INFORMATION ACQUISITION, EFFICIENCY, AND NON-FUNDAMENTAL VOLATILITY

Benjamin M. Hébert
Jennifer La'O

Information Acquisition, Efficiency, and Non-Fundamental Volatility
Benjamin M. Hébert and Jennifer La'O
NBER Working Paper No. 26771
February 2020
JEL No. C72,D62,D83

ABSTRACT

This paper analyzes non-fundamental volatility and efficiency in a class of large games (including e.g. linear-quadratic beauty contests) that feature strategic interaction and endogenous information acquisition. We adopt the rational inattention approach to information acquisition but generalize to a large class of information costs. Agents may learn not only about exogenous states, but also about endogenous outcomes. We study how the properties of the agents' information cost relate to the properties of equilibria in these games. We provide the necessary and sufficient conditions information costs must satisfy to guarantee zero non-fundamental volatility in equilibrium, and provide another set of necessary and sufficient conditions to guarantee equilibria are efficient. We show in particular that mutual information, the cost function typically used in the rational inattention literature, both precludes non-fundamental volatility and imposes efficiency, whereas the Fisher information cost introduced by Hébert and Woodford [2020] generates both non-fundamental volatility and inefficiency.

Benjamin M. Hébert
Graduate School of Business
Stanford University
655 Knight Way
Stanford, CA 94305
and NBER
bhebert@stanford.edu

Jennifer La'O
Columbia University
Department of Economics
1105A International Affairs Building
420 West 118th Street
New York, NY 10027
and NBER
jl4196@columbia.edu

1   Introduction

In many economic environments, agents make choices under incomplete information and have incentives to align their actions with both economic "fundamentals" and the actions of other agents [Morris and Shin, 2002, Angeletos and Pavan, 2007]. These games of strategic interaction form the underlying basis for many micro-founded macroeconomic environments; examples include firms' nominal price-setting decisions in New Keynesian models [Woodford, 2003], firms' real quantity choices in business cycle models [Angeletos and La'O, 2010, 2013], as well as investors' asset positions in models of financial trade [Grossman and Stiglitz, 1976, 1980].

In these games, agents' beliefs over exogenous fundamentals and the endogenous actions of others play a key role in determining equilibrium outcomes. But where do these beliefs come from and how are they formed? In this paper we investigate the endogenous acquisition of information within games of strategic interaction. We ask two questions. First, what properties of the agents' information acquisition costs guarantee that an equilibrium of the game does or does not exhibit non-fundamental volatility? Second, what properties of the agents' information acquisition costs guarantee that an equilibrium is constrained efficient?

A substantial literature has studied non-fundamental volatility and efficiency in exogenous information environments, that is, when agents form their beliefs based on an exogenously-given set of signals. In these environments, aggregate equilibrium outcomes are typically driven by both fundamentals as well as shocks orthogonal to fundamentals. The latter, "non-fundamental," shocks are rationalized as the result of errors in publicly-observed signals, or more generally, correlated errors in beliefs—standard components of generic information structures [Bergemann and Morris, 2013]. In fact, a robust positive prediction of these games is that the greater the strategic complementarity in actions (i.e. the greater the incentive to align actions with others), the greater the role of non-fundamental volatility in equilibrium outcomes [Morris and Shin, 2002].

Non-fundamental volatility can help explain short-run fluctuations in asset prices or business cycle activity that appear to be driven by "market sentiment" or "animal spirits" [Angeletos and La'O, 2010, 2013]. It is tempting to give a normative interpretation to these positive predictions and assume that non-fundamental volatility is a sign of inefficiency. Angeletos and Pavan [2007] demonstrate that, with exogenously given information structures, such conclusions are unwarranted.

But it is not obvious that non-fundamental volatility should be expected when agents acquire their information endogenously. It is also not clear whether agents will acquire their information efficiently, nor whether the questions of efficiency and non-fundamental volatility are related. Our paper seeks to address these questions.

This paper. We study a general class of large games of strategic interaction. A continuum of ex-ante identical agents take actions under incomplete information. Each agent has the incentive to align her own action with exogenous fundamentals as well as with the endogenous mean action. Agents must therefore form beliefs over these objects.

We allow information to be acquired endogenously. In particular, we adopt the rational inattention approach to costly information acquisition proposed by Sims [2003]. However, relative to the standard rational inattention framework, we make two important departures.

First, we do not assume the information acquisition cost is mutual information—the typical cost function introduced by Sims [2003] and used widely throughout the rational inattention literature. We instead consider a more general class of cost functions: costs that are "posterior-separable" in the terminology of Caplin, Dean, and Leahy [2019]. This class nests the standard mutual information cost function as a special case. However, it also includes other cost functions that have been proposed in the literature, including LLR cost function of Pomatto et al. [2018] and the Fisher information cost function proposed by Hébert and Woodford [2020].

Second, we do not restrict agents to acquiring information only about exogenous fundamentals. Instead, we follow Denti [2019] and allow the rationally-inattentive agents in our model to learn not only about exogenous states but *also* about endogenous mean actions. This modeling choice is motivated by the agents' incentives to align their actions with the endogenous mean action, and hence learn about it.

Thus, in our framework, a continuum of rationally-inattentive agents acquire information in a relatively unrestricted way about payoff-relevant states, payoff-irrelevant states, and endogenous mean actions. The payoff-irrelevant states are the potential source of "non-fundamental" volatility in our framework; they play a role similar to "noisy public signals" in exogenous information environments.

Within this context, we answer the two questions posed above. What properties of the agents' information cost structures guarantee that an equilibrium of the game does or does not exhibit non-fundamental volatility? And, what properties of the agents' information cost structures guarantee that an equilibrium exists that is or is not constrained efficient?[1] Through this analysis, we answer a third question: are these properties related? That is, is non-fundamental volatility synonymous with inefficiency?

We begin our analysis with a leading example: the classic linear-quadratic-Gaussian setting. In this setting, with either the Fisher information cost function or mutual information, agents optimally receive Gaussian signals. As a result, one might be tempted to think the two cost functions make identical predictions. We show that this is false—with mutual information, there is zero non-fundamental volatility and an efficient equilibrium exists, whereas with the Fisher information cost function the equilibrium exhibits non-fundamental volatility and is

---

[1]Angeletos and Sastry [2019] consider the related question of what properties of information costs are sufficient to ensure efficiency in a Walrasian context.

inefficient.

This is our first indication that the cost function matters. But it leaves open the question of what is it about these cost functions that lead to such divergent predictions. Our general model and analysis is focused on answering this question, and showing that non-fundamental volatility and efficiency are driven by separate properties.

Partial Monotonicity and Partial Invariance.   Whether or not a cost function leads to non-fundamental volatility or inefficiency depends on two key properties of the agents' information cost structures: what we call "partial monotonicity" and "partial invariance."

We introduce and define partial monotonicity and partial invariance as properties of posterior-separable cost functions, and in particular the divergences that define these cost functions. Loosely speaking, a divergence can be thought of as a measure of the "distance" between the prior and posterior. Partial monotonicity and partial invariance describe how this divergence responds to different transformations of the prior and posterior.

Suppose an agent is uncertain about a multi-dimensional aggregate state, and receives a signal that moves her posterior beliefs "away from" her prior in some dimension of the aggregate state. This signal is, in a sense, more informative than another signal that leaves posterior beliefs close to the prior in that dimension. This idea leads to a notion of monotonicity: a divergence is *monotonic* if the cost decreases as we make the posterior more like the prior in some dimension. But note that a divergence might be monotonic in some dimensions but not others—this is essentially our definition of "partial monotonicity."

Take for example, a two-dimensional state space, $s \in S$ and $r \in R$. We will say that a cost function is *monotonic in R* if the cost decreases when we replace the posterior's conditional distribution of r given s with the prior's conditional distribution of r given s. That is, if we make posteriors more like the prior in this one particular way, this decreases costs.

We define another concept we call "partial invariance." Take again our two-dimensional state space $s \in S$ and $r \in R$. We will say that a cost function is *invariant* in R if, for any prior and posterior with the same conditional distributions of r given s, the divergence between them is the same regardless of what that conditional distribution is. That is, if their conditional distributions of r given s are the same, only their marginal distributions on s matter for the information cost.

The advantage of defining partial monotonicity and partial invariance in this way is that it allows us to consider cost functions that are monotonic in one dimension, but not monotonic in others, or invariant in one dimension, but not invariant in others.

The forms of partial monotonicity and partial invariance that we introduce are generalizations of the invariance concept described in the literature on information geometry [Chentsov, 1982, Amari and Nagaoka, 2007] and that derived from the behavioral "invariance-

3

under-compression" axiom of Caplin, Dean, and Leahy [2019]. Invariance in this sense has been applied to particular economic applications by Hébert [2018] and Hébert and Woodford [2019].

The invariance and monotonicity properties described by these authors indicate whether a cost function is invariant or monotone with respect to *all possible* dimensions of the state space. In contrast, we show how the answers to the questions posed in this paper relate to invariance and monotonicity of the cost function with respect to *specific* dimensions of the state space. That is, the properties of equilibria in this class of games—namely, non-fundamental volatility and constrained efficiency—depend on the partial monotonicity and partial invariance properties of information costs that we define.

Results. In our framework agents may acquire information about payoff-relevant states, payoff-irrelevant states, and endogenous aggregate actions. We consider information costs that may be monotone or invariant in any, all, or none of those dimensions.

We first ask: under what conditions does there exist an equilibrium of the game that exhibits zero non-fundamental volatility? We find that monotonicity of the cost function in payoff-irrelevant states is necessary and sufficient to ensure the existence of such equilibria for all payoff functions, and generically necessary. The intuition behind this result is that with such a cost function, it is always cheaper for agents to condition their signals on the payoff-relevant states and aggregate actions than to condition their signals on the payoff-irrelevant states. But this form of monotonicity rules out "public signals," which are by definition not directly payoff-relevant but more easily observed than the payoff-relevant state itself. As a result, it prevents all shocks orthogonal to fundamentals from playing a role in equilibrium.

We then ask: under what conditions does an equilibrium exist that is constrained efficient? We first provide necessary and sufficient conditions for efficiency when information is exogenous, extending prior results by Angeletos and Pavan [2007] on efficiency in the use of information to our more general setting. We then show that, provided there is efficiency in the use of information, invariance of the cost function in endogenous aggregate actions is sufficient and generically necessary for the existence of a constrained efficient equilibrium.

The intuition behind this result is the following. When cost functions are not invariant in the aggregate action, agents' actions affect the ease with which other agents acquire information. For example, with the Fisher information cost function in our linear-quadratic example, if agents take more extreme actions in response to their signals, the endogenous aggregate action becomes less costly to observe. This is an externality—the planner would like to encourage more extreme actions to reduce information acquisition costs.

Finally, we answer the third question: are the answers to our first two questions related? Our results make evident that separate properties of the information cost determine whether or not the equilibrium exhibits non-fundamental volatility and whether or not it is efficient. By

4

precisely defining these properties—namely, partial monotonicity and partial invariance—we characterize the relationship between information costs and the properties of equilibria.

Mutual information, the standard cost function used in the rational inattention literature, is monotone and invariant in all dimensions; as a result, this cost leads to zero non-fundamental volatility and efficiency. However, alternative cost functions such as the Tsallis entropy cost [Caplin, Dean, and Leahy, 2019] and the Fisher information cost [Hébert and Woodford, 2020] have been proposed because they are better able to match observed experimental behavior (see, e.g., Dean and Neligh [2019]). The Fisher information cost results in both non-fundamental volatility and inefficiency, while the Tsallis entropy cost leads to non-fundamental volatility but an equilibrium remains efficient.

Related Literature.   A large literature has studied the positive and normative implications of large games of strategic interaction and incomplete information, and applied these insights to questions in macro, finance, and industrial organization (see Angeletos and Lian [2016] for a recent survey). Much of this literature assumes linear-quadratic payoffs, Gaussian priors, and exogenously specified Gaussian signals about exogenous states.

Several authors (e.g. Hellwig and Veldkamp [2009], Myatt and Wallace [2012], Colombo, Femminis, and Pavan [2014], Pavan [2016]) endogenize information acquisition in the linear-quadratic setting, allowing agents to choose the precision with which they observe an exogenously specified set of Gaussian signals about exogenous states. In these papers, the presence or absence of non-fundamental volatility depends on the assumed correlation structure of the exogenously given signals (and, with precision choice across multiple signals, agents' incentives to coordinate).

Other authors (e.g. Mackowiak and Wiederholt [2009], Paciello and Wiederholt [2014], Afrouzi [2019]) also endogenize information acquisition in this setting, but follow the rational inattention approach of Sims [2003]. These models do not assume a particular set of available signals; instead, agents can choose any signal structure, subject to a cost described by mutual information. With quadratic payoffs, Gaussian priors, and mutual information costs, the agent's optimal signal is a Gaussian signal about economic fundamentals. As a result, equilibria exhibit zero non-fundamental volatility.

Our paper also follows the rational inattention approach, but generalizes away from the mutual information cost function. As a result, we are able to accommodate non-fundamental volatility, building a bridge between these two seemingly distinct approaches.

Our study of efficiency builds on the work of Angeletos and Pavan [2007] and Colombo, Femminis, and Pavan [2014]. Angeletos and Pavan [2007] study the question of constrained efficiency in the class of linear-quadratic games with exogenous information structures. We extend their results to multi-dimensional settings with general payoff functions, and obtain

necessary and sufficient conditions for efficiency in the use of information in our game.

However, we shut down a key channel present in Angeletos and Pavan [2007] by assuming that only the cross-sectional mean of actions, but not the cross-sectional variance, enters payoffs. This is important for understanding our results vis-à-vis Colombo, Femminis, and Pavan [2014]. Colombo, Femminis, and Pavan [2014] study the efficiency of information acquisition within the Angeletos and Pavan [2007] linear-quadratic setting; they show that efficiency in the use of information does not guarantee efficiency in the acquisition of information because the dispersion of actions enters payoffs—an externality not internalized by agents. With this channel shut down in our game, their results imply that efficiency in the use of information should guarantee efficiency in the acquisition of information.

Instead, we find that efficiency in the use of information is not sufficient for efficiency in the acquisition of information because of a different externality: if agents' actions affect other agents' information costs. The inefficiency we highlight is closely related to the informational externality that arises when agents observe exogenous signals about endogenous objects such as prices, as in Laffont [1985], Angeletos and Pavan [2009], Amador and Weill [2010], Vives [2017], Angeletos, Iovino, and La'O [2020].

In a similar vein, Angeletos and Sastry [2019] allow rationally inattentive agents to learn from prices in a Walrasian context with complete markets over states and signal realizations. They find that invariance of the information cost is sufficient to ensure that a planner cannot improve allocations by sending a message that reduces information costs.[2]

In considering a large class of possible information costs in a rational inattention problem, and not just mutual information, we build on the work of Caplin, Dean, and Leahy [2019], Pomatto et al. [2018], and Hébert and Woodford [2020]. Our focus on games with agents who can acquire information about the endogenous actions of other agents builds on Denti [2019]. We adapt his approach to static games with a continuum of players. Relative to Denti [2019], the "largeness" feature of our class of games permits a simpler definition of equilibrium, which is essentially Bayesian Nash equilibrium in a static, simultaneous-move game. Our definition of equilibrium can also be thought of as the limit of the dynamic process of strategic information acquisition Denti [2019] introduces.

We begin with a linear-quadratic-Gaussian example to illustrate the role that the information cost plays in determining whether or not equilibria exhibit non-fundamental volatility and are efficient. Following this, we introduce and analyze the general class of models.

---

[2]That is, our paper and Angeletos and Sastry [2019] consider different planner's problems (among other differences). We employ the constrained efficiency concept in Angeletos and Pavan [2007] and Colombo, Femminis, and Pavan [2014] for abstract games in which the planner may only control the action functions and information choices of the players. Angeletos and Sastry [2019] consider a planner who can send messages to the agents, who can in turn learn about the content of the message (as opposed to prices or states directly). They ask a different question: whether in markets the price function is an efficient conveyer of information in the sense of Hayek [1945].

2 A Linear-Quadratic-Gaussian Example

In this section we use a linear-quadratic-Gaussian example to illustrate the impact that information costs can have on non-fundamental volatility and efficiency.

We consider a simple, stylized beauty-contest game, similar to the one studied in Morris and Shin [2002]. A unit-mass continuum of ex-ante identical agents indexed by $i \in [0, 1]$ attempt to choose an action to track both a fundamental state and the average action of the other agents. Let $s \in \mathbb{R}$ be the payoff-relevant dimension of the state, and let $a^i \in \mathbb{R}$ be the action of agent i. We define the aggregate action of all agents as $\bar{a} = \int_0^1 a^i di$. The payoff of an agent who takes action a^i when the aggregate action is $\bar{a}$ and the fundamental state is s is given by

$$u(a^i, \bar{a}, s) = -(1 - \beta)(a^i - s)^2 - \beta(a^i - \bar{a})^2 \tag{1}$$

where $\beta \in (0, 1)$ is a scalar.

The first component of (1) is a quadratic loss in the distance between the agent's action and the exogenous fundamental s; the second component is a quadratic loss in the distance between the agent's action and the aggregate action. The scalar β governs the extent of strategic interaction in this game; for this example, we assume $\beta > 0$ so that actions are "strategic complements." We also assume that $\beta < 1$ so that there is a unique pure-strategy Nash equilibria of this game under complete information, in which every agent takes the same action: $a^i = \bar{a} = s, \forall i$.

In games with exogenous information, agents receive costless signals about the aggregate state. Let $\omega^i \in \mathbb{R}$ be the realization of agent i's signal. With the quadratic payoffs in (1), each agent's optimal action is chosen according to the linear best response function

$$a^i(\omega^i) = \mathbb{E}[(1 - \beta)s + \beta\bar{a}|\omega^i], \tag{2}$$

where $\mathbb{E}[\cdot|\omega^i]$ denotes the agent's expectations conditional on ω^i.

