\omega)d\omega \\
&\approx \int_\Omega \Bigg(W(\omega, a(i)) + W_a(\omega, a(i))(\varepsilon a^{(1)}(i) + 0.5\varepsilon^2 a^{(2)}(i)) \\
&+ 0.5 W_{aa}(\omega, a(i))\varepsilon^2 (a^{(1)}(i), a^{(1)}(i))\Bigg)(\pi_i(\omega)+\varepsilon\eta_i(\omega))\mu_0(\omega)d\omega \\
&= \bar{W}(\pi) \; + \; \varepsilon \sum_i \Bigg(\int_\Omega (W(\omega, a(i))\eta_i(\omega) + W_a(\omega, a(i))a^{(1)}(i)\pi_i(\omega))\mu_0(\omega)d\omega \Bigg) \\
&+ 0.5\varepsilon^2 \sum_i \int_\Omega \Big(W_{aa}(\omega, a(i))(a^{(1)}(i), a^{(1)}(i))\pi_i(\omega) \\
&+ W_a(\omega, a(i))a^{(2)}(i)\pi_i(\omega) + W_a(\omega, a(i))a^{(1)}(i)\eta_i(\omega)\Big)\mu_0(\omega)d\omega
\end{aligned}
\tag{16}
$$

Since, by assumption, $\{\pi_i\}$ is an optimal information design, it has to be that the first order term in (18) is zero, while the second-order term is always non-positive. We can rewrite the first order term as

$$
\begin{aligned}
&\sum_i \Bigg(\int_\Omega (W(\omega, a(i))\eta_i(\omega) + W_a(\omega, a(i))a^{(1)}(i)\pi_i(\omega))\mu_0(\omega)d\omega \Bigg) \\
&= \sum_i \int_\Omega \Bigg(W(\omega, a(i)) \\
&- \Big(\int W_a(\omega_1, a(i))\pi_i(\omega_1)\mu_0(\omega_1)d\omega_1 \Big) \bar{G}_a(i)^{-1} \, G(a(i), \omega) \Bigg) \eta_i(\omega)\mu_0(\omega)d\omega
\end{aligned}
\tag{17}
$$

and hence it is zero for all considered perturbations if and only if

$$
\begin{aligned}
&W(\omega, a(k)) \; - \; \Big(\int W_a(\omega_1, a(k))\pi_k(\omega_1)\mu_0(\omega_1)d\omega_1 \Big) \bar{G}_a(k)^{-1} \, G(a(k), \omega) \\
&= \; W(\omega, a(k_1)) \; - \; \Big(\int W_a(\omega_1, a(k_1))\pi_{k_1}(\omega_1)\mu_0(\omega_1)d\omega_1 \Big) \bar{G}_a(k_1)^{-1} \, G(a(k_1), \omega)
\end{aligned}
\tag{18}
$$

Lebesgue-almost surely for $\omega \in I_1$. By Proposition 6, (18) also holds for all $\omega \in \Omega$. Hence, by Assumption 3, $a(k) = a(k_1)$, which contradicts Definition 3 of a regular equilibrium.

Proof of Theorem 8. Suppose a partition $\omega = \cup_k \Omega_k$ is optimal:

$$\int_{\Omega_k} G(a(k),\ \omega)\mu_0(\omega)d\omega \ = \ 0$$

defines $(a_n(k))$. Consider a small perturbation, whereby we move a small mass on a set $\mathcal{I} \subset \Omega_k$ to Ω_l. Then, the marginal change in $a_n(k)$ can be determined from

$$\begin{aligned}
0 \ &= \ \int_{\Omega_k} G(a(k),\ \omega)\mu_0(\omega)d\omega \ - \ \int_{\Omega_k \setminus \mathcal{I}} G(a(k,\mathcal{I}),\ \omega)\mu_0(\omega)d\omega \\
&\approx \ - \int_{\Omega_k} DG(a(k),\ \omega)\Delta a(k)\ \mu_0(\omega)d\omega \ + \int_{\mathcal{I}} G(a(k),\ \omega)\mu_0(\omega)d\omega\,,
\end{aligned}$$

implying that the first order change in a is given by

$$\Delta a(k) \ = \ (\bar{D}G(k))^{-1} \int_{\mathcal{I}} G(a(k),\ \omega)\mu_0(\omega)d\omega$$