Our focus is on games with endogenous information acquisition by rationally inattentive agents. For this example, we will consider two different information cost functions: mutual information and the Fisher information cost function. With linear-quadratic payoffs and Gaussian priors, both of these cost functions result in agents optimally choosing to observe a one-dimensional Gaussian signal [Hébert and Woodford, 2020]. As a result, our example falls into the tractable class of linear-quadratic-Gaussian games. However, we will see that these cost functions lead to different conclusions about the existence of non-fundamental volatility and whether equilibria are efficient.

We begin by discussing the issue of non-fundamental volatility, and then turn to the question of efficiency.

2.1 Non-Fundamental Volatility

We allow agents to flexibly acquire signals of both the payoff-relevant state, $s \in \mathbb{R}$, as well as a payoff-irrelevant state $r \in \mathbb{R}$, subject to a cost. We introduce the payoff-irrelevant states as potential sources of non-fundamental volatility in equilibrium. Let $x = (s, r)^T$ denote the aggregate state vector. To preserve the linear-quadratic-Gaussian structure of the model, we assume that all agents have a common Gaussian prior over (s, r), with a prior mean of zero for both variables and variance-covariance matrix of Σ. That is, $x \sim \mathcal{N}(0, \Sigma)$.

We consider equilibria in which the aggregate action $\bar{a}$ is a linear function of the aggregate state. In particular, we guess and verify the functional form

$$\bar{a} = \bar{\alpha}_s s + \bar{\alpha}_r r, \tag{3}$$

for some constants $\bar{\alpha}_s$ and $\bar{\alpha}_r$. Linearity of the aggregate action preserves the linear-quadratic Gaussian nature of the individual agents' problem. Under this assumption, the agent's best response after observing the signal ω^i is

$$a^i(\omega^i) = \mathbb{E}[(1 - \beta + \beta\bar{\alpha}_s)s + \beta\bar{\alpha}_r r | \omega^i] = \mathbb{E}[\psi^T x | \omega^i],$$

where ψ is a column vector of endogenous constants given by

$$\psi \equiv (1 - \beta + \beta\bar{\alpha}_s, \beta\bar{\alpha}_r)^T. \tag{4}$$

Consider now the individual agents' problem in this game. With both mutual information and the Fisher information cost function, agents will optimally choose to receive a one-dimensional[3] Gaussian signal,

$$\omega^i = \lambda^T x + \sigma \epsilon^i, \tag{5}$$

where λ is a vector describing what the agent chooses to learn about, σ^2 is the variance of the agent's signal, and ϵ^i is a standard normal shock, i.i.d. across agents. That is, both cost functions deliver Gaussian signals of the form in (5), but as we will show shortly, the two cost functions have different implications for the agent's optimal choices of (λ, σ).

Given such a signal, the agent's optimal action strategy follows from Bayesian updating (see e.g. Hébert and Woodford [2020]):

$$a^i(\omega^i) = \alpha\omega^i, \qquad \alpha = \underbrace{\frac{\psi^T \Sigma \lambda}{\lambda^T \Sigma \lambda}}_{\text{"beta" between } \lambda^T x \text{ and } \psi^T x} \underbrace{\frac{\sigma^{-2}}{\sigma^{-2} + (\lambda^T \Sigma \lambda)^{-1}}}_{\text{update on } \lambda^T x}. \tag{6}$$

---

[3]The one-dimensional nature of the signal is a consequence of the standard result in rational inattention problems that it is without loss of generality to equate signals with recommended actions.

The agent's unconditional expected payoff is $-\psi^T \Sigma^\omega(\lambda, \sigma)\psi$,[4] where $\Sigma^\omega(\lambda, \sigma)$ is the posterior variance-covariance matrix given by

$$(\Sigma^\omega(\lambda, \sigma))^{-1} = \Sigma^{-1} + \sigma^{-2}\lambda\lambda^T. \tag{7}$$

Due to the structure of the agents' optimal signals, this matrix is independent of the realized state and signal realizations, but depends on the agent's choice of (λ, σ).

Following Sims [2003] and Hébert and Woodford [2020], under both mutual and Fisher information, the cost of information acquisition can be written a function of the prior and posterior variance-covariance matrices Σ and $\Sigma^\omega(\lambda, \sigma)$. In either case, the agent chooses the parameters (λ, σ) of (5) subject to feasibility. The feasibility set is given by $\Gamma \equiv \{\lambda, \sigma : |\lambda| = |\psi|, \sigma \geq 0\}$. The first constraint on this set is a normalization. Expression (7) makes clear that scaling both λ and σ by the same constant does not change the posterior variance. To simplify notation, for our analysis of non-fundamental volatility we adopt the convention that $|\lambda| = |\psi|$. The second constraint that $\sigma \geq 0$ is called a "no-forgetting" constraint by Van Nieuwerburgh and Veldkamp [2010], and reflects the fact that the agent cannot reduce information costs by forgetting information.

Before continuing, we should note that with both of these cost functions, it is possible that no optimum over (λ, σ) exists, and the optimal policy is to receive a completely uninformative signal ($\sigma \to \infty$). This leads to an equilibrium in which all agents choose an action equal to their prior mean about s. In our results, we will assume that information costs are sufficiently small to rule out this possibility. We next characterize what kind of information the agents choose to gather.

Consider first the mutual information case.[5]

Example 1. *With mutual information, the problem of the agent is*

$$\min_{(\lambda, \sigma) \in \Gamma} \psi^T \Sigma^\omega(\lambda, \sigma)\psi + \theta\left[\ln(\det((\Sigma^\omega(\lambda, \sigma))^{-1})) - \ln(\det(\Sigma^{-1}))\right] \tag{8}$$

subject to (7), with $\theta > 0$. Substituting in (7), this simplifies to

$$\min_{(\lambda, \sigma) \in \Gamma} \psi^T \Sigma\psi - \sigma^{-2}\frac{(\psi^T\Sigma\lambda)^2}{1 + \sigma^{-2}\lambda^T\Sigma\lambda} + \theta\ln(1 + \sigma^{-2}\lambda^T\Sigma\lambda). \tag{9}$$

The first component of the agent's objective function is her unconditional expected payoff. The second component is the agent's cost of information acquisition under mutual information.

---

[4]The unconditional expected payoff is $-\psi^T\Sigma^\omega(\lambda, \sigma)\psi$ plus a constant. However, the constant term is unaffected by the individual agent's choices. We ignore it when considering the agent's optimal choice of (λ, σ), but must account for this constant when considering efficiency.

[5]See the proof of Proposition 1 for a more detailed derivation.

This is given by the difference in the log-determinant of the (inverse) posterior and prior variance-covariance matrices. The parameter $\theta > 0$ scales the cost of information.[6]

We contrast this problem to the case with the Fisher information cost function.

Example 2. *With Fisher information, the problem of the agent is*

$$\min_{(\lambda,\sigma)\in\Gamma} \psi^T \Sigma^\omega(\lambda,\sigma)\psi + \theta\left[\mathrm{tr}((\Sigma^\omega(\lambda,\sigma))^{-1}) - \mathrm{tr}(\Sigma^{-1})\right], \tag{10}$$

subject to (7), where $\mathrm{tr}(\cdot)$ *is the trace operator, with* $\theta > 0$. *Substituting in (7), this simplifies to*

$$\min_{(\lambda,\sigma)\in\Gamma} \psi^T \Sigma\psi - \sigma^{-2}\frac{(\psi^T\Sigma\lambda)^2}{1+\sigma^{-2}\lambda^T\Sigma\lambda} + \theta\sigma^{-2}|\lambda|^2. \tag{11}$$

The first component is again the agent's unconditional expected payoff. The second component is her cost of information acquisition under Fisher information. This given by the difference in the trace of the (inverse) posterior and prior variance-covariance matrices.

Therefore, the *only* difference between the mutual information and Fisher information cases is the functional form of information cost: the trace vs. the log-determinant. It is this difference that leads to distinct predictions for the agent's optimal choice of signal structure.

To understand the agent's optimal signal structure in either case, let $\tilde{a}^i \equiv \psi'x$ denote the agent's optimal action under *complete* information for a given ψ. This is the action the agent would choose if she herself faced no cost of information acquisition.

Under mutual information, the agent optimally receives a one-dimensional signal that *directly corresponds* to her optimal complete info action, $\tilde{a}^i$.[7] That is, the agent chooses

$$\lambda^* = \psi,$$

and as a result her signal is an unbiased, noisy version of $\tilde{a}^i$.

In contrast, under Fisher information, the agent optimally receives a one-dimensional signal that *maximally covaries* with $\tilde{a}^i$ under the resulting posterior.[8] That is, the agent's optimal choice of λ satisfies the fixed point:

$$\lambda^* \in \mathrm{argmax}_{\lambda:|\lambda|=|\psi|}\psi^T\Sigma^\omega(\lambda^*,\sigma^*)\lambda. \tag{12}$$

Both results are intuitive. It certainly seems logical to learn only about the optimal action $\tilde{a}^i$ and ignore everything else. On the other hand, it also seems perfectly natural to receive a signal that maximally covaries with $\tilde{a}^i$.

---

[6]Agents will find it optimal to gather some information if their uncertainty about the optimal action is sufficiently large relative to the parameter θ. See the proofs of Propositions 1 and 2 for precise conditions.

[7]This follows from the invariance properties of mutual information, but can also be shown using the first-order condition from (9).

[8]See Hébert and Woodford [2020]. This can also be shown using the first-order condition from (11).

In either case, the agent's optimal action is linear in her signal by (6), and her signal is linear in the underlying state (5). Together, and aggregating across individuals, this implies that the aggregate action is indeed linear in the state:

$$\bar{a} = \alpha^*(\lambda^*)^T x, \tag{13}$$

thereby verifying our guess in (3). Finally, in order for (13) to coincide with (3), the equilibrium vector $(\bar{\alpha}_s, \bar{\alpha}_r)^T$ must satisfy

$$(\bar{\alpha}_s, \bar{\alpha}_r)^T = \alpha^*\lambda^*, \tag{14}$$

and is therefore proportional to the agent's optimal choice of λ^*. That is, a linear equilibrium of this game is a collection of parameters $(\bar{\alpha}_r, \bar{\alpha}_s, \psi, \lambda, \sigma, \alpha)$ satisfying (4), (6), and (14), with (λ, σ) chosen optimally given ψ. This leads to the following result.

Proposition 1. *(i) Under mutual information, there exists a linear equilibrium. Any such equilibrium features zero non-fundamental volatility: $\bar{\alpha}_r = 0$.*

(ii) With the Fisher information cost, there exists a linear equilibrium. If r and s are correlated under the prior and θ is sufficiently small, any such equilibrium features non-fundamental volatility, $\bar{\alpha}_r \neq 0$.

Proof. See the appendix, 11.1. □

Because agents choose different types of signal structures across the two cases, this leads to distinct equilibrium properties. In the case of mutual information, the agent's optimal signal is one that directly tracks her complete-information action $\tilde{a}^i$, but ignores everything else. If all agents choose $\lambda^* = \psi$, in equilibrium ψ must satisfy the fixed point:

$$\psi = (1 - \beta)e_1 + \beta\alpha^*\psi.$$

where $e_1 = (1,0)^T$ and α^* is a scalar function strictly less than one by equation (6). It immediately follows that in any equilibrium, the second element of ψ must be zero; hence, $\bar{\alpha}_r = 0$. Therefore, with mutual information, the equilibrium aggregate action does not depend on the payoff-irrelevant state r.

In the case of Fisher information, the agent chooses a signal to maximize the covariance between the signal ω^i and optimal action $\psi^T x$. If s and r are correlated (if the off-diagonal elements of Σ are non-zero), then this immediately rules out equilibria with zero non-fundamental volatility. To see this, suppose by contradiction there exists an equilibrium in which $\bar{\alpha}_r = 0$, so that the second element of ψ is zero. In this case, as long as s and r are correlated, the agent chooses a signal such that the second element of λ^* is non-zero: this choice maximizes the covariance between her signal and her complete-information action (see (12)). But if the second element of λ^* is non-zero, then by (14), $\bar{\alpha}_r$ must be non-zero, a contradiction.

11

We conclude that all linear equilibria in the case of Fisher information generically feature non-fundamental volatility. The only cases in which $\bar{\alpha}_r$ may be zero in equilibrium are the non-generic cases in which s and r have zero correlation, or when no information is gathered in equilibrium.

Mutual information and Fisher information therefore generate starkly different predictions for the existence of non-fundamental volatility in equilibrium in the simple beauty-contest game. The distinction stems from what these costs imply for agents' optimal signal structures. In the mutual information case, agents track their optimal action directly, but ignore everything else—it would be costly to do otherwise. When all agents behave in this way, in equilibrium there is no room for the payoff-irrelevant state r to affect equilibrium actions.

In contrast, in the Fisher information case, agents instead find it optimal to receive a signal that depends at least somewhat on r, as this maximizes covariance with s. That is, it is cheaper for agents to learn about s by partially observing r, rather than learn about s directly. This optimal cost-saving behavior is what opens the door for variation in r to affect individual actions and thereby, in equilibrium, aggregate actions.

One possible interpretation of r in this context is a noisy public signal. In exogenous information environments, agents learn through costless public signals and base their actions upon them. Common errors in these signals lead to variation in aggregate actions that is orthogonal to fundamentals. In the Fisher information case, r plays the exact same role: learning about r is not costless, but it is a "cheaper" way of learning about s than learning about s alone.

2.2 Constrained Efficiency

Next, we consider the question of constrained efficiency in the beauty contest game. For simplicity, we abstract from the payoff-irrelevant states r in our previous example and assume $s \in \mathbb{R}$ is the only exogenous aggregate state.

To discuss efficiency in the context of our beauty contest game, we proceed in three steps. The first step is to ask whether the game is constrained efficient under exogenous information, that is, when agents cannot choose their information structure. The second step is to ask whether the game is constrained efficient under endogenous information when agents may acquire information about the exogenous aggregate state s. The third step is to ask whether the game is constrained efficient under endogenous information when agents can learn not only about the exogenous aggregate state s but also about the endogenous aggregate action $\bar{a}$.

Step 1. We begin by asking first whether the equilibrium is constrained efficient under exogenous information. Agents receive noisy signals $\omega^i \sim N(s, \sigma^2)$ about $s \in \mathbb{R}$.

Consider an equilibrium in which the aggregate action is a linear function of the state, $\bar{a} = \bar{\alpha}_s s$. The agent chooses a strategy $a^i(\omega^i)$ in order to maximize her expected payoffs,

$$V(\sigma, \bar{\alpha}_s) = \max_{a^i(\omega^i)} \mathbb{E}[u(a^i(\omega), \bar{\alpha}_s s, s)|\omega^i] \tag{15}$$

The agent's first-order condition to this problem is what gives rise to the linear private best response function reported in (2). A symmetric equilibrium, then, is an individual strategy $a(\omega)$ (the same for all agents) and an aggregate action coefficient $\bar{\alpha}_s$, such that optimality conditions (2) hold along with, by the law of large numbers,

$$\bar{\alpha}_s s = \mathbb{E}[a\left(\omega\right)|s] \qquad \forall s \in \mathbb{R} \tag{16}$$

Can a planner who controls how each agent responds to her own signal, but is unable to share information across agents, improve welfare relative to the non-cooperative equilibrium? To answer this question, we follow Angeletos and Pavan [2007] and solve a constrained planner's problem. The constrained planner chooses strategies $a(\omega)$ in order to maximize expected utility across all agents,

$$W(\sigma) = \max_{a(\omega)} \int_i \mathbb{E}\left[u(a(\omega), \bar{\alpha}_s s, s)\right] di \tag{17}$$

subject to the constraint that the aggregate action $\bar{\alpha}_s s$ satisfies (16). Taking first-order conditions with the linear-quadratic payoffs of (1), we find that the efficient "best response" function dictated by the planner is given by

$$a(\omega) = \mathbb{E}[(1 - \beta)s + \beta\bar{\alpha}_s s|\omega],$$

and thereby coincides exactly with the agent's private best response function in (2).

We conclude that in this particular game, by construction, the equilibrium use of information is efficient. See Angeletos and Pavan [2007] and Section 7 of this paper for a detailed discussion of the intuition behind this result.

Step 2. We now ask whether the equilibrium is constrained efficient under endogenous information acquisition. We begin by allowing agents to choose the standard deviation of their signals, σ, subject to a cost. Let $C(\sigma^i)$ be the cost of receiving a signal with standard deviation σ^i. This cost could be based on mutual information, the Fisher information cost, or any other cost—the details will not matter for our argument.

Consider the individual agent's problem. Each agent takes $\bar{\alpha}_s$ as exogenous, and chooses both an action strategy α^i and a standard deviation σ^i. Fixing the agent's choice of σ^i, her choice of α^i is exactly as in (15). The agent therefore chooses σ^i to maximize her private value minus her information cost,

$$\sigma^i \in \mathrm{argmax}_{\sigma \geq 0} V(\sigma, \bar{\alpha}_s) - C(\sigma).$$

Can a planner who controls both how much information an agent acquires and how each agent responds to her own signal improve welfare relative to the non-cooperative equilibrium? Consider a planner that chooses a pair (α^*, σ^*) to maximize expected utility across all agents, less their information cost.

Again, fix the planner's choice of σ^* and observe that the problem of choosing α^* is exactly the planner's problem with exogenous information, (17). The planner therefore chooses σ^* to maximize the social value minus the information cost,

$$\sigma^* \in \operatorname{argmax}_{\sigma \geq 0} W(\sigma) - C(\sigma).$$

It follows almost immediately that the planner's optimal (α^*, σ^*) is also an equilibrium. To see this, again fix σ and consider the inner problem of both the agent and the planner. We know that the planner's optimal choice of α coincides with the equilibrium α in this "σ-subgame." This in fact is true for any σ-subgame, and is simply a restatement of the result that the equilibrium use of information is efficient. It follows that for any σ, the social value of σ forms an upper envelope of the private value:

$$W(\sigma) = \max_{\bar{\alpha}_s} V(\sigma, \bar{\alpha}_s). \tag{18}$$

That is, even if we were to ignore the mean-consistency requirement (16), the $\bar{\alpha}_s$ that maximizes the private value is the same $\bar{\alpha}_s$ that solves the planner's problem and thereby satisfies mean consistency.

But observe from this equation that the planner's optimal choice of σ^* coincides with the agent's best response to $\bar{\alpha}_s^* \in \arg\max_{\bar{\alpha}_s} V(\sigma^*, \bar{\alpha}_s)$; therefore the solution to the planner's problem is an equilibrium. We conclude that in this particular game, efficiency in the use of information implies efficiency in the acquisition of information when agents receive signals only about exogenous states. Note that this result relies on the fact that only the mean action, and not higher moments of the action distribution, enters the agents' payoffs in (1); see Colombo, Femminis, and Pavan [2014] for details.

Step 3. We now allow agents to learn not only about the exogenous fundamental s but also about the endogenous aggregate action of other agents, $\bar{a}$, as in Denti [2019].

To mirror our previous discussion of non-fundamental volatility, let $x = (s, \bar{a})^T$ denote the vector of objects the agents can learn about. To preserve the linear-quadratic-Gaussian structure of the model, we continue to assume $s \sim \mathcal{N}(0, \sigma_0^2)$ and consider linear equilibria of the form $\bar{a} = \bar{\alpha}_s s$. With either mutual or Fisher information, it remains the case that agents optimally receive a one-dimensional Gaussian signal ω^i of the form given by (5).

Unlike the previous step, agents may now learn about the aggregate action $\bar{a}$. Consequently, as in our analysis of non-fundamental volatility, agents have choices about both the noise in

their signal (σ) and what they learn about (λ). In this case, because the aggregate action $\bar{a}$ is part of the state vector x, the payoff-relevant dimension of the state is exogenous, $\psi = (1 - \beta, \beta)^T$. Note also that $x = \chi s$, where we define $\chi = (1, \bar{\alpha}_s)^T$. As a result, x is Gaussian, $x \sim \mathcal{N}(0, \Sigma)$ with the degenerate variance-covariance matrix $\Sigma = \sigma_0^2 \chi \chi^T$.

Let us now fix the agent's choice of σ, and consider the optimal choice of λ under both mutual information and the Fisher information cost. For this analysis, we adopt the normalization that $\chi^T \lambda = 1$, and consider the optimal choice of λ in (8) and (10) from the previous section (because these formulas are valid even with a degenerate prior). With this normalization, the information cost depends on the choice of λ, but the expected payoff component of these two equations does not.

With mutual information, as discussed in the previous section, an optimal choice of λ is proportional to the payoff-relevant dimension of the state space. With our normalization, this is $\lambda^* = (\psi^T \chi)^{-1} \psi$. In contrast, with the Fisher information cost, we can see from the first-order condition of (10) that $\lambda^* = (\chi^T \chi)^{-1} \chi$. Again, with mutual information, the agent receives a signal directly about the payoff-relevant dimension of the state, whereas with the Fisher information cost the agent maximizes covariance with the payoff-relevant dimension of the state.

Plugging in these choices for λ^* back into (8) and (10), we see that the cost functions in terms of σ can be written in the mutual information case as

$$C(\sigma) = \theta \ln(1 + \sigma^{-2} \sigma_0^2),$$

and in the Fisher information case as

$$C(\sigma, \bar{\alpha}_s) = \frac{\theta \sigma^{-2}}{1 + \bar{\alpha}_s^2}. \tag{19}$$

Let us now consider a planner who can choose (α^*, σ^*) as in step 2 of our analysis.[9] The following result is then immediate.

Proposition 2. *(i) With mutual information, an optimum of the planner (α^*, σ^*) is also an equilibrium. (ii) With the Fisher information cost, an equilibrium exists, but if θ is sufficiently small any optimum of the planner (α^*, σ^*) is not an equilibrium.*

Proof. See the appendix, 11.2. $\square$

The result for mutual information follows directly from our analysis in step 2 and Colombo, Femminis, and Pavan [2014]. As long as agents use their information efficiently in any σ-subgame, then the private value of information is the same as the social value. As a result, agents behave exactly as the planner would dictate.

---

[9]That is, we do not allow the planner to choose λ^*. This is for expositional clarity and does not affect our results.

This is not the case with the Fisher information cost. In (19), the agent's cost of information acquisition depends on the strategies of others, and in particular it is decreasing in $\bar{\alpha}_s$. Thus, if agents make their actions more sensitive to their signals, this increases the sensitivity of the aggregate action to the aggregate state, which in turn decreases the cost of information acquisition for all agents. Agents do not internalize this effect when making their own individually-optimal decisions. The planner on the other hand takes this externality into account when maximizing welfare, and as a result, dictates a higher α^* and lower σ^* relative to the non-cooperative equilibrium. Note that this externality exists only when agents gather information in equilibrium, hence our assumption that θ is sufficiently small.

The aforementioned externality does not exist when agents can only gather information about exogenous states (step 2). Why does it arise here? Agents in this economy learn from both the fundamental state and the aggregate action. With the Fisher information cost, it is cheaper for agents to observe the aggregate action when it is highly sensitive to the aggregate state, that is, when $\bar{\alpha}_s$ is larger. Fisher information thereby incorporates a scale effect into costs: when aggregate actions are more extreme, they become more salient in the eyes of other agents, and thereby less costly to observe. It is this scale effect on costs that the agents do not internalize. Mutual information, on the other hand, is invariant to these scale effects, and as a result this externality is absent.[10]

2.3 Summary and Layout

We have shown that these two information cost structures have significantly different equilibrium implications in this simple linear-quadratic beauty contest game. Under mutual information, the equilibrium features zero non-fundamental volatility and is constrained efficient. With the Fisher information cost, the equilibrium exhibits non-fundamental volatility and is constrained inefficient. But what is it about these two cost functions that leads to such divergent predictions? We formally address this question in the general framework that follows.

The remainder of this paper is organized as follows. In Section 3 we define the general class of large games that we study. In Section 4 we define equilibria and prove its existence. In Section 5 we introduce and define certain properties of cost functions that we call partial monotonicity and partial invariance. In Section 6 we characterize under what conditions equilibria do or do not feature non-fundamental volatility. In Section 7 we define equilibria and efficiency under exogenous information and characterize under what conditions there is efficiency in the use of information. In Section 8 we show what needed in addition to efficiency in use to obtain efficiency with endogenous acquisition of information. In Section 9 we conclude. All proofs are

---

[10]Angeletos and Sastry [2019] construct a related example in which the variance of prices enters information costs, and show that this leads in their Walrasian setting to multiple, Pareto-ranked equilibria.

provided in the Appendix.

3   The General Game

We study large games of strategic interaction. These games generalize our simple example in several respects. First, agents' action spaces can be multi-dimensional, payoff functions are not necessarily quadratic, and the aggregate action is not necessarily linear in the state variables. Second, we do not restrict ourselves to Gaussian structures: agents can have arbitrary priors. Third, we study a large class of information cost functions which includes mutual information and Fisher information but also includes many others. Fourth, we allow for learning about both payoff-irrelevant states and endogenous actions, rather than consider these things separately as in our simple examples. In all of these respects, our general environment nests the example of the previous section; however, to avoid certain technical complications and simplify our exposition, we assume that the exogenous state space is finite.

3.1   Agents, Actions, and Payoffs

There is a unit mass continuum of agents, indexed by $i \in [0, 1]$. Agent i chooses her action, $a^i \in A \subseteq \mathbb{R}^L$. Let $\bar{a} \in \bar{A} \subseteq \mathbb{R}^L$ be the vector of aggregate actions, defined as the average action chosen by agents:

$$\bar{a} = \int_0^1 a^i di.$$

There is a finite set of exogenous payoff-relevant states, $s \in S$. These states, along with aggregate actions $\bar{a}$, determine the agent's payoffs. Agents have payoff function $u : A \times \bar{A} \times S \to \mathbb{R}$; that is, an agent who takes action $a^i \in A$ in state $s \in S$ when the aggregate action is $\bar{a} \in \bar{A}$ receives payoff $u(a^i, \bar{a}, s)$. Note that individual agents—each of whom is infinitesimal—do not take into account how their own action affects the aggregate action when making their own strategic choices. This assumption is a defining feature of "large games."

We impose the following regularity assumption on the payoff functions and action space.

Assumption 1. *A is non-empty, convex, and compact, and $u(a^i, \bar{a}, s)$ is continuously differentiable on $A \times \bar{A}$ for all $s \in S$.*

Assumption 1 will be sufficient, but not necessary, for our results.[11] In particular, our results could readily be extended to games with finitely many actions.

The last primitives of our environment are the agents' information acquisition technologies, which we describe next.

---

[11]Note that Assumption 1 guarantees the existence of mixed strategy Nash equilibria under complete information in games with continuous actions spaces. See, e.g., Fudenberg and Tirole [1991] theorem 1.2.

3.2 Shocks and Information Acquisition

Shocks and Priors. In addition to the payoff-relevant, or "fundamental," states, $s \in S$, we allow for a finite set of payoff-irrelevant states, $r \in R$.

Agents are endowed with a common prior $\mu_0(s, r)$ over the exogenous states. Let $\mathcal{U}_0 \equiv \Delta(S \times R)$ denote the space of probability measures over the exogenous states, with $\mu_0 \in \mathcal{U}_0$. Note that the payoff-irrelevant states can be independent of the fundamental states, in which case they can be interpreted as pure noise, or correlated, in which case they case they can be interpreted as noisy signals about fundamentals. These states are payoff-irrelevant by definition, but are a potential source of non-fundamental volatility in equilibrium.

We define $\bar{\alpha} : S \times R \to \bar{A}$ as a function mapping exogenous states to an aggregate action. Let $\bar{\mathcal{A}}$ be the space of all such functions.[12] One may think of $\bar{\alpha} \in \bar{\mathcal{A}}$ as the "aggregate strategy," as this function will be determined endogenously by aggregating over the individual agents' strategies.

We will allow agents to learn not only about the exogenous states, but also about endogenous aggregate actions. Agents will optimally choose which objects to pay attention to; in order to facilitate this choice, we specify their prior over the larger $S \times R \times \bar{A}$ space. Let $\bar{\mathcal{U}} \equiv \Delta(S \times R \times \bar{A})$ denote the space of probability measures over this space.

We construct the agents' prior on this larger space from their prior on the exogenous states μ_0 and the aggregate strategy $\bar{\alpha}$ as follows. Let $\phi_{\bar{A}} : \mathcal{U}_0 \times \bar{\mathcal{A}} \to \bar{\mathcal{U}}$ denote a mapping from any pair $\mu_0, \bar{\alpha}$ to its induced probability measure, defined as

$$\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}(s, r, \bar{a}) = \mu_0(s, r)\,\delta(\bar{a} - \alpha(s, r)), \quad \forall s \in S, r \in R, \bar{a} \in \bar{A}, \tag{20}$$

where $\delta(\cdot)$ is the Dirac delta function. We define the space of all probability measures that may be generated on $S \times R \times \bar{A}$ by some pair $(\mu_0, \bar{\alpha}), \mathcal{U} \subset \bar{\mathcal{U}}$, as

$$\mathcal{U} = \left\{ \mu \in \bar{\mathcal{U}} : \exists \quad \mu_0 \in \mathcal{U}_0 \quad \text{and} \quad \bar{\alpha} \in \bar{\mathcal{A}} \quad \text{s.t.} \quad \mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\} \right\}.$$

Given a prior $\mu_0 \in \mathcal{U}_0$ and an aggregate action function $\bar{\alpha} \in \bar{\mathcal{A}}$, the induced prior on the larger space $\mu \in \mathcal{U}$ is given by $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$.

Agents' strategies. We now consider the strategies of the agents. In games of imperfect information with exogenous signals, an individual agent chooses her own action based on the realization of her own signal $\omega^i \in \Omega$, where Ω is a signal alphabet. We assume Ω has a cardinality weakly greater than $\mathbb{R}^L$ (and hence the action space). In these games, an individual agent's pure strategy is a mapping from signals to actions, $\alpha^i : \Omega \to A$. Let $\mathcal{A}$ be the space of all possible pure strategies.

---

[12]We are restricting the aggregate action to be a deterministic function of (s, r). However, r could include elements that are independent of the rest of (s, r) and can be regarded as sunspots.

In games with endogenous information acquisition, agents also choose their signal structure. A signal structure is a conditional probability distribution function

$$\nu^i : S \times R \times \bar{A} \to \Delta\left(\Omega\right).$$

where $\Delta\left(\Omega\right)$ is the space of probability measures on Ω. That is, $\nu^i\left(\omega|s,r,\bar{a}\right)$ denotes agent i's probability of observing signal $\omega \in \Omega$ conditional on $(s,r,\bar{a})$. Let $\mathcal{V}^\Omega$ be the space of all such functions; the superscript indicates the signal alphabet. In what follows, it will also be useful to consider signal structures that depend only on the exogenous states. We define $\mathcal{V}_0^\Omega \subset \mathcal{V}^\Omega$ as the set of signal structures whose conditional probabilities depend only on the exogenous states. That is,

$$\mathcal{V}_0^\Omega = \{\nu \in \mathcal{V}^\Omega : \ \nu^i(\omega|s,r,\bar{a}) = \nu^i(\omega|s,r,\bar{a}') \ \forall s \in S, r \in R, \bar{a}, \bar{a}' \in \bar{A}\}.$$

Note that our setup allows agents to learn about the mean (or aggregate) action, but precludes agents from learning about any other *particular* agent's action.[13] To summarize, with endogenous information acquisition, an individual agent chooses both an action strategy $\alpha^i \in \mathcal{A}$ and a signal structure $\nu^i \in \mathcal{V}^\Omega$ in order to maximize his or her payoffs subject to a cost of information acquisition.[14] We discuss these costs after introducing some notation for signal probabilities and posteriors.

Posterior distributions. Take any signal structure $\nu \in \mathcal{V}^\Omega$ and prior $\mu \in \mathcal{U}$; together these induce a joint distribution on $S \times R \times \bar{A} \times \Omega$. The marginal distribution on Ω associated with this joint distribution is the agent's unconditional probability of observing signal $\omega \in \Omega$,

$$\pi\{\nu,\mu\}\left(\omega\right) = \sum_{s \in S, r \in R} \int_{\bar{A}} \nu\left(\omega|s,r,\bar{a}\right) \mu\left(s,r,\bar{a}\right) d\bar{a}, \tag{21}$$

with $\pi\{\nu,\mu\} \in \Delta\left(\Omega\right)$.

The joint distribution also induces posteriors over the $S \times R \times \bar{A}$ space. The agent's posterior over $(s,r,\bar{a})$ conditional on observing any signal $\omega \in \Omega$, is given by

$$\mu^\omega\{\nu,\mu\}\left(s,r,\bar{a}\right) = \frac{\nu\left(\omega|s,r,\bar{a}\right)\mu\left(s,r,\bar{a}\right)}{\pi\{\nu,\mu\}\left(\omega\right)}, \tag{22}$$

consistent with Bayes' rule and assuming $\pi\{\nu,\mu\}\left(\omega\right) > 0$. Note that, if $\mu \in \mathcal{U}$ (meaning that $\bar{a}$ is deterministic conditional on (s,r) under μ), then $\mu^\omega\{\nu,\mu\} \in \mathcal{U}$ for all $\nu \in \mathcal{V}^\Omega, \omega \in \Omega$. We adopt the convention that, for zero probability signals, posteriors are equal to priors.

We use these objects to define "uninformative" and "informative" signal structures.

---

[13]This choice, made for tractability, is motivated by the "largeness" feature of the game.

[14]By standard arguments in the rational inattention literature, it will be without loss of generality to identify signals with actions, and hence to assume pure as opposed to mixed action strategies.

Definition 1. *A signal structure ν is **uninformative** if for all $\omega \in \Omega$ such that $\pi\{\nu, \mu\}(\omega) > 0$, $\mu^\omega\{\nu, \mu\} = \mu$. A signal structure is **informative** if it is not uninformative, that is, if there exists an ω such that $\pi\{\nu, \mu\}(\omega) > 0$ and $\mu^\omega\{\nu, \mu\} \neq \mu$.*

Thus, informative signal structures are those that move posteriors away from the prior. Armed with these definitions, we next describe the cost of information.

Cost of information acquisition. Agents face a cost of acquiring informative signal structures. We generalize the standard rational inattention setup and define the cost of information acquisition by a function

$$C^\Omega : \mathcal{V}^\Omega \times \mathcal{U} \to \mathbb{R}_+.$$

That is, given a prior μ, an agent i which chooses signal structure $\nu^i \in \mathcal{V}^\Omega$ incurs information costs $C^\Omega(\nu^i, \mu)$, where the superscript Ω indicates the signal alphabet over which the agent chooses its signal structure.

It is without loss of generality to impose the following assumptions on the cost function.

Assumption 2. *For all $\mu \in \mathcal{U}$,*

1. *The cost function $C^\Omega(\nu, \mu)$ is zero if the signal structure ν is uninformative.*

2. *Take $\nu \in \mathcal{V}^\Omega$ and $\nu' \in \mathcal{V}^{\Omega'}$ for two signal alphabets Ω and Ω'. If ν Blackwell-dominates ν' in the sense of Blackwell [1953], then $C^\Omega(\nu, \mu) \geq C^{\Omega'}(\nu', \mu)$.*

3. *The cost function $C^\Omega(\nu, \mu)$ is convex in ν.*

As discussed by Caplin and Dean [2015], and invoking Lemma 1 of Hébert and Woodford [2019],[15] these assumptions are without loss of generality. That is, any behavior that could be observed for a rationally inattentive agent with a cost function not satisfying the second and third conditions could also be observed for a rationally inattentive agent with a cost function satisfying those conditions, and the first condition is a normalization. The intuition for this result comes from the possibility of the agent pursuing mixed strategies over actions conditional on a signal realization and over choices of signal structures.

Our next assumption requires that the information costs we study are continuous. This assumption is phrased in a somewhat technical fashion in order to account for the possibility that the signal space Ω is not a finite set. Observe by the finiteness of $S \times R$ and $\bar{A} \subseteq \mathbb{R}^L$ that $\bar{\mathcal{A}}$ can be viewed as a subset of $\mathbb{R}^{L \times |S| \times |R|}$ and endowed with the standard (Euclidean) topology.