Thus, the change in welfare is[21]

$$\begin{aligned}
\Delta W \ &= \ \int_{\Omega_k} W(a(k),\ \omega)\mu_0(\omega)d\omega \ - \ \int_{\Omega_k \setminus \mathcal{I}} W(a(k,\mathcal{I}),\ \omega)\mu_0(\omega)d\omega \\
&+ \int_{\Omega_l} W(a(l),\ \omega)\mu_0(\omega)d\omega \ - \ \int_{\Omega_l \cup \mathcal{I}} W(a(l,\mathcal{I}),\ \omega)\mu_0(\omega)d\omega \\
&\approx \ - \int_{\Omega_k} DW(a(k),\ \omega)\Delta a(k)\mu_0(\omega)d\omega + \int_{\mathcal{I}} W(a(k),\ \omega)\mu_0(\omega)d\omega \\
&- \int_{\Omega_l} DW(a(l),\ \omega)\Delta a(l)\mu_0(\omega)d\omega - \int_{\mathcal{I}} W(a(l),\ \omega)\mu_0(\omega)d\omega \\
&= \ -\bar{D}W(k)(\bar{D}G(k))^{-1} \int_{\mathcal{I}} G(a(k),\ \omega)\mu_0(\omega)d\omega + \int_{\mathcal{I}} W(a(k),\ \omega)\mu_0(\omega)d\omega \\
&+ \bar{D}W(l)(\bar{D}G(l))^{-1} \int_{\mathcal{I}} G(a(l),\ \omega)\mu_0(\omega)d\omega - \int_{\mathcal{I}} W(a(l),\ \omega)\mu_0(\omega)d\omega
\end{aligned}$$

[21]Note that DW is a horizontal (row) vector.

This expression has to be non-negative for any $\mathcal{I}$ of positive Lebesgue measure. Thus,

$$
\begin{aligned}
&- \bar{D}W(k)(\bar{D}G(k))^{-1}G(a(k),\ \omega) + W(a(k),\ \omega) \\
&+ \bar{D}W(l)(\bar{D}G(l))^{-1}G(a(l),\ \omega)\ -\ W(a(l),\ \omega)\ \geq\ 0
\end{aligned}
$$

for Lebesgue almost any $\omega \in \Omega_k$. Q.E.D.

Proof of Lemma 9. By Lemma 17, the set of partitions is compacts and hence we can find a subsequence $\{\Omega_k(\varepsilon_j)\}$ converging to some partition $\{\Omega_k(0)\}$. We have

$$
0\ =\ \int_{\Omega_k(\varepsilon)} G(a(k,\varepsilon),\varepsilon\omega)\mu_0(\omega)d\omega\ =\ \int_{\tilde{\Omega}_k(\varepsilon)} G(a(k,\varepsilon),\omega)\mu_0(\omega)d\omega
$$

Now,

$$
\begin{aligned}
0\ &=\ \int_{\Omega_k(\varepsilon)} G(a(k,\varepsilon),\varepsilon\omega)\mu_0(\omega)d\omega \\
&=\ G(a(k,\varepsilon),0)M(\Omega_k(\varepsilon))\ +\ \varepsilon\, G_\omega(a(k,\varepsilon),0)\, M_1(\Omega_k(\varepsilon))\ +\ O(\varepsilon^2)\,.
\end{aligned}
\tag{19}
$$

Let us show that $a(k,\varepsilon) - a(k,0)\ =\ O(\varepsilon)$. Suppose the contrary. Then there exists a sequence $\varepsilon_m \to 0$ such that $\|a(k,\varepsilon) - a(k,0)\|\varepsilon^{-1}\ \to \infty$. We have

$$
\begin{aligned}
&G(a(k,\varepsilon),0)\ -\ G(a(k,0),0)\ =\ \int_0^1 DG(a(k,0) + t(a(k,\varepsilon) - a(k,0)))(a(k,\varepsilon) - a(k,0))dt \\
&\geq\ c\|a(k,\varepsilon) - a(k,0)\|
\end{aligned}
$$

for some $c > 0$ due to the continuity and non-degeneracy of $DG(0) = DG(a(k,0))$. Dividing (20) by ε, we get a contradiction.