---

[15]Lemma 1 of Hébert and Woodford [2019] allows us to replace the Caplin and Dean [2015] "mixture feasibility" condition with convexity. Hébert and Woodford [2019] prove it in the context of a finite signal alphabet, but nothing in the proof depends on the alphabet being finite.

Assumption 3. *Under the topology of weak convergence on $\mathcal{V}^\Omega$ and $\mathcal{U}$, the cost function C^Ω is continuous in the product topology of $\mathcal{V}^\Omega \times \mathcal{U}$.*

Assumption 2 by itself implies continuity in ν (due to convexity), holding fixed μ. What Assumption 3 adds is the requirement of continuity in (ν, μ).[16]

Posterior-Separability. Finally, we restrict attention to information costs that are "posterior-separable," in the terminology of Caplin, Dean, and Leahy [2019]. Posterior-separable cost functions can be written as the expected value of a divergence between the agents' posterior and prior beliefs. A divergence is a measure of how "close" or "far" two distributions are from one another.[17] To capture the idea that the action of other agents might influence the cost of information, we define these divergences on the larger space of probability measures, $\mathcal{U}$.

Take any signal structure $\nu \in \mathcal{V}^\Omega$ and prior $\mu \in \mathcal{U}$. A posterior-separable cost function is a cost function that can be written as

$$C^\Omega(\nu, \mu) = \int_\Omega \pi\{\nu, \mu\}(\omega)\, D(\mu^\omega\{\nu, \mu\} || \mu) d\omega, \tag{23}$$

where $D : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ is a divergence from the agents' prior μ to posterior μ^ω, convex in its first argument and continuous on $\mathcal{U} \times \mathcal{U}$.[18]

Assumption 4. *The cost function C^Ω is posterior-separable as defined by equation (23), with a divergence D that is continuously differentiable in both arguments.*

Mutual information, the standard rational inattention cost function, is posterior separable. The associated divergence is the Kullback-Leibler divergence, defined in our context as

$$D_{KL}(\mu'||\mu) = \sum_{s \in S, r \in R} \int_{\bar{A}} \mu'(s, r, \bar{a}) \ln\left(\frac{\mu'(s, r, \bar{a})}{\mu(s, r, \bar{a})}\right) d\bar{a}. \tag{24}$$

Other posterior-separable cost functions include the Tsallis entropy cost function proposed by Caplin et al. [2019], versions of the LLR cost function proposed by Pomatto et al. [2018],

---

[16]Note that continuity in (ν, μ) implies continuity in $(\nu, \bar{\alpha})$ holding fixed μ_0.

[17]A divergence is a function of two probability measures that is weakly positive and zero if and only if the measures are identical. Unlike a distance, a divergence does not need to be symmetric, nor does not necessarily satisfy the triangle inequality.

[18]Convexity in the first argument is implied by Assumption 2 and continuity (under the weak topology) by Assumption 3. Also note that we have defined the divergence D on $\mathcal{U}$ rather than the entire space $\bar{\mathcal{U}} = \Delta(S \times R \times \bar{A})$; all priors and posteriors in our problem will remain in $\mathcal{U}$, and therefore it is unnecessary to define the divergence on the entire space. By the finiteness of S, R and the definition of $\mathcal{U}$, elements of $\mathcal{U}$ can be represented as a subset of $\mathbb{R}^{|S| \times |R| \times (L+1)}$, and as a result differentiability for D can be defined in the usual way. Finally, observe that, under the assumption of posterior-separability, the cost function depends on the signal alphabet only through the domain of integration.

and the neighborhood-based cost functions proposed by Hébert and Woodford [2020]. Hébert and Woodford [2019] show that all differentiable information cost functions can be locally approximated by a differentiable posterior-separable cost function.

Hereinafter, we impose Assumption 1 on action spaces and payoffs and Assumptions 2, 3, and 4 on information costs. Before turning to our definition of equilibrium, we briefly discuss one implication of these assumptions.

Recall that the set $\mathcal{U}$ is defined by the property that the aggregate action $\bar{a}$ is deterministic conditional on the exogenous states (s, r). The following lemma demonstrates that, as a result, for any signal structure $\nu \in \mathcal{V}^{\Omega}$, there is another signal structure that does not condition on $\bar{a}$, $\nu' \in \mathcal{V}_0^{\Omega}$, that generates the same signal probabilities and posteriors, and consequently (by Assumption 4) the same information cost.

Lemma 1. *Given any $\nu \in \mathcal{V}^{\Omega}$ and $\mu \in \mathcal{U}$, there is signal structure $\nu' \in \mathcal{V}_0^{\Omega}$ such that $\pi\{\nu, \mu\} = \pi\{\nu', \mu\}$ and, for all $\omega \in \Omega$, $\mu^{\omega}\{\nu, \mu\} = \mu^{\omega}\{\nu', \mu\}$.*

Proof. See the appendix, 11.3. □

The intuition for this result is the following: zero-probability $(s, r, \bar{a})$ events have no impact on either unconditional signal probabilities or posteriors, and therefore do not change information costs. Because these two signal structures result in identical posteriors, we will see that the resulting distribution of actions in our game are also identical, and therefore from an individual agent's perspective the two signal structures are equivalent.

Why, then, do we consider the possibility that agents acquire information about the endogenous actions of the others? The answer is that, despite the fact that it is without loss of generality for an agent to choose a signal structure in $\mathcal{V}_0^{\Omega}$, the information cost of this signal structure might nevertheless be influenced by the actions of others. To see this, observe that by (28), the prior $\mu \in \mathcal{U}$ depends the aggregate strategy $\bar{\alpha} \in \bar{\mathcal{A}}$. As a result, the divergence D in (23) might depend on $\bar{\alpha}$ even if the signal structure ν does not condition on $\bar{a}$, because both the prior and posteriors will be affected by $\bar{\alpha}$. This channel—aggregate actions affecting information costs—is exactly the one illustrated by our beauty contest example with Fisher information. We will eventually show that this can lead to an externality in our general game.

4 Equilibrium Definition and Existence

We proceed by defining our equilibrium concept and proving equilibrium existence.

To streamline our exposition, we first invoke the usual result in rational inattention problems that it is without loss of generality to equate signals with actions. To do so, let $\sigma : \Omega \to \Delta(A)$ denote a mixed strategy: a mapping from signal realizations $\omega \in \Omega$ to distributions

over actions, $\Delta(A)$. Given any mixed strategy $\sigma : \Omega \to \Delta(A)$ and signal structure $\nu^\Omega \in \mathcal{V}^\Omega$, we may define the induced conditional distribution over actions, $\nu^A : S \times R \times \bar{A} \to \Delta(A)$, as

$$\nu^A(a|s,r,\bar{a}) = \int_\Omega \sigma(a|\omega)\nu^\Omega(\omega|s,r,\bar{a})d\omega. \tag{25}$$

Let $\mathcal{V}^A$ be the set of all conditional distributions over actions. Rather than write the problem of the agent as a choice over both σ and ν^Ω, we condense these choices and write the problem of the agent as a choice over the conditional distribution of actions, $\nu^A \in \mathcal{V}^A$.

Each agent is infinitesimal, meaning that the agent treats the joint distribution of payoff-relevant states, payoff-irrelevant states, and aggregate actions as exogenous. We will look for symmetric Bayesian Nash equilibria in which all agents play best responses to the equilibrium action function $\bar{\alpha} \in \bar{\mathcal{A}}$, all agents choose the same conditional distribution of actions ν^A, and the equilibrium action function is generated by these conditional distributions.

We denote a symmetric strategy profile by

$$\xi \equiv \{\nu^A, \bar{\alpha}\},$$

consisting of identical strategies $\nu^A \in \mathcal{V}^A$ for all agents and an aggregate action function $\bar{\alpha} \in \bar{\mathcal{A}}$. We furthermore assume that, conditional on $(s,r,\bar{a})$, the realizations of signals across agents are independent. That is, it is only the distributions of actions, not the realizations, that are identical across agents. This independence of realizations allows us to apply the law of large numbers and require that each agent's average action a^i be consistent with the mean action $\bar{a}$ in the population [Uhlig, 1996]. We impose this as follows.

Definition 2. *A symmetric strategy profile ξ is **mean-consistent** if it satisfies*

$$\int_A a^i \nu^A(a^i|s,r,\bar{\alpha}(s,r))da^i = \bar{\alpha}(s,r) \quad \forall s \in S, r \in R, \ s.t. \ \mu_0(s,r) > 0. \tag{26}$$

By assuming that signal realizations are independent across agents, we are not ruling out correlation in equilibrium actions. Instead, we are simply imposing that the only channel by which agents may correlate is through their choice of how their actions condition on $(s,r,\bar{a})$. In particular, the exogenous state r allows agents to coordinate their actions on public signals or sunspots.

We define an equilibrium in our game as follows.

Definition 3. *Given a common prior $\mu_0 \in \mathcal{U}_0$, a **symmetric Bayesian Nash equilibrium** (BNE) of the game is a mean-consistent strategy profile ξ such that agents' strategies $\nu^A \in \mathcal{V}^A$ are best responses*

$$\nu^A \in \mathrm{argsup}_{\nu' \in \mathcal{V}^A} \sum_{s \in S, r \in R} \int_{\bar{A}} \left[\int_A u\left(a^i, \bar{a}, s\right) \nu'\left(a^i|s,r,\bar{a}\right) da^i \right] \mu\left(s,r,\bar{a}\right) d\bar{a} - C^A\left(\nu', \mu\right), \tag{27}$$

where

$$\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}. \tag{28}$$

23

Our equilibrium concept is based on Denti [2019]. Equilibrium Definition 3 is a hybrid of a Bayesian Nash equilibrium and a Rational Expectations equilibrium. It is a Bayesian Nash equilibrium in the sense that all agents play best responses given their beliefs, which we assume to be formed according to Bayes' rule. It is a Rational Expectations equilibrium [Grossman, 1976, Grossman and Stiglitz, 1980, 1976] in the sense that agents' may learn from endogenous aggregate variables while simultaneously choosing their own actions; thus beliefs are functions of endogenous actions while actions are functions of endogenous beliefs. Consistency of beliefs and actions is imposed by mean-consistency of strategy profiles (Definition 2) along with (28), which is the usual requirement that agents best-respond to the equilibrium strategies of other agents.

Our first result is that such an equilibrium exists.

Proposition 3. *A symmetric BNE of the game exists.*

Proof. See the appendix, 11.4. □

The proof of this result uses Kakutani's fixed point theorem in the usual way, relying on the finiteness of $S \times R$, the continuity of the utility function, and the convexity and continuity of the information cost function.

Having now established that our equilibria exist, we begin to study these equilibria. We first investigate under what circumstances equilibria do or do not exhibit "non-fundamental" volatility; we then study the circumstances under which equilibria are constrained efficient, defining constrained efficiency as coinciding with the solution to a particular planner's problem (which we will define subsequently).

For both of these results, the focus of our investigation will be the relationship between properties of the information cost function, in particular its associated divergence D, and the properties of the equilibrium. We therefore begin by defining the properties of information costs that will be the focus of our analysis.

5   Partial Monotonicity and Partial Invariance

In this section we define two concepts, "partial monotonicity" and "partial invariance," as properties of divergences. Mutual information exhibits both of these properties; Fisher information exhibits neither property. We will later show how these two properties of divergences are related to properties of the equilibria.

5.1   Coarsening and Embedding

We begin by introducing two types of operators, coarsenings and embeddings, which we define with respect to different partitions of the state space. Coarsenings and embeddings are ways

of moving back and forth between joint and marginal distributions. Specifically, coarsenings take joint distributions and transform them into marginal distributions, while embeddings transform a marginal into a joint by adding a conditional distribution.

Consider first a "coarsening" that removes information about $r \in R$. We let $\mathcal{U}_R \equiv \Delta(S \times \bar{A})$ denote the space of probability measures on $(S \times \bar{A})$ and $\mu_R \in \mathcal{U}_R$ denote a particular distribution on this space. The subscript R indicates the dimension (of the larger space) that is missing, a convention we follow below. We define a coarsening function $\gamma_R : \mathcal{U} \to \mathcal{U}_R$ by

$$\gamma_R\{\mu\}(s,\bar{a}) = \sum_{r \in R} \mu(s,r,\bar{a}), \qquad \forall s \in S, \bar{a} \in \bar{A}. \tag{29}$$

This operator takes a probability distribution $\mu \in \mathcal{U}$ on the larger space $(S \times R \times \bar{A})$ and projects it onto the smaller state space $(S \times \bar{A})$. As indicated by its subscript, this coarsening "throws out" all information about $r \in R$, conditional on $(s,\bar{a})$.

Consider now the opposite transformation, an "embedding," that adds information about $r \in R$. Letting $\hat{\phi}_R(r|s,\bar{a})$ denote a conditional distribution for $r \in R$, conditional on $(s,\bar{a})$, we define an embedding function $\phi_R : \mathcal{U}_R \to \mathcal{U}$ associated with the conditional distribution function $\hat{\phi}_R : S \times \bar{A} \to \Delta(R)$ by,

$$\phi_R\{\mu_R\}(s,r,\bar{a}) = \hat{\phi}_R(r|s,\bar{a}) \mu_R(s,r,\bar{a}), \quad \forall s \in S, r \in R, \bar{a} \in \bar{A}. \tag{30}$$

This operator takes a probability measure on the smaller space $(S \times \bar{A})$ and embeds it into the larger space $(S \times R \times \bar{A})$ using the information contained in $\hat{\phi}_R$. It thereby "adds" information about $r \in R$, conditional on $(s,\bar{a})$.[19]

Any embedding ϕ_R is associated with a particular conditional distribution $\hat{\phi}_R$, and in turn any conditional distribution function defines a particular embedding. Thus, while there is only one way to coarsen from $\mathcal{U}$ to $\mathcal{U}_R$, there are many possible ways to embed from $\mathcal{U}_R$ to $\mathcal{U}$. Let Φ_R denote the set of all possible embeddings from $\mathcal{U}_R$ to $\mathcal{U}$.

5.2   Monotonicity and Invariance in R

Armed with these definitions of coarsenings and embeddings with respect to R, we are now able to define our concepts of partial monotonicity and partial invariance with respect to R.

Consider a composition of the coarsening operator γ_R and a *specific* embedding $\phi_R \in \Phi_R$. Let $\eta_R : \mathcal{U} \times \mathcal{U} \to \mathcal{U}$ denote the operation that coarsens its first argument in R, then embeds using the conditional distribution defined by its second argument. We define this compositional operator as follows:

$$\eta_R\{\mu_1, \mu_2\}(s,r,\bar{a}) = \begin{cases} \frac{\mu_2(s,r,\bar{a})}{\gamma_R\{\mu_2\}(s,\bar{a})} \gamma_R\{\mu_1\}(s,\bar{a}) & \text{if } \gamma_R\{\mu_2\}(s,\bar{a}) > 0 \\ 0 & \text{if } \gamma_R\{\mu_2\}(s,\bar{a}) = 0. \end{cases} \tag{31}$$

---

[19]To ensure that the resulting distribution on $\Delta(S \times R \times \bar{A})$ remains in $\mathcal{U}$, we require that for all $s \in S$ and $\bar{a}, \bar{a}' \in \bar{A}$, the supports of $\hat{\phi}_R(\cdot|s,\bar{a})$ and $\hat{\phi}_R(\cdot|s,\bar{a}')$ do not intersect.

To apply this operation, we require that $\gamma_R\{\mu_1\}$ be absolutely continuous with respect to $\gamma_R\{\mu_2\}$, which we denote by $\gamma_R\{\mu_1\} \ll \gamma_R\{\mu_2\}$.

This compositional operator first takes the probability measure $\mu_1 \in \mathcal{U}$ on the larger space and projects it onto the smaller state space $(S \times \bar{A})$ by coarsening it in R. It then takes the resulting distribution and embeds it back into the larger space using *not* the conditional distribution defined by $\mu_1 \in \mathcal{U}$, but instead the conditional distribution defined by $\mu_2 \in \mathcal{U}$. That is, it essentially replaces the conditional distributions of μ_1 (of $r \in R$ conditional on $s, \bar{a}$) with those of μ_2. The end result is a distribution that is arguably "more like" μ_2 than μ_1 was originally. Note that by construction, $\eta_R\{\mu, \mu\} = \mu$ for any $\mu \in \mathcal{U}$.

With this compositional operator, we define "monotonicity in R" as follows.

Definition 4. *A divergence $D : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ is **monotone in R** if for all $\mu, \mu^\omega \in \mathcal{U}$ such that $\gamma_R\{\mu^\omega\} \ll \gamma_R\{\mu\}$,*

$$D(\mu^\omega \| \mu) \geq D(\eta_R\{\mu^\omega, \mu\} \| \mu). \tag{32}$$

Recall that a divergence is a non-negative function of the prior and the posterior with no requirements other than being equal to zero if and only if the prior and posterior are the same. The above property compares the divergence of the prior μ to the posterior μ^ω after replacing the posterior's conditional distributions of $r \in R$, conditional on $(s, \bar{a})$, with the divergence before this replacement. Monotonicity captures the idea that if we make the posterior more like the prior in this sense, then this reduces the divergence from the prior to the posterior.

We next define a different concept, "invariance in R."

Definition 5. *A divergence $D : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ is invariant with respect to Φ_R, or **invariant in R**, if for all $\mu, \mu^\omega, \mu' \in \mathcal{U}$ such that $\gamma_R\{\mu^\omega\} \ll \gamma_R\{\mu\}$ and $\gamma_R\{\mu'\} \ll \gamma_R\{\mu\}$,*

$$D(\eta_R\{\mu^\omega, \mu\} \| \mu) = D(\eta_R\{\mu^\omega, \mu'\} \| \eta_R\{\mu, \mu'\}). \tag{33}$$

The above property compares the divergence of the prior μ to the posterior μ^ω after replacing *both* the prior and the posterior's conditional distributions (of $r \in R$ conditional on $s, \bar{a}$), which may originally differ, with an identical conditional distribution from μ'. Invariance captures the idea that if the prior and posterior share a common conditional distribution, the exact values of this conditional distribution should not matter for the divergence from the prior to the posterior. This is different than monotonicity, which requires that replacing the posterior's conditional distributions of $r \in R$ with that of the prior reduces their divergence. Invariance and monotonicity together require the divergence to shrink to the same value for *all possible* conditional distributions (i.e. embeddings).

5.3 Invariance in $\bar{A}$

We have thus far defined coarsenings, embeddings, monotonicity, and invariance only with respect to partitions in $r \in R$. However, we can define these concepts with respect to other

dimensions as well, e.g. in $s \in S$ or $\bar{a} \in \bar{A}$. For the purposes of our exercise, we now consider coarsenings and embeddings with respect to partitions in $\bar{a} \in \bar{A}$ and define invariance in $\bar{A}$ along the same lines as our previous definition of invariance in R.

Consider $\mathcal{U}_0 \equiv \Delta\left(S \times R\right)$, the space of probability measures over the exogenous state space. We define a coarsening function, $\gamma_{\bar{A}} : \mathcal{U} \to \mathcal{U}_0$, by

$$\gamma_{\bar{A}}\{\mu\}\left(s, r\right) = \int_{\bar{A}} \mu\left(s, r, \bar{a}\right) d\bar{a}, \quad \forall s \in S, r \in R. \tag{34}$$

This operator takes a probability distribution $\mu \in \mathcal{U}$ on the larger space $\left(S \times R \times \bar{A}\right)$ and projects it onto the smaller state space $\left(S \times R\right)$. As indicated by its subscript, this coarsening "throws out" all information about $\bar{a} \in \bar{A}$, conditional on $\left(s, r\right)$.

Consider now an embedding that adds information about $\bar{a} \in \bar{A}$. We have in fact already defined such an embedding; the function $\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$ defined in (20) is an embedding that maps probability measures $\mu_0 \in \mathcal{U}_0$ on the smaller (exogenous) state space $\left(S \times R\right)$ to measures on the larger one. The associated conditional distribution function for this embedding is given by

$$\hat{\phi}_{\bar{A}}\left(\bar{a}|s, r\right) = \delta\left(\bar{a} - \bar{\alpha}\left(s, r\right)\right), \quad \forall s \in S, r \in R, \bar{a} \in \bar{A}, \tag{35}$$

that is, the degenerate distribution induced by the aggregate action function $\bar{\alpha}\left(s, r\right)$. Any aggregate action function $\bar{\alpha} \in \bar{A}$ defines a particular embedding from $\mathcal{U}_0$ to $\mathcal{U}$ with associated conditional distribution (35), and the set $\mathcal{U}$ is defined by the property that for all $\mu \in \mathcal{U}$, there exists a $\mu_0 \in \mathcal{U}_0$ and $\bar{\alpha} \in \bar{A}$ such that $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$. We can therefore think of $\bar{A}$ as defining the set of all possible embeddings from $\mathcal{U}_0$ to $\mathcal{U}$.

Armed with these definitions of coarsenings and embeddings with respect to $\bar{a} \in \bar{A}$, we may now define our concept of invariance in $\bar{A}$, in a manner that is exactly analogous to our defining of invariance in R (Definition 5). Consider again a composition of the coarsening operation and a *specific* embedding, $\eta_{\bar{A}} : \mathcal{U} \times \mathcal{U} \to \mathcal{U}$, that coarsens its first argument in $\bar{a} \in \bar{A}$, then embeds using the conditional distribution defined by its second argument. This operator is

$$\eta_{\bar{A}}\{\mu_1, \mu_2\}\left(s, r, \bar{a}\right) = \begin{cases} \frac{\mu_2\left(s, r, \bar{a}\right)}{\gamma_{\bar{A}}\{\mu_2\}\left(s, \bar{a}\right)}\gamma_{\bar{A}}\{\mu_1\}\left(s, \bar{a}\right) & \text{if } \gamma_{\bar{A}}\{\mu_2\}\left(s, \bar{a}\right) > 0 \\ 0 & \text{if } \gamma_{\bar{A}}\{\mu_2\}\left(s, \bar{a}\right) = 0, \end{cases} \tag{36}$$

which is the exact analog of (31). With this operator, we define invariance in $\bar{A}$ as follows.

Definition 6. *A divergence* $D : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ *is invariant with respect to* $\bar{A}$*, or* **invariant in** $\bar{A}$*, if for all* $\mu, \mu^\omega, \mu' \in \mathcal{U}$ *such that* $\gamma_{\bar{A}}\{\mu^\omega\} \ll \gamma_{\bar{A}}\{\mu\}$ *and* $\gamma_{\bar{A}}\{\mu'\} \ll \gamma_{\bar{A}}\{\mu\}$,

$$D(\eta_{\bar{A}}\{\mu^\omega, \mu\}||\mu) = D(\eta_{\bar{A}}\{\mu^\omega, \mu'\}||\eta_{\bar{A}}\{\mu, \mu'\}). \tag{37}$$

In our game, different aggregate strategies $\bar{\alpha}, \bar{\alpha}' \in \bar{A}$ will lead to different priors $\mu, \mu' \in \mathcal{U}$. However, if the divergence D is invariant in $\bar{A}$, we will demonstrate below that these different

priors do not lead to different information costs for signal structures that condition only on exogenous states ($\nu \in \mathcal{V}_0^\Omega$, as in Lemma 1). Consequently, with this form of invariance, the aggregate action strategy will not matter for agents' information costs, as in the linear-quadratic-Gaussian example with mutual information.

5.4 Relation to the standard definition of invariance.

This leads to our next point, which is the relationship between the forms of partial monotonicity and invariance we have defined and the stronger form of invariance discussed in other papers. The literature has focused on divergences that are simply "invariant," meaning that they are both monotone and invariant with respect to *all* possible coarsenings and embeddings between a larger space and a smaller space. These invariant divergences have been described in the information geometry literature (see Chentsov [1982], Amari and Nagaoka [2007]), and employed in economics by Hébert [2018] and Hébert and Woodford [2019]. Another term for coarsening is "compression," and invariant divergences have a close connection to the invariance-under-compression axiom described in Caplin, Dean, and Leahy [2019]. Mutual information, in particular, is invariant in this stronger sense, and hence is invariant and monotone in both R and $\bar{A}$. Fisher information, in contrast, is not invariant or monotone in either of these senses, provided that agents are allowed to learn about endogenous actions.

Our generalization to partial monotonicity and partial invariance allows us to study divergences that are invariant to some but *not necessarily all* partitions of the state space. In particular our concepts of partial monotonicity and partial invariance are defined with respect to partitions in certain dimensions, e.g. R and $\bar{A}$. Our framework thereby allows for divergences that may, for example, be invariant in R, but not invariant in $\bar{A}$, and vice-versa.

For example, consider the divergence defined by

$$D\left(\mu^\omega||\mu\right) = \theta_1 D_{KL}\left(\mu^\omega||\mu\right) + \theta_2 \sum_{r \in R}(\sum_{s \in S} \int_{\bar{A}} \mu(s,r,\bar{a})d\bar{a})D_{KL}\left(\mu^{\omega r}||\mu^r\right),$$

where D_{KL} is the Kullback-Leibler divergence and, for values of r occurring with positive probability,

$$\mu^r(s,\bar{a}) = \frac{\mu(s,r,\bar{a})}{\sum_{s \in S} \int_{\bar{A}} \mu(s,r,\bar{a})d\bar{a}}, \qquad \mu^{\omega r}(s,\bar{a}) = \frac{\mu^\omega(s,r,\bar{a})}{\sum_{s \in S} \int_{\bar{A}} \mu^\omega(s,r,\bar{a})d\bar{a}}.$$

Here, $\mu^r \in \mathcal{U}_R$ and $\mu^{\omega r} \in \mathcal{U}_R$ are the prior and posterior distributions on $\Delta(S \times \bar{A})$ condition on observing $r \in R$.

When $\theta_2 = 0$, this divergence is simply the KL divergence. When $\theta_2 > 0$, there is an extra penalty for having a distribution conditional on r under μ^ω that deviates from the distribution conditional on r under the prior μ. In the limit as $\theta_2 \to \infty$, the cost to learn about r remains unchanged, but it becomes impossible to learn anything aside from r. In this limit, every agent

will optimally choose to learn about r and only r even if r is not payoff-relevant (that is, if $\bar{a}$ does not depend on r), provided that r is in some way correlated with the payoff-relevant variables. Even away from this limit, it will generally be cheaper for the agent to choose signal structures that vary in r, even if learning about r is less useful than learning directly about s. One very straightforward interpretation of this cost function is that r is a public signal, and it is cheaper for agents to observe this public signal than to observe either the fundamentals or other agents' actions.

From this discussion, it is immediately apparent that this divergence is neither invariant nor monotone in R. However, because of the invariance of the Kullback-Leibler divergence, it is also straightforward to observe that this divergence is invariant and monotone in $\bar{A}$. We therefore conclude from this example that cost functions with one form of partial invariance but not the other exist and are potentially interesting.

Armed with these definitions, we next demonstrate that monotonicity in R and invariance in $\bar{A}$ are critical in determining the equilibrium properties of our game.

6   Non-Fundamental Volatility in Equilibrium

In this section we consider the question of whether the equilibrium depends on the payoff-irrelevant exogenous state r. We begin by defining a notion of measurability; we will say an equilibrium is s-measurable if the agents' signal structures ν^A and the aggregate strategy $\bar{\alpha}$ do not depend on r.

Definition 7. *An aggregate strategy $\bar{\alpha} \in \bar{A}$ is **s-measurable** if, for all $s \in S$ and $r, r' \in R$, $\bar{\alpha}(s,r) = \bar{\alpha}(s,r')$. A symmetric BNE (Definition 3) $\xi = \{\nu^A, \bar{\alpha}\}$ is **s-measurable** if $\bar{\alpha}$ is s-measurable and if, for all $s \in S$ and $r, r' \in R$, $\nu^A(a^i|s,r,\bar{\alpha}(s,r)) = \nu^A(a^i|s,r',\bar{\alpha}(s,r'))$.*

When the aggregate strategy $\bar{\alpha}$ depends on r, the associated equilibrium exhibits non-fundamental volatility in the sense that outcomes depend on non-payoff-relevant exogenous states. When it does not, the associated equilibrium exhibits zero non-fundamental volatility. By definition, if an equilibrium is s-measurable, it exhibits zero non-fundamental volatility. However, if an equilibrium is not s-measurable, it could either be that the aggregate strategy $\bar{\alpha}$ is not s-measurable, or that $\bar{\alpha}$ is s-measurable but the signal structure ν^A conditions on r. In the first of these situations, the economy exhibits non-fundamental volatility.

The second situation is non-generic. If ν^A conditions on r, then by the mean-consistency condition (Definition 2) we would generically expect that $\bar{\alpha}(s,r)$ depends on r. The special case in which this would not occur is if, under ν^A, r influences the higher moments of the agents' actions but not the mean action. This case is non-generic, in the sense that by slightly perturbing the agents' utility function, we can construct a new game with an equilibrium

involving an $\bar{\alpha}$ function that does depend on r. We discuss this issue in more detail in the Appendix, Section 10. In what follows, we will treat the question of whether or not the equilibrium is s-measurable as equivalent to the question of whether or not $\bar{\alpha}$ is s-measurable, which determines whether or not there is non-fundamental volatility.

6.1   Equilibrium Implications of Monotonicity in R

We begin by demonstrating that if the divergence D is monotone in R, an s-measurable equilibrium (featuring zero non-fundamental volatility) exists.

Monotonicity in R has the implication that if an agent does not care about r per se, only how it affects $\bar{a}$, then there is no reason for the agent to acquire any information about r. Recall that in our framework a signal structure $\nu \in \mathcal{V}^{\Omega}$ is a conditional distribution. That is, $\nu(\omega|s, r, \bar{a})$ is the probability of observing signal $\omega \in \Omega$ conditional on the realization of $(s, r, \bar{a})$.

Let us now define an operator $\nu_R : \mathcal{V}^{\Omega} \times \mathcal{U} \to \mathcal{V}^{\Omega}$ that removes from $\nu \in \mathcal{V}^{\Omega}$ the conditioning of the signal on r, while preserving the probabilities of each $\omega \in \Omega$ conditional on $(s, \bar{a})$,

$$
\nu_R\{\nu, \mu\}(\omega|s, r, \bar{a}) = \frac{\sum_{r' \in R}(\nu(\omega|s, r', \bar{a})\mu(s, r', \bar{a}))}{\gamma_R\{\mu\}(s, \bar{a})} \ \forall s \in S, \bar{a} \in \bar{A} \ s.t. \ \gamma_R\{\mu\}(s, \bar{a}) > 0.
$$

From this definition, we can observe immediately that $\nu_R\{\nu, \mu\}(\omega|s, r, \bar{a}) = \nu_R\{\nu, \mu\}(\omega|s, r', \bar{a})$ for all $r, r' \in R$, which is to say that the signal structure does not condition on r. It also follows immediately that $\pi\{\nu_R\{\nu, \mu\}, \mu\} = \pi\{\nu, \mu\}$, which is to say that the unconditional probabilities of each signal are preserved by this operator.

Our next lemma answers the following question: when is it always less costly (in terms of the information cost) to avoid conditioning signals on r?

Lemma 2. *(i) If the divergence associated with the cost function C^{Ω} is monotone in R, then for all priors $\mu \in \mathcal{U}$ and all signal structures $\nu \in \mathcal{V}^{\Omega}$,*

$$
C^{\Omega}(\nu_R\{\nu, \mu\}, \mu) \leq C^{\Omega}(\nu, \mu).
$$

(ii) If, for all priors $\mu \in \mathcal{U}$ and all signal structures $\nu \in \mathcal{V}^{\Omega}$,

$$
C^{\Omega}(\nu_R\{\nu, \mu\}, \mu) \leq C^{\Omega}(\nu, \mu),
$$

then the divergence is monotone in R.

Proof. See the appendix, 11.5. $\square$

Part (i) of Lemma 2 states that if the divergence is monotone in R, then the minimally-informative signal structure is also the least-costly. One implication of this result is conditional

30

independence: because the agents' payoff never depends on r conditional on $(s, \bar{a})$, the agent's optimal signal ω will be independent of r conditional on $(s, \bar{a})$. Part (ii) of Lemma 2 states the converse: if removing conditioning on r always reduces information costs, then the divergence is monotone in R. Put together, Lemma 2 tells us that monotonicity in R is equivalent to the statement that paying attention to r is always costly. We now use this result to show that monotonicity in R is sufficient for the existence of s-measurable equilibria.

Proposition 4. *If the divergence associated with the cost function C^Ω is monotone in R, then an s-measurable symmetric BNE of the game exists.*

Proof. See the appendix, 11.6. □

Our proof is essentially a restatement of our existence proof combined with an application of part (i) of Lemma 2. The key observation is that with monotonicity in R, agents optimally choose actions that conditional on $(s, \bar{a})$, are independent of r. This is because conditional on $(s, \bar{a})$, an agent has no reason to acquire information about r: this would only increase the agent's information costs with no benefit. As a result, if agents face an s-measurable aggregate action function $\bar{\alpha}$, they best-respond with a policy whose mean action is indeed s-measurable.

The sufficient conditions in Proposition 4 are in fact stronger than necessary: divergences need not be R-monotone on *all* priors. Instead, it suffices for divergences to be R-monotone only on priors that may occur in an s-measurable equilibrium. Intuitively, only these priors matter when conjecturing the existence of an s-measurable equilibrium.

To weaken the conditions Proposition 4, we define $\mathcal{U}^{s-meas}(\mu_0) \subseteq \mathcal{U}$ to be the set of priors that may be **generated by s-measurable $\bar{\alpha}$ given the exogenous prior** $\mu_0 \in \mathcal{U}_0$,

$$\mathcal{U}^{s-meas}(\mu_0) = \left\{ \mu \in \mathcal{U} : \exists \bar{\alpha} \in \bar{\mathcal{A}} \text{ s.t. } \mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\} \text{ and } \bar{\alpha}(s, r) = \bar{\alpha}(s, r') \; \forall s \in S, r, r' \in R \right\},$$

and define $\mathcal{U}^{s-meas} = \cup_{\mu_0 \in \mathcal{U}_0} \mathcal{U}^{s-meas}(\mu_0)$ as the set of all s-measurable priors. We are now able to state a version of Proposition 4 demonstrating that monotonicity in R on s-measurable priors is both necessary and sufficient to ensure the existence of an s-measurable symmetric BNE, regardless of the utility function.

Proposition 5. *An s-measurable symmetric BNE of the game exists for all utility functions u satisfying Assumption 1 if and only if the divergence associated with the cost function C^Ω is monotone in R on all priors $\mu \in \mathcal{U}^{s-meas}(\mu_0)$.*

Proof. See the appendix, 11.7. □

The intuition for this result is essentially the converse of our previous result. With non-monotonicity in R, there are priors $\mu \in \mathcal{U}$ such that agents optimally choose actions that, conditional on $(s, \bar{a})$, are *not* independent of r. This is because even if the agent has no particular

31

concern for the value of r, she finds it cheaper to obtain signals correlated with r than to gather no information about r at all. As a result, each agent best-responds with a policy that, conditional on $(s, \bar{a})$, varies in r, provided that the agent gathers any information at all.

We conclude that R-monotonicity on s-measurable priors is the key condition that guarantees zero non-fundamental volatility. When this condition is violated, there are utility functions that will generate non-fundamental volatility in equilibrium (because an equilibrium exists by Proposition 3, and it is not s-measurable by Proposition 5).

To further extend our results, we consider cost functions that are generically non-monotone in R. We define the "opposite" of monotonicity in R, relying on the "only-if" aspect of Lemma (2). Consider any informative signal structure $\nu \in \mathcal{V}_0^\Omega$ that does not condition on r. We will say that a cost function C^Ω is "generically non-monotone in R" if, generically on the set of priors $\mu \in \mathcal{U}$, the signal structure ν is *not* the least-costly of all the signal structures that coarsen to ν, except at isolated points. Our use of the term generic follows Geanakoplos and Polemarchakis [1986] and Farhi and Werning [2016].