Define

$$
a^{(1)}(k)\ \equiv\ -DG(0)^{-1}G_\omega(a(k),0)M_1(\Omega_k(0))\ =\ -\mathcal{G}\, M_1(\Omega_k(0))\,.
$$

Let us now show that $a(k,\varepsilon) - a(k,0) \; = \; \varepsilon a^{(1)}(k) \; + \; o(\varepsilon)$. Suppose the contrary. Then, $\|\varepsilon^{-1}(a(k,\varepsilon) - a(k,0)) - a^{(1)}(k)\| \; > \; c$ for some $c > 0$ along a sequence of $\varepsilon \to 0$. By (20),

$$
\begin{aligned}
0 \; &= \; \int_{\Omega_k(\varepsilon)} G(a(k,\varepsilon), \varepsilon\omega)\mu_0(\omega)d\omega \\
&= \; G(a(k,\varepsilon),0)M(\Omega_k(\varepsilon)) \; + \; \varepsilon\, G_\omega(a(k,\varepsilon),0)\, M_1(\Omega_k(\varepsilon)) \; + \; O(\varepsilon^2) \\
&= \; \varepsilon DG(0)\varepsilon^{-1}(a(k,\varepsilon) - a(k,0))M(\Omega_k(\varepsilon)) \; + \; \varepsilon\, G_\omega(a(k),0)\, M_1(\Omega_k(\varepsilon)) \; + \; O(\varepsilon^2)\,,
\end{aligned}
\tag{20}
$$

and we get a contradiction taking the limit as $\varepsilon \to 0$. Q.E.D.

Proof of Theorem 10. Consider the real analytic hyper-surface

$$
\begin{aligned}
\Psi_{k,l}(\varepsilon) \; = \; \{\omega \in \Omega : \; &-\bar{D}W(k,\varepsilon)(\bar{D}G(k,\varepsilon))^{-1}G(a(k,\varepsilon),\; \varepsilon\omega) + W(a(k,\varepsilon),\; \varepsilon\omega) \\
&+ \bar{D}W(l,\varepsilon)(\bar{D}G(l,\varepsilon))^{-1}G(a(l,\varepsilon),\; \varepsilon\omega) \; - \; W(a(l,\varepsilon),\; \varepsilon\omega) \; = \; 0\}\,.
\end{aligned}
$$

We have

$$
\begin{aligned}
\bar{D}W(k,\varepsilon) \; &= \; \int_{\Omega_k(\varepsilon)} DW(a(k,\varepsilon),\varepsilon\omega)\mu_0(\omega)d\omega \\
&= \; \int_{\Omega_k(\varepsilon)} (DW(0) + \varepsilon\omega^\top DW_\omega(0)^\top + \varepsilon a^{(1)}(k)^\top D^2W(0) + o(\varepsilon))\mu_0(\omega)d\omega \\
&= \; (DW(0) + \varepsilon(a^{(1)}(k))^\top D^2W(0) \; + \; \varepsilon\, M_1(\Omega_k(0))^\top (DW_\omega(0))^\top)M(\Omega_k(\varepsilon)) \; + \; o(\varepsilon) \; \in \mathbb{R}^{1\times(Nm)}\,.
\end{aligned}
$$

At the same time, an analogous calculation implies that

$$
\bar{D}G(k,\varepsilon) \; = \; (DG(0) + \varepsilon(a^{(1)}(k))^\top D^2G(0) \; + \; \varepsilon\, DG_\omega(0)M_1(\Omega_k(0)))M(\Omega_k(\varepsilon)) \; + \; o(\varepsilon)
$$

Here, $DG(0) = (\partial G_i/\partial a_j)_{i,j=1}^A$ and hence

$$
(DG_\omega(0)M_1(\Omega_k(0)))_{i,j} \; = \; \sum_k \frac{\partial^2 G_i}{\partial a_j \partial \omega_k} M_{1,k}\,,
$$

and, similarly,

$$((a^{(1)}(k))^\top D^2G(0))_{i,j} \;=\; \sum_l (a^{(1)}(k))_l \frac{\partial^2 G_l}{\partial a_i \partial a_j} \;\in \mathbb{R}^{(Nm)\times(Nm)}\,.$$

Thus, the inverse of DG can be computed as

$$\begin{aligned}
&M(\Omega_k(\varepsilon))\bar{D}G(k,\varepsilon)^{-1}\\
&= \; DG(0)^{-1} - DG(0)^{-1}\varepsilon\Big((a^{(1)}(k))^\top D^2G(0) \;+\; \varepsilon\, DG_\omega(0)M_1(\Omega_k(0))\Big)DG(0)^{-1} \;+\; o(\varepsilon)\,,
\end{aligned}$$

and therefore

$$\begin{aligned}
&\bar{D}W(k,\varepsilon)(\bar{D}G(k,\varepsilon))^{-1} \;=\; DW(0)DG(0)^{-1}\\
&+\; \varepsilon(M_1^\top DW_\omega(0)^\top DG(0)^{-1} \;+\; (a^{(1)}(k))^\top D^2W(0)DG(0)^{-1})\\
&-\varepsilon DW(0)DG(0)^{-1}\Big((a^{(1)}(k))^\top D^2G(0) \;+\; DG_\omega(0)M_1(\Omega_k(0))\Big)DG(0)^{-1} +\; o(\varepsilon)\\
&=\; DW(0)DG(0)^{-1}\\
&+\; \varepsilon(M_1^\top DW_\omega(0)^\top DG(0)^{-1} \;-\; M_1^\top \mathcal{G}^\top D^2W(0)DG(0)^{-1})\\
&-\varepsilon DW(0)DG(0)^{-1}\Big(-M_1^\top \mathcal{G}^\top D^2G(0) \;+\; DG_\omega(0)M_1\Big)DG(0)^{-1} +\; o(\varepsilon)\\
&=\; DW(0)DG(0)^{-1} + \varepsilon\Gamma + o(\varepsilon)\,,
\end{aligned}$$

where

$$\begin{aligned}
\Gamma \;&=\; M_1^\top DW_\omega(0)^\top DG(0)^{-1}\\
&-M_1^\top \mathcal{G}^\top D^2W(0)DG(0)^{-1} - DW(0)DG(0)^{-1}\Big(-M_1^\top \mathcal{G}^\top D^2G(0) \;+\; DG_\omega(0)M_1\Big)DG(0)^{-1}\,.
\end{aligned}$$

Define

$$\tilde{a}^{(1)}(k,\varepsilon) \;\equiv\; \varepsilon^{-1}(a(k,\varepsilon) - a(k,0)) \;=\; a^{(1)}(k) \;+\; o(1)\,.$$

Let also

$$G^{(2)}(k) \;\equiv\; 0.5\varepsilon^2 (a^{(1)}(k)^\top D^2 G(0) a^{(1)}(k) + \; 2\omega^\top DG_\omega(0) a^{(1)}(k) \;+\; \omega^\top G_{\omega\omega}(0)\omega)$$

so that

$$G(a(k,\varepsilon),\; \varepsilon\omega) \;-\; (\varepsilon DG(0)\tilde{a}^{(1)}(k,\varepsilon) + \varepsilon G_\omega(0)\omega) \;=\; \varepsilon^2 G^{(2)}(k) \;+\; o(\varepsilon^2)\,,$$

where we have used that $G(0) = 0$. While we cannot prove that $\varepsilon\tilde{a}^{(1)}(k) = o(\varepsilon^2)$, we show that this term cancels out. We have