Definition 8. *A cost function C^Ω is **generically non-monotone in R** if, generically on the set $\mu \in \mathcal{U}^{s-meas}$, for all informative signal structures $\nu \in \mathcal{V}_0^\Omega$ such that $\nu = \nu_R\{\nu, \mu\}$, there exists a $\nu' \in \mathcal{V}_0^\Omega$ with $\nu = \nu_R\{\nu', \mu\}$ such that*

$$C^\Omega\left(\nu', \mu\right) < C^\Omega\left(\nu, \mu\right).$$

We have defined generic non-monotonicity as a property of the cost function C^Ω as opposed to of the divergence purely for convenience. Note that monotonicity and generic non-monotonicity are not exhaustive classes of cost functions; cost functions might exhibit monotonicity in R for some priors but not others. We have little to say about whether s-measurable equilibria will or will not exist in this case.

The reason we need a notion of generic non-monotonicity, as opposed to non-monotonicity for all priors, is illustrated in our linear-quadratic-Gaussian example with the Fisher information cost function. In that example, if under the prior s and r are independent, then even with the Fisher information cost function the equilibrium will have zero non-fundamental volatility. This situation is non-generic in the sense that even small amounts of correlation between s and r will restore the result that the equilibrium features non-fundamental volatility.

Before presenting our result, we need one more definition. We will say that an equilibrium is deterministic if the aggregate action is constant and agents do not acquire any information.

Definition 9. *A symmetric BNE (Definition 3) $\xi = \{\nu^A, \bar{\alpha}\}$ is **deterministic** if, for all $s, s' \in S$ and $r, r' \in R$, $\bar{\alpha}(s, r) = \bar{\alpha}(s', r')$ and $\nu^A(a^i | s, r, \bar{\alpha}(s, r)) = \nu^A(a^i | s', r', \bar{\alpha}(s', r'))$.*

As we discussed in our linear-quadratic-Gaussian example, if the costs of acquiring any information exceed the benefits (which can happen in that example with both mutual information and Fisher information), there can be equilibria in which no information is

acquired and the aggregate action is identical in all states. This is also a possibility in our general game, but as our next result shows, if the cost function is non-monotone in R and the equilibrium features some information gathering, then it will generically have non-fundamental volatility.

Proposition 6. *If the cost function is generically non-monotone in R on the set of priors $\mathcal{U}^{s-meas}$, then generically all symmetric BNE of the game are either not s-measurable or are deterministic.*

Proof. See the appendix, 11.9. □

6.2 Interpretation and Remarks

When do equilibria exhibit non-fundamental volatility? In exogenous information environments, non-fundamental volatility originates as errors in public signals. Noisy public signals, or more generally correlated errors in beliefs, are natural components of generic information structures [Bergemann and Morris, 2013]. In these environments, agents costlessly observe public signals; as long as public signals contain information about fundamentals, agents condition their actions on it. As a result, errors in these signals orthogonal to fundamentals affect equilibrium outcomes.

However, under endogenous information acquisition, what appears to be a rather natural property to impose on cost structures—monotonicity in R—leads to a surprising and strong result: zero non-fundamental volatility in equilibrium. That is, if agents have no reason per se to obtain information about $r \in R$, and paying attention to r only increases costs, then in equilibrium agents will optimally choose to ignore r. As a result, actions are conditionally independent of r, and equilibria feature zero non-fundamental volatility. In fact, as we show in Section 2, mutual information—the typical cost function used in the rational inattention literature—produces this result.

To break this—to eliminate s-measurable equilibria altogether—we show that one must break monotonicity in R. If cost functions are non-monotonic in R, then agents pay attention to r, even conditional on $s, \bar{a}$. And, because the realization of r is common across all agents, it introduces correlated errors in agents' actions, resulting in non-fundamental volatility in equilibrium outcomes.

The variables $r \in R$ thereby play the role of "noisy public signals:" they capture the idea that it is comparatively cheap for agents to observe r as opposed to receiving signals *only* about fundamentals s. In fact, note that in the limit in which r is completely costless, they are identical to costless public signals (e.g. as in the exogenous information case). Away from this limit, the variables $r \in R$ are not costless, but are "salient."

Our beauty contest example in Section 2 demonstrates that with Fisher information, the equilibrium generically features non-fundamental volatility. It should now be clear that the

underlying reason for why Fisher information and mutual information lead to such different equilibrium properties is due to this particular difference in cost structure: mutual information is monotonic in R, while Fisher information is not.

7 Efficiency under Exogenous Information

We next turn to the question of efficiency, and the connection between the cost functions and informational externalities. To study inefficiencies with endogenous information acquisition, we must first isolate inefficiencies that may arise in any other aspect of the game. In this section, we consider a version of the game in which information is incomplete but exogenous. If the equilibrium is constrained efficient in the game under exogenous information, then agents *use* their information efficiently.

We first establish a sufficient condition that ensures the use of information is efficient. Angeletos and Pavan [2007] provide a related condition in a linear-quadratic setting with one-dimensional action spaces (like our beauty contest example). Our analysis in this section generalizes their result to our game, which features multi-dimensional actions and general payoffs. After we establish that our condition is sufficient, we show that it is also necessary to guarantee constrained efficiency for all possible priors and signal structures.

In the subsequent section, we assume efficiency in the *use* of information, and ask what more is needed to guarantee efficiency in the *acquisition* of information.

7.1 The Game under Exogenous Information

In the spirit of Lemma 1, we define our game with exogenous information by endowing agents with a given signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$. These signal structures are exogenous both in the sense that agents do not choose them and in the sense that they condition only on exogenous state variables.[20] We formalize the agent's choice set under exogenous information in a slightly unusual way, to emphasize the connection between games with exogenous information and games with endogenous information.

All other features of the game—payoffs, action spaces, exogenous priors–remain the same. Throughout this section we continue to impose Assumption 1 on action spaces and payoffs. Because the agent no longer chooses his own signal structure and there are no information costs, our assumptions on information acquisition costs are not applicable in this section.

---

[20]It is well-known that in games with exogenous information structures, allowing for signals about endogenous objects introduces an informational externality; see e.g. Laffont [1985], Angeletos and Pavan [2009], Amador and Weill [2010], Vives [2017], Angeletos, Iovino, and La'O [2020]. In order to maintain a clean benchmark, we abstract from such externalities in this section.

Choice sets with exogenous signal structures. Recall that any mixed strategy $\sigma : \Omega \to \Delta(A)$ and exogenously given signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$ together define a conditional distribution over actions, $\nu^A \in \mathcal{V}_0^A$, by (25). In the game with endogenous information, we wrote the agent's problem as a choice over his conditional distribution of actions, $\nu^A \in \mathcal{V}^A$; see Definition 3. In the exogenous information game, we instead endow the agent with a signal structure ν^Ω; as a result, the agent's only choice is his mixed strategy σ.

But now observe that the mapping σ can be thought of as a "garbling" in the sense of Blackwell [1953]. That is, the conditional distributions ν^A that arise from the signals ν^Ω and mixed strategies σ are weakly Blackwell-dominated by the conditional distributions ν^Ω. In fact, by Blackwell's theorem, the set of conditional action distributions $\nu^A \in \mathcal{V}_0^A$ that can be feasibly created by any mixed strategy σ are precisely those that are Blackwell-dominated by ν^Ω.

For any signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$, let $\mathcal{B}^A(\nu^\Omega) \subset \mathcal{V}_0^A$ denote the convex subset of conditional action distributions in $\mathcal{V}_0^A$ that are Blackwell-dominated by ν^Ω. In the game with exogenous information, we can write the agent's problem as a choice over conditional action distributions $\nu^A \in \mathcal{B}^A(\nu^\Omega)$, given the exogenously endowed signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$.

7.2 Equilibrium and Efficiency Definitions

We are now in a position to define equilibrium and efficiency with exogenous information. As in our definition of equilibrium in the original game with endogenous information (Definition 3), we focus on symmetric Nash equilibria in which all agents choose the same action strategy.

Definition 10. (Exogenous information game.) *Given a common prior $\mu_0 \in \mathcal{U}_0$ and exogenous signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$, a **symmetric Bayesian Nash equilibrium** of the game under exogenous information is a mean-consistent strategy profile ξ such that agents' strategies $\nu^A \in \mathcal{V}_0^A$ are best responses*

$$\nu^A \in \sup_{\nu' \in \mathcal{B}^A(\nu^\Omega)} \sum_{s \in S, r \in R} \int_{\bar{A}} \left[\int_A u\left(a^i, \bar{a}, s\right) \nu'\left(a^i | s, r, \bar{a}\right) da^i \right] \mu\left(s, r, \bar{a}\right) d\bar{a}.$$

with $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$.

Our equilibrium definition under exogenous information mirrors our equilibrium definition under endogenous information. In fact, Definitions 10 and 3 are nearly identical—the only difference is that with endogenous information, agents face a convex cost of information (C^A), whereas with exogenous information, their choice of signal structure is restricted to a convex set ($\mathcal{B}^A(\nu^\Omega)$). Viewed from this perspective, the game under exogenous information is not very different from the game under endogenous information, and our proof of existence applies almost unchanged.

Proposition 7. *A symmetric BNE of the game under exogenous information exists.*

Proof. See the appendix, 11.10. □

Having established equilibrium existence, we next consider constrained efficiency. We define constrained efficiency in this game as the solution to a particular planner's problem. Our definition enforces symmetry in strategies as a constraint on the planner, mirroring our symmetric BNE definition.

Definition 11. (Exogenous information game.) *Given a common prior $\mu_0 \in \mathcal{U}_0$ and exogenous signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$, a symmetric strategy profile $\xi^* = (\nu^{A*}, \bar{\alpha}^*)$ is **constrained efficient** if it solves*

$$\sup_{\nu^A \in \mathcal{B}^A(\nu^\Omega), \bar{\alpha} \in \bar{\mathcal{A}}} \sum_{s \in S, r \in R} \int_{\bar{A}} \left[\int_A u\left(a^i, \bar{a}, s\right) \nu^A \left(a^i | s, r, \bar{a}\right) da^i \right] \mu\left(s, r, \bar{a}\right) d\bar{a}$$

subject to $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$ and mean-consistency (Definition 2).

Our notion of constrained efficiency follows that in Angeletos and Pavan [2007].[21] In our environment, the planner chooses a mean-consistent symmetric strategy profile ξ in order to maximize welfare. This planner can tell agents how to *use* their information, but is constrained in that she may not endow agents with more information nor transfer information from one agent to another. The planner therefore treats the convex set $\mathcal{B}^A(\nu^\Omega)$ as a *feasibility* constraint on the set of possible strategies.

7.3 A Sufficient Condition for Constrained Efficiency

As already stated, the purpose of this section is to obtain sufficient conditions for efficiency in the use of information in our setting, as Angeletos and Pavan [2007] have done for linear-quadratic games. This entails ensuring that no externalities exist in the game under exogenous information.

A symmetric BNE of the game under exogenous information will not necessarily be constrained efficient in the sense of Definition 11. Even under complete information, standard "payoff externalities" may arise in any environment in which agents' actions affect the payoffs of others. Classic examples of payoff externalities include pollution, monopolistic competition, public goods provision, network spillovers (vaccines), etc.[22]

To understand the source of potential externalities in our game, consider first the special case in which all agents have complete information, and suppose for expositional purposes that the utility function is differentiable and all actions are interior. In this case, the first-order conditions of an individual agent i and the planner with respect to that agent's action

---

[21]See also Radner [1962] and Vives [1988].

[22]Of course, payoff externalities also do not necessarily lead to inefficiency—as the classic welfare theorems demonstrate in the case of pecuniary externalities.

are, respectively,

$$\frac{\partial u(\cdot)}{\partial a^i} = 0, \qquad \text{and} \qquad \frac{\partial u(\cdot)}{\partial a^i} + \int \frac{\partial u(\cdot)}{\partial \bar{a}} dj = 0, \qquad \forall i \in [0,1], s \in S.$$

These first-order conditions are similar except that the planner considers the impact of agent i's action on other agents, $\int \frac{\partial u(\cdot)}{\partial \bar{a}} dj$, whereas the agent does not. This term represents a standard payoff externality: the effect of agent i's action on the aggregate action, which in turn affects all other agents' payoffs. This effect is not internalized by agent i and is therefore absent from her first-order condition. Constrained efficiency requires that this term be equal to zero given the equilibrium actions of the other agents.

Consider the same heuristic argument when agents have incomplete and dispersed information. In the exogenous information game, the first-order conditions of an individual agent i and the planner with respect to that agent's action are, respectively:

$$\mathbb{E}\left[\frac{\partial u(\cdot)}{\partial a^i} \middle| \omega^i\right] = 0, \qquad \text{and} \qquad \mathbb{E}\left[\frac{\partial u(\cdot)}{\partial a^i} + \int \frac{\partial u(\cdot)}{\partial \bar{a}} dj \middle| \omega^i\right] = 0, \qquad \forall i \in [0,1], \omega^i \in \Omega \qquad (38)$$

These conditions are similar to those in the complete information game; again the second term in the planner's condition is absent from the agent's optimality condition. Now, however, efficiency requires that this condition hold for all possible realizations of the agent's signal, $\omega^i \in \Omega$. Thus, in order for the equilibrium to be efficient under exogenous information, there must be no payoff externalities from the perspective of any agent in the economy conditional on *any* realization of his or her signal.

The question, then, is under what circumstances this equation will hold in equilibrium. Angeletos and Pavan [2007] show that this condition requires more than what is required for efficiency under complete information. That is, a game can be efficient under complete information, but inefficient under incomplete and dispersed information. This is because under complete information, the equilibrium actions of all agents are identical and equal to the aggregate action. In this case efficiency only requires that the planner's and the agents' optimality conditions coincide *at the equilibrium aggregate action*. With incomplete and dispersed information, efficiency requires that (38) hold for all signal realizations in equilibrium given the optimal strategies of agents ν^A and the associated aggregate action function $\bar{\alpha}$.

With a one-dimensional action space and linear-quadratic payoffs, e.g. as in our beauty contest example, Angeletos and Pavan [2007] demonstrate that

$$\frac{\partial^2}{\partial a^i \partial \bar{a}} u(a^i, \bar{a}, s) = -\frac{\partial^2}{\partial \bar{a}^2} u(a^i, \bar{a}, s)$$

is sufficient to guarantee efficiency under exogenous information.[23] Likewise, our main result in this section provides a sufficient condition for efficiency with exogenous information to hold

---

[23]In the context of our beauty contest game in Section 2, this condition holds for the quadratic payoff function in (1).

given a utility function u defined on an action space A and set of fundamental states S, for all finite sets R, priors $\mu_0 \in \mathcal{U}_0$, and signal structures $\nu^\Omega \in \mathcal{V}_0^\Omega$.

Proposition 8. *Suppose there exists a function $G : \bar{A} \times S \to \mathbb{R}$ and a function $g : A \times S \to \mathbb{R}$ such that*

$$u(a^i, \bar{a}, s) = g(a^i; s) + G(\bar{a}; s) + (a^i - \bar{a}) \cdot \nabla G(\bar{a}; s), \tag{39}$$

where $\nabla G(\bar{a}; s)$ denotes the gradient of G with respect to its first argument, and that u satisfies Assumption 1.[24] Then there exists a symmetric BNE with exogenous signals (Definition 10) that is constrained efficient (Definition 11).

Proof. See the appendix, 11.11. □

The proof of sufficiency is relatively straightforward. The first-derivative of the payoff function in (39) with respect to $\bar{a}$ is given by

$$\frac{\partial u(\cdot)}{\partial \bar{a}} = \mathbf{H}(G(\bar{a}; s))(a^i - \bar{a}).$$

where $\mathbf{H}(G(\cdot))$ is the Hessian matrix of second-order partial derivatives of G. Integrating this expression across agents results in a gradient of zero, as required by (38). It is also immediately apparent from this expression that, if u were quadratic and the action space were one-dimensional (so that $\mathbf{H}(G(\bar{a}; s))$ is a constant), then our condition would reduce to the Angeletos and Pavan [2007] condition.

Note that condition (39) is a condition on payoffs u alone; it is sufficient for constrained efficiency with exogenous information regardless of the details of the information structure. Note also that the convexity or concavity of G determines whether this functional form exhibits strategic complementarity of substitutability. For example, in the one-dimensional context, if G is convex, then ∇G is increasing in $\bar{a}$, and larger actions a become preferable as $\bar{a}$ increases.

Armed with this result, we now know how to "shut down" these externalities, so as to isolate externalities arising from information acquisition in our game with endogenous information—the focus of our following section. Prior to that, however, we will show that our condition is not only sufficient but in a certain sense necessary to guarantee constrained efficiency.

7.4 A Necessary Condition for Constrained Efficiency

In this sub-section, we demonstrate that the functional form described in Proposition 8 is necessary for constrained efficiency to hold regardless of the details of the information structure, within a large class of utility functions.

---

[24]That is, g is continuously differentiable on A for all $s \in S$, and G is continuously twice-differentiable on $\bar{A}$ for all $s \in S$.

To simplify our analysis, we focus on payoffs that are concave in $\bar{a}$; this is equivalent under the functional form of (39) (and in Angeletos and Pavan [2007]) to assuming weak strategic complementarity. The class of utility functions we consider in this sub-section is the set of strictly concave and sufficiently smooth utility functions that guarantee optimal actions on the interior of A. The two assumptions below formalize the properties of this class of utility functions.

Assumption 5. *For all $s \in S$, the utility function $u(a^i, \bar{a}, s)$ is strictly concave and continuously twice-differentiable on $(a^i, \bar{a})$.*

Assumption 6. *There exists a set A^* in the relative interior of A such that for all $a' \in A \setminus A^*$ there exists an $a \in A^*$ satisfying $u(a, \bar{a}, s) > u(a', \bar{a}, s) \, \forall s \in S, \bar{a} \in \bar{A}$.*

The utility function in our beauty contest fall into this class given any finite set S, for a sufficiently large action space A. That is, at some point, actions are so far from optimal given any exogenous state that they are dominated regardless of what the other agents do. As a result, in any equilibrium, actions are guaranteed to be interior. To understand why we want to ensure that optimal actions are interior, observe that if they are not, allocations can be efficient despite the presence of externalities. For example, an excessive incentive for private agents to consume gasoline does not lead to inefficiency if the planner would choose maximal gasoline consumption anyways.