$$\begin{aligned}
&-\bar{D}W(k,\varepsilon)(\bar{D}G(k,\varepsilon))^{-1} G(a(k,\varepsilon),\; \varepsilon\omega) + W(a(k,\varepsilon),\; \varepsilon\omega) \\
&\approx\; -\bar{D}W(k,\varepsilon)(\bar{D}G(k,\varepsilon))^{-1}\Big(\varepsilon DG(0)\tilde{a}^{(1)}(k,\varepsilon) + \varepsilon G_\omega(0)\omega + \varepsilon^2 G^{(2)}(k) \;+\; o(\varepsilon^2)\Big) \\
&+\Bigg(W(0) + \varepsilon DW(0)\tilde{a}^{(1)}(k,\varepsilon) + \varepsilon W_\omega(0)\omega \\
&+ 0.5\varepsilon^2\Big((a^{(1)}(k))^\top D^2 W(0) a^{(1)}(k) + \omega^\top W_{\omega,\omega}(0)\omega + 2(a^{(1)}(k))^\top DW_\omega(0)\omega\Big) \;+\; o(\varepsilon^2)\Bigg) \\
&=\; -\Big(DW(0)DG(0)^{-1} \;+\; \varepsilon\Gamma + o(\varepsilon)\Big) \\
&\times \Big(\varepsilon DG(0)\tilde{a}^{(1)}(k,\varepsilon) + \varepsilon G_\omega(0)\omega + \varepsilon^2 G^{(2)}(k) \;+\; o(\varepsilon^2)\Big) \\
&+\Bigg(W(0) + \varepsilon DW(0)\tilde{a}^{(1)}(k) + \varepsilon W_\omega(0)\omega \\
&+ 0.5\varepsilon^2\Big((a^{(1)}(k))^\top D^2 W(0) a^{(1)}(k) + \omega^\top W_{\omega,\omega}(0)\omega + 2(a^{(1)}(k))^\top DW_\omega(0)\omega\Big) + o(\varepsilon^2)\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \; W(0) \\
&+ \; \varepsilon\Bigg(- DW(0)DG(0)^{-1}\Big(DG(0)\tilde{a}^{(1)}(k,\varepsilon) + G_\omega(0)\omega\Big) \; + DW(0)\tilde{a}^{(1)}(k,\varepsilon) + W_\omega(0)\omega\Bigg) \\
&+ \varepsilon^2\Bigg(- DW(0)DG(0)^{-1}\varepsilon^2 G^{(2)}(k) - \Gamma\Big(DG(0)a^{(1)}(k) + G_\omega(0)\omega\Big) \\
&+ 0.5\Big((a^{(1)}(k))^\top D^2W(0)a^{(1)}(k) + \omega^\top W_{\omega,\omega}(0)\omega + 2a^{(1)}(k)^\top DW_\omega(0)\omega\Big)\Bigg) \; + \; o(\varepsilon^2)\,.
\end{aligned}
$$

We have

$$
\begin{aligned}
&\Gamma\Big(DG(0)a^{(1)}(k) + G_\omega(0)\omega\Big) \; = \; \Big(M_1^\top DW_\omega(0)^\top DG(0)^{-1} \\
&- M_1^\top \mathcal{G}^\top D^2W(0)DG(0)^{-1} - DW(0)DG(0)^{-1}\Big(- M_1^\top \mathcal{G}^\top D^2G(0) \; + \; DG_\omega(0)M_1\Big)DG(0)^{-1}\Big) \\
&\times G_\omega(0)(\omega - M_1) \\
&= \; \Big(M_1^\top DW_\omega(0)^\top - M_1^\top \mathcal{G}^\top D^2W(0) - DW(0)DG(0)^{-1}\Big(- M_1^\top \mathcal{G}^\top D^2G(0) \; + \; DG_\omega(0)M_1\Big)\Big) \\
&\times \mathcal{G}(\omega - M_1) \; = \; M_1^\top \mathcal{D}_1 \mathcal{G}(\omega - M_1)
\end{aligned}
$$

where

$$
\mathcal{D}_1 \; = \; DW_\omega(0)^\top - DW(0)DG(0)^{-1}DG_\omega(0) - (\mathcal{G}^\top D^2W(0) - \mathcal{G}^\top DW(0)DG(0)^{-1}D^2G(0)) \in \mathbb{R}^{L\times(Nm)}
$$

and where the three-dimensional tensor multiplication is understood as follows:

$$
\begin{aligned}
M_1^\top DW(0)DG(0)^{-1}DG_\omega(0) \; &= \; \sum_k M_{1,k} DW(0)DG(0)^{-1}DG_{\omega_k}(0) \\
M_1^\top \mathcal{G}^\top DW(0)DG(0)^{-1}DG_\omega(0) \; &= \; \sum_k (\mathcal{G}M_1)_k DW(0)DG(0)^{-1}DG_{a_k}(0)\,.
\end{aligned}
$$

Rewriting, we get

$$
\begin{aligned}
&W(0) \; + \; \varepsilon\Bigg(- DW(0)DG(0)^{-1}G_\omega(0)\omega + W_\omega(0)\omega\Bigg) \\
&+ \varepsilon^2\Bigg(- DW(0)DG(0)^{-1}\varepsilon^2 G^{(2)}(k) - M_1^\top \mathcal{D}_1\mathcal{G}(\omega - M_1) \\
&+ 0.5\Big((a^{(1)}(k))^\top D^2W(0)a^{(1)}(k) + \omega^\top W_{\omega,\omega}(0)\omega + 2a^{(1)}(k)^\top DW_\omega(0)\omega\Big)\Bigg) \; + \; o(\varepsilon^2) \\
&= \; W(0) \; + \; \varepsilon\Bigg(- DW(0)\mathcal{G}\omega + W_\omega(0)\omega\Bigg) \\
&+ \varepsilon^2\Bigg(- DW(0)DG(0)^{-1}\varepsilon^2 G^{(2)}(k) - M_1^\top \mathcal{D}_1\mathcal{G}(\omega - M_1) \\
&+ 0.5\Big(M_1^\top\mathcal{G}^\top D^2W(0)\mathcal{G}M_1 + \omega^\top W_{\omega,\omega}(0)\omega - 2(\mathcal{G}M_1)^\top DW_\omega(0)\omega\Big)\Bigg) \; + \; o(\varepsilon^2)\,.
\end{aligned}
$$

Now,

$$
\varepsilon^2 G^{(2)}(k) \; = \; 0.5(M_1^\top\mathcal{G}^\top D^2G(0)\mathcal{G}M_1 \; - \; 2(\mathcal{G}M_1)^\top DG_\omega(0)\omega + \omega^\top G_{\omega\omega}(0)\omega)\,.
$$

Thus, the desired expression is given by

$$
W(0) \; + \; \varepsilon\Bigg(- DW(0)\mathcal{G}\omega + W_\omega(0)\omega\Bigg) \; + \; \varepsilon^2(0.5M_1^\top\mathcal{A}M_1 + M_1^\top\mathcal{B}\omega \; + \; \omega^\top\mathcal{C}\omega)
$$

where we have defined

$$
\begin{aligned}
\mathcal{A} \;&\equiv\; -DW(0)DG(0)^{-1}\mathcal{G}^\top D^2G(0)\mathcal{G} + 2\mathcal{D}_1\mathcal{G} + \mathcal{G}^\top D^2W(0)\mathcal{G} \\
&=\; -DW(0)DG(0)^{-1}\mathcal{G}^\top D^2G(0)\mathcal{G} + 2\Big(DW_\omega(0)^\top - DW(0)DG(0)^{-1}DG_\omega(0) \\
&- (\mathcal{G}^\top D^2W(0) - DW(0)DG(0)^{-1}\mathcal{G}^\top D^2G(0))\Big)\mathcal{G} + \mathcal{G}^\top D^2W(0)\mathcal{G} \\
&=\; \mathcal{G}^\top DW(0)DG(0)^{-1}D^2G(0)\mathcal{G} - \mathcal{G}^\top D^2W(0)\mathcal{G} + 2(DW_\omega(0)^\top\mathcal{G} - \mathcal{G}^\top DW(0)DG(0)^{-1}DG_\omega(0)) \in \mathbb{R}^{L\times L} \\
\mathcal{B} \;&\equiv\; \mathcal{G}^\top DW(0)DG(0)^{-1}DG_\omega(0) - \mathcal{D}_1\mathcal{G} - \mathcal{G}^\top DW_\omega(0) \\
&=\; \mathcal{G}^\top DW(0)DG(0)^{-1}DG_\omega^\top(0) - \Big(\mathcal{G}^\top DW_\omega(0) - \mathcal{G}^\top DW(0)DG(0)^{-1}DG_\omega(0) \\
&- (\mathcal{G}^\top D^2W(0)\mathcal{G} - DW(0)DG(0)^{-1}\mathcal{G}^\top D^2G(0))\mathcal{G}\Big) - DW_\omega(0)^\top\mathcal{G}
\end{aligned}
$$