Our necessity result shows that, if we are given a set S, action space A, and utility function u satisfying Assumptions 1, 5, and 6, and are furthermore told that for all finite sets R, priors $\mu_0 \in \mathcal{U}_0$, and signals $\nu^\Omega \in \mathcal{V}_0^\Omega$, there is a constrained efficient symmetric BNE, then the utility function must satisfy (39) on the relevant part of its domain. To understand this last qualifier, let us suppose we are given such a utility function u, and let us define given this utility function the set $A^{ce} \subset A$ of actions that occur in some constrained efficient symmetric BNE. That is, if $a \in A^{ce}$, there is some (R, μ_0, ν^Ω) such that in a constrained efficient symmetric BNE this action occurs with positive probability.

Let us now consider a different utility function $\tilde{u}$, with the property that $\tilde{u}(a^i, \bar{a}, s) = u(a^i, \bar{a}, s)$ for all $a^i \in A^{ce}, \bar{a} \in \bar{A}$, and $s \in S$, and $\tilde{u}(a^i, \bar{a}, s) < u(a^i, \bar{a}, s)$ otherwise. It is immediately apparent that the set of constrained efficient symmetric BNE given $\tilde{u}$ is identical to the set of constrained efficient symmetric BNE given u, because all we have done is make socially sub-optimal actions even worse. Consequently the set of actions A^{ce} is identical for the two utility functions, and if u guarantees constrained efficiency, so does $\tilde{u}$.

What we learn from this example is that it cannot be necessary for u to satisfy (39) on all of its domain to guarantee constrained efficiency. Instead, we will show that A^{ce} is a convex subset of A, and that it is necessary for u to satisfy (39) on $A^{ce} \times A^{ce} \times S$ to guarantee constrained efficiency. We begin by formally defining A^{ce}.

Definition 12. *Fix a set S, action space A, and utility function u. An action $a \in A$ is an element of the set A^{ce} if there exists a finite set R, prior $\mu_0 \in \mathcal{U}_0$, and signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$ such that a constrained efficient symmetric BNE $(\nu^A, \bar{\alpha})$ exists with $\nu^A(a|s, r, \bar{\alpha}(s, r)) > 0$ for some $s \in S, r \in R$ with $\mu_0(s, r) > 0$.*

That is, $a \in A^{ce}$ if there is some situation that rationalizes this action in some circumstances as part of a constrained efficient equilibrium. Armed with this definition, we provide our necessary condition to guarantee constrained efficiency with exogenous signals.

Proposition 9. *Take as given the set of fundamental states S, action space A, and a utility function u satisfying Assumptions 1, 5, and 6. If for all finite sets R, priors $\mu_0 \in \mathcal{U}_0$ and signal structures $\nu^\Omega \in \mathcal{V}_0^\Omega$ there exists a constrained efficient symmetric BNE, then $u(a^i, \bar{a}, s)$ satisfies (39) for some functions g, G on all $(a^i, \bar{a}, s) \in A^{ce} \times A^{ce} \times S$, with G twice-differentiable and convex.*

Proof. See the appendix, 11.12. □

Condition (39) is therefore both sufficient and necessary to guarantee efficiency in the use of information. Let us now assume it, and ask what more is required to guarantee efficiency in the acquisition of information.

8 Efficiency under Endogenous Information

We now return to the original game with endogenous information acquisition as described in Section 3. In addition to Assumption 1 on action spaces and payoffs, we reinstate Assumptions 2-4 on the costs of information acquisition. We define constrained efficiency as the solution to a certain planner's problem in this setting as follows.

Definition 13. *Given a common prior $\mu_0 \in \mathcal{U}_0$, a symmetric strategy profile $\xi^* = (\nu^{A*}, \bar{\alpha}^*)$ is* ***constrained efficient*** *if it solves*

$$\sup_{\nu^A \in \mathcal{V}^A, \bar{\alpha} \in \bar{\mathcal{A}}} \sum_{s \in S, r \in R} \int_{\bar{A}} \left[\int_A u\left(a^i, \bar{a}, s\right) \nu^A\left(a^i|s, r, \bar{a}\right) da^i \right] \mu\left(s, r, \bar{a}\right) d\bar{a} - C^A\left(\nu^A, \mu\right)$$

subject to $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$ and mean-consistency (Definition 2).

Our definition of constrained efficiency under endogenous information mirrors our definition of constrained efficiency under exogenous information, but with a convex cost on ν^A in the place of a restriction to a convex set. This cost enters the planner's problem; that is, the planner acknowledges that the agents face information acquisition costs.

The distinction between exogenous and endogenous information games in our context makes clear why, at least potentially, a new externality may arise in the game with endogenous

40

information acquisition. In the exogenous information game, agents strategies' could affect other agents *only* through payoffs. This was by construction: each agent's convex set $\mathcal{B}^A(\nu^\Omega)$ of strategies was independent of other agents' actions.[25] In contrast, in the endogenous information game, the cost function C^A can in general depend on the aggregate action function $\bar{\alpha}$. This is true even though it is without loss of generality for agents to choose signals about exogenous states only ($\nu^A \in \mathcal{V}_0^A$), as discussed in Section 3. This dependence opens the door to externalities in the acquisition of information: agents may not internalize how their own strategy affects the information acquisition costs of others. In what follows, we derive under what conditions such an externality can arise.

8.1 Efficiency in the Acquisition of Information

In order for there to be inefficiency in the acquisition of information, there must be an externality by which the actions of one agent affect the welfare of others. In our framework, such an externality can exist only via the the aggregate action function $\bar{\alpha}$.[26] Recall our definition of posterior-separable cost functions C^A in (23), and note that in what follows we will use the action space A as the signal alphabet, to align our results with our definitions of equilibrium and constrained efficiency. In principle, C^A can vary with the aggregate action function $\bar{\alpha}$ via both the unconditional signal probabilities $\pi\{\nu^A, \mu\}$ and the divergence D.

However, Lemma 1 shows that it is without loss of generality to assume that $\nu^A \in \mathcal{V}_0^A$, meaning that ν^A does not in fact condition on $\bar{a}$. As a result, the distribution of $\bar{a}$ does not affect the unconditional signal probabilities holding fixed such a signal structure, meaning that for all $\mu_0 \in \mathcal{U}_0$, $\nu^A \in \mathcal{V}_0^A$, and $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,

$$\pi\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\} = \pi\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\}.$$

Moreover, $\nu^A \in \mathcal{V}_0^A$ has implications for the structure of the posteriors. Let $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$, and observe that the posterior after receiving the signal recommending action $a \in A$, $\mu^a \in \mathcal{U}$, is, for an arbitrary $\bar{a}_0 \in \bar{A}$,

$$\mu^a\{\nu^A, \mu\}(s, r, \bar{a}) = \frac{\nu^A(a|s, r, \bar{a}_0)\, \mu_0(s, r)\, \delta(\bar{a} - \bar{\alpha}(s, r))}{\pi\{\nu^A, \mu\}(a)},$$

and therefore satisfies, for all all $\mu \in \mathcal{U}$ and $\nu^A \in \mathcal{V}_0^A$,

$$\mu^a\{\nu^A, \mu\} = \eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu\}$$

---

[25]This was due to the assumption that $\nu^\Omega \in \mathcal{V}_0^\Omega$. If this were not the case, then this set would in general vary with the aggregate action function $\bar{\alpha}$, as it does in Laffont [1985], Angeletos and Pavan [2009], Amador and Weill [2010], Vives [2017], Angeletos, Iovino, and La'O [2020].

[26]As discussed in our beauty contest example, our framework precludes the possibility that the cross-sectional dispersion of actions affects agents' payoffs. This channel is the central focus of Colombo, Femminis, and Pavan [2014]; we abstract from it here.

where $\eta_{\bar{A}}$ is the composition of a coarsening and embedding we used to define invariance with respect to $\bar{A}$ (see (36) and Definition 6). That is, the conditional distribution of $\bar{a}$ given (s, r) under the posterior μ^a is the same as the conditional distribution under the prior μ.

From this observation, it is a small step to show that, if D is invariant with respect to $\bar{A}$ and $\nu^A \in \mathcal{V}_0^A$, changing $\bar{\alpha}$ does not change the divergence between the posterior and prior. Consequently, in this case, $\bar{\alpha}$ does not change the information cost. Conversely, if the information cost is unaffected by $\bar{\alpha}$ regardless of the values of $\nu^A \in \mathcal{V}_0^A$ and $\mu_0 \in \mathcal{U}_0$, it must be the case that the divergence D is invariant in $\bar{A}$. The following lemma summarizes these results.

Lemma 3. *(i) If the divergence associated with the cost function C^A is invariant in $\bar{A}$, then for all priors $\mu_0 \in \mathcal{U}_0$ and all signal structures $\nu^A \in \mathcal{V}_0^A$, and all $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,*

$$C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha} \} \right) = C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha}' \} \right).$$

(ii) If, for all priors $\mu_0 \in \mathcal{U}_0$ and all signal structures $\nu^A \in \mathcal{V}_0^A$, and all $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,

$$C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha} \} \right) = C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha}' \} \right),$$

then the divergence is invariant in $\bar{A}$.

Proof. See the appendix, 11.14. □

Thus, if the divergence associated with the cost function is invariant with respect to $\bar{\mathcal{A}}$, then there is no channel by which agents' actions affect another agents' cost of information. Consequently, if the game is constrained efficient under exogenous information (which we guarantee by assuming (39) holds), it will also be constrained efficient under endogenous information acquisition.

Proposition 10. *In the game with endogenous information, a constrained-efficient symmetric BNE exists if u satisfies the conditions of Proposition 8 and the divergence D is invariant with respect to $\bar{A}$. There exists a convex function $\bar{G} : \bar{\mathcal{A}} \to \mathbb{R}$, which is determined by g and C^A, such that if the function*

$$\bar{G}(\bar{\alpha}) - \sum_{s \in S, r \in R} \mu_0(s, r) G(\bar{\alpha}(s, r); s)$$

is convex (where g and G are defined as in Proposition 8), then all symmetric BNE are constrained efficient.

Proof. See the appendix, 11.15. □

Proposition 10 provides sufficient conditions under which a constrained efficient equilibrium exists in the game with endogenous information. If the game is efficient under exogenous information, and if agents have cost functions with associated divergences that are

invariant with respect to $\bar{\mathcal{A}}$, then there are no externalities related to the cost of information acquisition.

Our proof of Proposition 10 shows something stronger: that all symmetric BNE are critical points of the planner's problem. If strategic complementarities are not too strong (G is not too convex), then the social planner's problem is concave, and consequently all symmetric BNE are maximizers, and hence constrained efficient. The requirement that strategic complementarities not be too strong is the analog of our assumption that $\beta < 1$ in our beauty contest example. Note also that even in games of complete information, with strong strategic complementarities, there are in general multiple Pareto-ranked equilibria.

A related result appears in work by Angeletos and Sastry [2019], who study constrained efficiency of equilibria in a Walrasian setting. Angeletos and Sastry [2019] consider two cases: one in which agents can track only the exogenous state, and one in which agents can also track endogenous prices. They show that the two cases are equivalent, and equivalent to the planner's problem they consider (which is different than our planner's problem), if agents' information costs are invariant in the standard sense (e.g. as in mutual information). The condition we offer in Proposition 10 is weaker than the invariance condition in Angeletos and Sastry [2019] as it only imposes invariance with respect to embeddings in $\bar{\mathcal{A}}$, rather than all possible embeddings. This is a useful distinction in our setting when considering necessity, which we turn to next.

8.2 Inefficiency in Information Acquisition.

We begin by defining a notion of generic non-invariance. Loosely speaking, a cost function is generically non-variant with respect to $\bar{\mathcal{A}}$ if perturbations to the aggregate strategy affect the cost of information.

Definition 14. *A cost function C^Ω is **generically non-invariant with respect to** $\bar{\mathcal{A}}$ **at** $\mu_0 \in \mathcal{U}_0$ if, for all $(\nu^\Omega \in \mathcal{V}_0^\Omega, \bar{\alpha} \in \bar{\mathcal{A}})$ except possibly at a set of isolated points, there exists an $\omega^j \in \Omega$ with*

$$\sum_{s \in S, r \in R} \nu^\Omega \left(\omega^j | s, r, \bar{a}_0 \right) \nabla_{\bar{\alpha}} \{ C^\Omega \left(\nu^\Omega, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha} \} \right) \}(s, r) \neq \vec{0}$$

With generic non-invariance with respect to $\bar{\mathcal{A}}$, there is an interaction between agent's actions and other agents' cost of information. Our necessity result shows that this leads to inefficiency generically in the space of utility functions that would otherwise guarantee efficiency under exogenous information. As in our necessity proof in the exogenous information case, it is convenient to impose an assumption that guarantees interior actions (Assumption 6).

Proposition 11. *Fix the sets S, R, and A, and let C^Ω be a cost function that is generically non-invariant with respect to $\bar{\mathcal{A}}$ at the prior $\mu_0 \in \mathcal{U}_0$. Generically on the space of utility functions u satisfying Assumptions 1 and 6 and the sufficient conditions of Proposition 8, all symmetric BNE are either deterministic or not constrained-efficient.*

43

Proof. See the appendix, 11.16. □

Even when the game is efficient under exogenous information, an externality arises in the endogenous information game if the cost function is non-invariant in $\bar{\mathcal{A}}$. When agents choose how their conditional distribution of actions correlates with the exogenous states and other agents' actions, they affect the conditional distribution of the aggregate action. If the cost function is generically non-invariant in $\bar{\mathcal{A}}$, then the equilibrium conditional distribution of the aggregate action affects all agents' information acquisition costs. Agents do not internalize this information cost effect, and as a result, the equilibrium is inefficient.

In this case the planner chooses a different signal structure than the one that arises in equilibrium—one that instead internalizes the aforementioned information cost externality. In contrast, if the agents' cost structures are invariant with respect to $\bar{\mathcal{A}}$, then no such externality arises, and the planner's solution coincides with the equilibrium strategy profile.

Our formulation of this proposition side-steps the issue of deterministic equilibria (Definition 9), which also arose in our analysis of non-fundamental volatility. In a deterministic equilibrium, agents do not gather information, and consequently their information costs are zero. As a result, if both the planner and the agents are at this corner, there is no externality even if the cost function is generically non-invariant.

8.3 Interpretation and Remarks

Does efficiency under exogenous information imply efficiency under endogenous information acquisition? In our game, the answer is no. Propositions 10 and 11 demonstrate that efficiency under endogenous information requires an extra condition on information costs: namely, invariance with respect to $\bar{\mathcal{A}}$.

This condition arises from the influence that agents' actions have on other agents' information costs. Such a channel is natural if we imagine that agents learn in part by observing other agents' actions; in this case, it is intuitive to suppose that some actions reveal more information than other actions.

With exogenous signals, which we interpret as restricting conditional action distributions to a convex set, the importance of the distinction between signals about exogenous states and signals about endogenous objects has been highlighted in several papers; see, e.g. Laffont [1985], Angeletos and Pavan [2009], Amador and Weill [2010], Vives [2017], Angeletos, Iovino, and La'O [2020]. When agents receive exogenous signals about endogenous objects, an information-aggregation externality arises: agents do not take into account how their own use of information affects the information content of these signals. In these environments, the planner may wish for agents to use their information in a way that differs from what is privately-optimal, in order to improve the aggregation of information in the endogenous signals.

With endogenous acquisition of information, these issues are subtler. Many authors in the literature have assumed both that agents can only acquire signals about exogenous states and that the cost of such signals does not depend on the actions of other agents. That is, they have implicitly assumed invariance with respect to $\bar{\mathcal{A}}$. Our analysis makes clear that by doing so, one is effectively substituting a fixed convex set (the exogenous signals case) with a convex cost function invariant with respect to $\bar{\mathcal{A}}$. We show that this in and of itself unsurprisingly does not alter whether the game is efficient or inefficient. Put simply, in this context, efficiency in the use of information automatically implies efficiency in the acquisition of information. Although this result may not be not well-known in this literature, it has been shown in some specific contexts.[27]

Our results offer a different perspective. Assuming that agents receive signals that condition only on exogenous states is without loss of generality (Lemma 1), but it does not imply that agents are not "learning about the actions of others" because those actions are in equilibrium a deterministic function of the exogenous states. Put another way, when the prior $\mu \in \mathcal{U}$ over exogenous states and endogenous outcomes is degenerate, and it always is in our framework, which variables the signal structure conditions on cannot be interpreted as describing what the agent is paying attention to. Instead, it is the invariance or lack thereof of the information cost *with respect to* $\bar{\mathcal{A}}$ that describes whether agents are attending to the exogenous state, endogenous outcomes, or some combination thereof. Moreover, it is this condition that determines the constrained efficiency properties of equilibria.

9 Conclusion

In this paper, we have explored the relationship between information cost functions and the properties of equilibria in large games with strategic interaction and rationally inattentive agents. Under the assumption of posterior separability, we have demonstrated the close connection between certain properties of information cost functions—namely, partial monotonicity and partial invariance—and whether or not the equilibrium is efficient/exhibits non-fundamental volatility. We have interpreted these forms of invariance as describing whether or not it is possible to learn directly about the actions of others and whether or not public signals are available to the agent.

Efficiency holds only when cost functions are invariant in endogenous actions and when there are no externalities under exogenous information. The class of utility functions that rules out externalities under exogenous information has a particular functional form; this form generalizes the characterization of Angeletos and Pavan [2007] to games that are not necessarily

---

[27]See e.g. Online Appendix A of Angeletos and La'O [2020], the analysis in Angeletos and Sastry [2019] of non-price-tracking economies, or the analysis of Colombo, Femminis, and Pavan [2014] when action dispersion does not enter payoffs.

linear-quadratic. Non-fundamental volatility exists when information costs are non-monotone in payoff-irrelevant dimensions of the exogenous state space. The standard rational inattention cost function, mutual information, leads to both efficiency and zero non-fundamental volatility. In contrast, Fisher information, an alternative cost function proposed by Hébert and Woodford [2020], leads to both non-fundamental volatility and inefficiency.