Here, the first term is given by

$$
(DW(0)DG(0)^{-1}DG_\omega^\top(0))_{i,j} \;=\; \sum_k((DW(0)DG(0)^{-1})_k\frac{\partial^2 G_k}{\partial a_i\partial\omega_j}
$$

Q.E.D.

B Proofs for Linearized Partitions: One Dimension

Proof of Proposition 11. In this case, the region Ω_k takes the form

$$
\Omega_k \;=\; \{\omega:\; (\tilde{q}_k - \tilde{q}_l)^\top\begin{pmatrix}\nabla_\omega Z(0)\\ \nabla_\omega G(0)\end{pmatrix}\omega \;\geq\; \tilde{v}_k - \tilde{v}_l\;\forall\; l\}
$$

for all $l \neq k$, while the system takes the form

$$\begin{aligned}
\tilde{q}_k &= \nabla^2 V(Z(0), G(0))\, A_k^{(1)} \\
\tilde{v}_k &= 0.5(A_k^{(1)})^\top \nabla^2 V(Z(0), G(0))\, A_k^{(1)} \\
A_k^{(1)} &= \begin{pmatrix} \nabla Z_\omega(0) \\ \nabla G_\omega(0) \end{pmatrix} x_k \\
x_k &= \int_{\Omega_k^*} \omega \mu_0(\omega) d\omega / M(\Omega_k^*)\,.
\end{aligned}$$

Thus,

$$\tilde{v}_k \;=\; x_k^2 \nu, \tilde{q}_k^\top \begin{pmatrix} \nabla Z_\omega(0) \\ \nabla G_\omega(0) \end{pmatrix} \;=\; \nu x_k,$$

and the partition take the form

$$\Omega_k \;=\; \{\omega :\; \nu(x_k - x_l)\omega \;\geq\; 0.5(x_k^2 - x_l^2)\nu \;\forall l\}\,.$$

Let us order x_k in the increasing order. If $\nu > 0$, then this means

$$(x_k - x_l)\omega \;\geq\; 0.5(x_k^2 - x_l^2)$$

for all $l \neq k$, which means $\omega \geq 0.5(x_k + x_l)$ for all $x_l \leq x_k$ and $\omega \leq 0.5(x_k + x_l)$ for all $x_l \geq x_k$. That is, $\Omega_k = ((x_{k-1} + x_k)/2, (x_k + x_{k+1})/2)$. By contrast, if $\nu < 0$, then the sets is empty. This is intuitive: If $\widetilde{V}$ is convex, then there are gains from concavification. But if it is concave, then it is optimal not to reveal any information.

Then, the system takes the form

$$x_k \;=\; \frac{\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \omega \mu_0(\omega) d\omega}{\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \mu_0(\omega) d\omega}\,.$$

Or, equivalently,

$$\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \omega\mu_0(\omega)d\omega \ - \ x_k \int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \mu_0(\omega)d\omega \ = \ 0\,.$$

Q.E.D.

Proof of Proposition 14. Let x_k^* be the uniform partition of $[0,1]$. Define

$$\eta_k^{(p)} \ \equiv \ \int_{(x_{k-1}^*+x_k^*)/2}^{(x_k^*+x_{k+1}^*)/2} \omega^p\eta(\omega)d\omega\,.$$

Then, define

$$\alpha_k^* \ \equiv \ \frac{8}{x_{k+1}^*-x_{k-1}^*}(x_k^*\eta_k^{(0)} \ - \ \eta_k^{(1)})\,.$$

Then, we can rewrite (21) as

$$\int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \omega d\omega \ - \ x_k \int_{(x_{k-1}+x_k)/2}^{(x_k+x_{k+1})/2} \mu_0(\omega)d\omega \ = \ \varepsilon\frac{x_{k+1}^*-x_{k-1}^*}{8}\alpha_k^* \ + \ O(\varepsilon^2)\,. \tag{21}$$

That is,

$$\frac{1}{8}(x_{k+1}-x_{k-1})(x_{k-1}+x_{k+1}-2x_k) \ = \ \varepsilon\frac{x_{k+1}^*-x_{k-1}^*}{8}\alpha_k^* \ + \ O(\varepsilon^2)$$

Let $x_k = x_k^* + \varepsilon\tilde{x}_k + O(\varepsilon^2)$. Then, $x_{k-1}+x_{k+1}-2x_k = O(\varepsilon)$ and hence

$$\frac{1}{8}(x_{k+1}^*-x_{k-1}^*)\varepsilon(\tilde{x}_{k-1}+\tilde{x}_{k+1}-2\tilde{x}_k) \ = \ \varepsilon\frac{x_{k+1}^*-x_{k-1}^*}{8}\alpha_k^* \ + \ O(\varepsilon^2)\,.$$

Thus, we get the system

$$\tilde{x}_{k-1}+\tilde{x}_{k+1}-2\tilde{x}_k \ = \ \alpha_k^*\,.$$

Thus, the precision of the signal is determined by α_k^* : if $\alpha_k^* > 0$, then the precision is locally increasing, and is decreasing otherwise. By direct calculation, if η is monotone increasing on the k'th interval then α_k^* is negative. Otherwise, it is positive.

The solution to this system is given by

$$\tilde{x}_k \;=\; k\tilde{x}_1 \;+\; \sum_{m=1}^{k-1}(k-m)\alpha_k^*\,,$$

Q.E.D.

Proof of Proposition 15. Let us rewrite the system (11) as

$$F(\mu_0(\cdot),\{x_k\},\Lambda_2) \;=\; 0\,. \tag{22}$$

Suppose first that all solutions $\{x_k^1\}_{k=1}^K$ to (22) are regular in the sense that $\nabla_{\{x_k\}_{k=1}^K}F$ is non-degenerate and $F(\mu_0(\cdot),\{x_k^1\},\Lambda_2)$ is continuously differentiable in Λ_2 at $\Lambda_2 = 0$. Under this assumption, the implicit function theorem immediately yields that all solutions to (11) solve (13) for sufficiently small Λ_2.

Suppose that the claim of the proposition is not correct. Then, there exists a sequence $\Lambda_2(\ell) \to 0$ as $\ell \to \infty$, and solutions $\{\tilde{x}_k^{\mathcal{V}}(\ell)\}_{k=1}^K$ with non-zero $\tilde{x}_k^2(\ell)$ that solve (11). By continuity, we can pick a sub-sequence that converges to a solution to (11) with $\Lambda_2 = 0$, and the claim in the previous paragraph implies the required.

Finally, the Sard theorem implies that all solutions $\{x_k^1\}_{k=1}^K$ to (22) are regular for generic $\mu_0(\omega)$ in the linear manifold defined by condition 14 The proof is complete.

More generally, it is all about the fixed number K: distorting the rectangular structure introduces randomization into the partition. So, if we look a K on the essential part, then the gain cannot be offset, and hence it is better not to reveal anything about the non-essential part (all appealing to the sub-optimality of randomization).

Q.E.D.

Proof of Proposition 16. Consider the optimal partition for revealing information only about $\omega_1^{\mathcal{V}}$. Then, it maximizes the part of (15) with positive eigenvalues, while at the same time revealing no information about $\omega_2^{\mathcal{V}}$. At the same time, revealing any information about $\omega_2^{\mathcal{V}}$ would introduce randomization into the revealed information about $\omega_1^{\mathcal{V}}$ (which we know is sub-optimal) while at the same time reducing the utility component with negative eigenvalues.

Q.E.D.