References

Hassan Afrouzi. Strategic inattention, inflation dynamics, and the non-neutrality of money. *Columbia mimeo*, 2019. 1

Manuel Amador and Pierre-Olivier Weill. Learning from prices: Public communication and welfare. *Journal of Political Economy*, 118(5):866–907, 2010. 1, 20, 25, 8.3

Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191. American Mathematical Soc., 2007. 1, 5.4

George-Marios Angeletos and Jennifer La'O. Noisy business cycles. In *NBER Macroeconomics Annual 2009, Volume 24*, pages 319–378. University of Chicago Press, 2010. 1

George-Marios Angeletos and Jennifer La'O. Sentiments. *Econometrica*, 81(2):739–779, 2013. 1

George-Marios Angeletos and Jennifer La'O. Optimal monetary policy with informational frictions. *Journal of Political Economy (forthcoming)*, 2020. 27

George-Marios Angeletos and Chen Lian. Incomplete information in macroeconomics: Accommodating frictions in coordination. *Handbook of Macroeconomics*, 2:1065–1240, 2016. 1

George-Marios Angeletos and Alessandro Pavan. Efficient use of information and social value of information. *Econometrica*, 75(4):1103–1142, 2007. 1, 1, 1, 2, 2.2, 2.2, 7, 7.2, 7.3, 7.3, 7.3, 7.4, 9

George-Marios Angeletos and Alessandro Pavan. Policy with dispersed information. *Journal of the European Economic Association*, 7(1):11–60, 2009. 1, 20, 25, 8.3

George-Marios Angeletos and Karthik Sastry. Inattentive economies: General equilibrium and welfare theorems. *MIT mimeo*, 2019. 1, 1, 2, 10, 8.1, 27

George-Marios Angeletos, Luigi Iovino, and Jennifer La'O. Efficiency and policy with endogenous learning. *MIT/Columbia/Bocconi mimeo*, 2020. 1, 20, 25, 8.3

Arindam Banerjee, Xin Guo, and Hui Wang. On the optimality of conditional expectation as a bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, 2005. 11.12

Dirk Bergemann and Stephen Morris. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 2013. doi: 10.3982/ECTA11105. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11105. 1, 6.2

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, 24(2):265–272, 1953. 2, 7.1

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203, 2015. 3.2, 15

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. *Columbia mimeo*, 2019. 1, 1, 1, 1, 3.2, 3.2, 5.4

Nikolai Nikolaevich Chentsov. *Statistical decision rules and optimal inference*. Number 53. American Mathematical Soc., 1982. 1, 5.4

Luca Colombo, Gianluca Femminis, and Alessandro Pavan. Information acquisition and welfare. *The Review of Economic Studies*, 81(4):1438–1483, 2014. 1, 2, 2.2, 2.2, 26, 27

Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. *Columbia mimeo*, 2019. 1

Tommaso Denti. Unrestricted information acquisition. *Cornell University mimeo*, 2019. 1, 1, 2.2, 4

Emmanuel Farhi and Iván Werning. A theory of macroprudential policies in the presence of nominal rigidities. *Econometrica*, 84(5):1645–1704, 2016. 6.1

Drew Fudenberg and Jean Tirole. Game theory, 1991. *Cambridge, Massachusetts*, 393(12):80, 1991. 11

John D Geanakoplos and Heraklis M Polemarchakis. Existence, regularity, and constrained suboptimality of competitive allocations. *Essays in Honor of Kenneth J. Arrow: Volume 3, Uncertainty, Information, and Communication*, 3:65, 1986. 6.1

Sanford Grossman. On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance*, 31(2):573–585, 1976. doi: 10.1111/j.1540-6261.1976.tb01907.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1976.tb01907.x. 4

Sanford J. Grossman and Joseph E. Stiglitz. Information and competitive price systems. *The American Economic Review Papers and Proceedings*, 66(2):246–253, 1976. 1, 4

Sanford J Grossman and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980. 1, 4

F. A. Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945. ISSN 00028282. URL http://www.jstor.org/stable/1809376. 2

Benjamin Hébert. Moral hazard and the optimality of debt. *The Review of Economic Studies*, 85 (4):2214–2252, 2018. 1, 5.4

Benjamin Hébert and Michael Woodford. Rational inattention when decisions take time. *Unpublished manuscript*, October 2019. 1, 3.2, 15, 3.2, 5.4

Benjamin Hébert and Michael Woodford. Neighborhood-based information costs. Technical report, National Bureau of Economic Research, 2020. (document), 1, 1, 1, 2, 2.1, 2.1, 8, 3.2, 9, 11.1, 11.1, 11.1, 11.2

Christian Hellwig and Laura Veldkamp. Knowing what others know: Coordination motives in information acquisition. *Review of Economic Studies*, 76(1):223–251, 2009. 1

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101 (6):2590–2615, 2011. 11.11

Jean-Jacques M Laffont. On the welfare analysis of rational expectations equilibria with asymmetric information. *Econometrica*, 53(1):1–29, January 1985. 1, 20, 25, 8.3

Bartosz Mackowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, June 2009. 1

Stephen Morris and Hyun Song Shin. Social value of public information. *American Economic Review*, 92(5):1521–1534, 2002. 1, 2

David P Myatt and Chris Wallace. Endogenous information acquisition in coordination games. *The Review of Economic Studies*, 79(1):340–374, 2012. 1

Luigi Paciello and Mirko Wiederholt. Exogenous information, endogenous information, and optimal monetary policy. *Review of Economic Studies*, 81(1):356–388, 2014. 1

Alessandro Pavan. Attention, coordination, and bounded recall. *Northwestern University mimeo*, 2016. 1

Luciano Pomatto, Philipp Strack, and Omer Tamuz. The cost of information. *arXiv preprint arXiv:1812.04211*, 2018. 1, 1, 3.2

R. Radner. Team decision problems. *The Annals of Mathematical Statistics*, 33(3):857–881, 1962. 21

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3): 665 – 690, 2003. ISSN 0304-3932. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information. 1, 1, 2.1

Harald Uhlig. A law of large numbers for large economies. *Economic Theory*, 8(1):41–50, 1996. 4

Stijn Van Nieuwerburgh and Laura Veldkamp. Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2):779–805, 2010. 2.1

Xavier Vives. Aggregation of information in large cournot markets. *Econometrica*, 56(4):851–76, 1988. 21

Xavier Vives. Endogenous public information and welfare in market games. *Review of Economic Studies*, 84(2):935–963, 2017. 1, 20, 25, 8.3

Michael Woodford. Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 2003. 1

10 Non-s-Measurability and Generic Non-Fundamental Volatility

Let us suppose that, given some utility function u satisfying Assumption 1 and divergence D satisfying our assumptions on information costs, there exists an equilibrium $(\nu^A, \bar{\alpha})$ that is not s-measurable (Definition 7), but satisfies $\bar{\alpha}(s, r) = \bar{\alpha}(s, r')$ for all $s \in S$ and $r, r' \in R$. By Lemma 1, it is without loss of generality to suppose that $\nu^A \in \mathcal{V}_0^A$.

Let us now define a sequence of functions $f_\epsilon : A \to A$ with the following properties: $f_\epsilon(a)$ is strictly increasing and differentiable in each dimension of $A \subset \mathbb{R}^L$, strictly convex in a for all $\epsilon > 0$, and satisfies $\lim_{\epsilon \to 0^+} f_\epsilon(a) = a$ for all $a \in A$. Observe that each of these functions is invertible. Given such a sequence of functions, define the utility functions

$$u_\epsilon(a', \bar{a}_\epsilon(s, r), s) = u(f_\epsilon^{-1}(a'), \bar{\alpha}(s, r), s) \, \forall s \in S, r \in R, a' \in A,$$

where, for some $\bar{a}_0 \in \bar{A}$,

$$\bar{a}_\epsilon(s, r) = \int_A a' \nu^A(f_\epsilon^{-1}(a')|s, r, \bar{a}_0) da'.$$

49

By construction, this $\bar{\alpha}_\epsilon$ and $\nu^A_\epsilon(a'|s, r, \bar{a}') = \nu^A(f_\epsilon^{-1}(a')|s, r, \bar{a}_0)$ are an equilibrium of the game with the game with the u_ϵ utility functions.

But now observe that we can rewrite $\bar{a}_\epsilon$ as

$$\bar{a}_\epsilon(s, r) = \int_A f_\epsilon(a')\nu^A(a|s, r, \bar{a}_0)da'.$$

Because the higher moments of the distribution of a given (s, r) depend on r under ν^A, we can find a sequence of functions such that, for some arbitrary s, $\bar{a}_\epsilon(s, r) \neq \bar{a}_\epsilon(s, r')$ for all $\epsilon > 0$ and $r \neq r'$.

We conclude that the case in which the equilibrium is not s-measurable but the economy nevertheless exhibits zero non-fundamental volatility is non-generic, in the sense that arbitrary perturbations to the utility function will generate non-fundamental volatility in this case.

11 Appendix A: Proofs

11.1 Proof of Proposition 1

The problems defined by (7), (8), and (10) follow from the results of Hébert and Woodford [2020]. We first derive (9) and (11).

By the Sherman-Morrison lemma and (7),

$$\psi^T \Sigma^\omega(\lambda, \sigma)\psi = \psi^T \Sigma \psi - \sigma^{-2}\frac{(\psi^T \Sigma \lambda)^2}{1 + \sigma^{-2}\lambda^T \Sigma \lambda}.$$

Using (7) and the matrix determinant lemma,

$$\det((\Sigma^\omega(\lambda, \sigma))^{-1}) = (1 + \sigma^{-2}\lambda^T \Sigma \lambda) \det(\Sigma^{-1}),$$

and therefore with mutual information, using,

$$\ln(\det((\Sigma^\omega(\lambda, \sigma))^{-1})) - \ln(\det(\Sigma^{-1})) = \ln(1 + \sigma^{-2}\lambda^T \Sigma \lambda).$$

With Fisher information, using (7),

$$\text{tr}((\Sigma^\omega(\lambda, \sigma))^{-1}) - \text{tr}(\Sigma^{-1}) = \text{tr}(\sigma^{-2}\lambda\lambda^T)$$
$$= \sigma^{-2}|\lambda|^2.$$

Note that in what follows we will assume $(\sigma^*)^{-2} > 0$ (i.e. that the agents acquire information). With mutual information, the agents will never find it optimal to choose $\sigma^* = 0$, and we will show below that this is also true with the Fisher information cost.

The optimal action with mutual information described by (6) follows from standard Bayesian updating (the fact that $\lambda = \psi$ is known in the literature, and follows from the first-order condition of (9)). In this case,

$$\alpha^* = \frac{(\sigma^*)^{-2}\lambda^T\Sigma\lambda}{(\sigma^*)^{-2}\lambda^T\Sigma\lambda + 1} < 1.$$

We can rewrite (9) as

$$\max_{\alpha^* \in [0,1]} -(\psi^T\Sigma\psi)\alpha^* - \theta\ln(1 - \alpha^*),$$

noting that $\alpha^* = 1$ will never be optimal. The first order condition yields

$$\frac{\theta}{1 - \alpha^*} - (\psi^T\Sigma\psi) \le 0, \tag{40}$$

with equality if $\alpha^* > 0$. We conclude that α^* is continuous in ψ.

From Hébert and Woodford [2020], the optimal action with Fisher information described by (6) simplifies to

$$\alpha^* = \theta^{\frac{1}{2}}|\psi|(\sigma^*)^{-2}.$$

This could also be derived from the first-order condition characterizing λ^*. Note here that our normalization of λ and σ is different from the one used in Hébert and Woodford [2020]. Let the Hébert and Woodford [2020] parameters be $\hat{\lambda} = |\psi|^{-1}\theta^{-\frac{1}{2}}\lambda$ and $\hat{\sigma} = |\psi|^{-1}\theta^{-\frac{1}{2}}\sigma$, noting that a factor of two is omitted due to a difference in the definition of θ. The above equation follows from Hébert and Woodford [2020] and

$$\hat{\lambda}(\hat{\sigma})^{-2} = |\psi|\theta^{\frac{1}{2}}\lambda\sigma^{-2}.$$

The optimal value of $(\sigma^*)^{-2}$ satisfies, by Hébert and Woodford [2020],

$$|(\Sigma^{-1} + \theta^{-1}\hat{\sigma}^{-2}I)^{-1}\psi|^2 \le \theta,$$

with equality if the signal is informative. If the signal is informative, this is

$$|(\Sigma^{-1} + |\psi|^2(\sigma^*)^{-2}I)^{-1}\psi|^2 = \theta,$$

or

$$|(\Sigma^{-1} + \theta^{-\frac{1}{2}}|\psi|\alpha^*I)^{-1}\psi|^2 = \theta \tag{41}$$

It follows that

$$|(\theta^{-\frac{1}{2}}|\psi|\alpha^*I)^{-1}\psi|^2 > \theta,$$

which is

$$\alpha^* < 1.$$

If the signal is uninformative, $\alpha^* = 0$. It follows that α^* is continuous in ψ.

Equilibrium requires that

$$\psi = (1 - \beta)e_1 + \beta \alpha^* \lambda^*,$$

where $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. For both mutual information and Fisher information, α^* and λ^* are continuous in ψ under the optimal policies. It follows that this equation can be understood as a continuous fixed point problem in ψ. Because $\alpha^* < 1$ in both cases and $\beta \in (0, 1)$, this fixed point problem defines a mapping from the unit sphere to the unit sphere. By Brouwer's fixed point theorem, a fixed point exists, and this fixed point characterizes an equilibrium of our linear-quadratic game.

In the case of mutual information, the fixed point satisfies

$$\psi = (1 - \beta)e_1 + \beta \alpha^* \psi,$$

and it immediately follows that the second element of ψ ($\bar{\alpha}_r$) must be zero.

In the case of the Fisher information cost, from Hébert and Woodford [2020] we have (assuming some information is gathered)

$$\hat{\lambda} = (\theta \Sigma^{-1} + (\hat{\sigma})^{-2} I)^{-1} \psi, \tag{42}$$

where I is the identify matrix. Note that this could also be derived from the first-order condition of (10).

It follows immediately that $\bar{\alpha}_r = 0$ cannot be a fixed point if r and s are correlated, provided that some information is gathered. If it were, we would require that the second element of ψ was zero, and hence that the second element of λ is zero. But the off-diagonal element

$$(\theta \Sigma^{-1} + (\hat{\sigma})^{-2} I)^{-1}$$

will be zero if and only if r and s have zero correlation under Σ.

No information will be gathered with the Fisher information cost if $|\Sigma \psi| \leq \theta$. In this case, we must have $\psi = (1 - \beta)e_1$, and therefore a sufficient condition to ensure information gathering is

$$|\Sigma e_1| > \frac{\theta}{1 - \beta}.$$

11.2 Proof of Proposition 2

With the degenerate covariance matrix $\Sigma = \sigma_0^2 \chi \chi^T$, we can simplify (9) and (11). Conjecture that $\chi^T \lambda \neq 0$, so that our normalization is innocuous. In this case, we have

$$\psi^T \Sigma \psi - \sigma^{-2} \frac{(\psi^T \Sigma \lambda)^2}{1 + \sigma^{-2} \lambda^T \Sigma \lambda} = \sigma_0^2 (\psi^T \chi)^2 - \sigma^{-2} \frac{\sigma_0^4 (\psi^T \chi)^2}{1 + \sigma_0^2 \sigma^{-2}},$$

which does not depend on λ. In the mutual information case,

$$C(\sigma) = \theta \ln(1 + \sigma^{-2}\lambda^T \Sigma \lambda) = \theta \ln(1 + \sigma_0^2 \sigma^{-2}),$$

as stated in the text. Hence it follows that the choice of λ is irrelevant in this case, and it is without loss of generality to assume $\lambda \propto \psi$ (which is the limit from the non-degenerate case).

With the Fisher information cost, the agent solves

$$C(\sigma) = \min_{\lambda : \chi^T \lambda = 1} \theta \sigma^{-2} |\lambda|^2,$$

which is $\lambda = |\chi|^{-2}\chi$. Note that this is also the limit of the non-degenerate case (as can be seen from (12) or the explicit solution in (42)). Plugging this into the cost yields

$$\begin{aligned} C(\sigma, \bar{\alpha}_s) &= \theta \sigma^{-2} |\chi|^{-2} \\ &= \theta \sigma^{-2}(1 + \bar{\alpha}_s^2)^{-1}. \end{aligned}$$

Let us consider the fixed point that defines equilibrium. We have (as in the text)

$$\bar{a} = \bar{\alpha}_s s = \alpha^* (\lambda^*)^T x = \alpha^* (\lambda^*)^T \chi s,$$

and therefore by our normalization

$$\bar{\alpha}_s = \alpha^* \in [0, 1).$$

By equations (40) and (41) in the proof of Proposition 1, α^* is continuous in χ. This implies equilibrium existence with either cost function. That is, given a $\bar{\alpha}_s \in [0, 1]$, there exists a best response $\alpha^* \in [0, 1]$, continuous in $\bar{\alpha}_s$, and therefore by Brouwer's fixed point theorem a fixed point exists. A sufficient condition for the Fisher information case to involve information-gathering is, from Hébert and Woodford [2020], $|\Sigma\psi| > \theta$. To rule out equilibrium in which $\bar{\alpha}_s = 0$, it is therefore sufficient to assume

$$\sigma_0^2 > \frac{\theta}{1 - \beta}.$$

By the argument in the text (i.e. that the social planner would choose a mean-consistent $\bar{\alpha}_s$ even if not required to), in the mutual information case the planner solves

$$\max_{\sigma \geq 0, \bar{\alpha}_s \in \mathbb{R}} V(\sigma, \bar{\alpha}_s) - C(\sigma),$$

which is also an equilibrium of the problem in which the agents solve

$$\max_{\sigma \geq 0} V(\sigma, \bar{\alpha}_s) - C(\sigma)$$

and $\bar{\alpha}_s$ is determined by mean-consistency. Observe also that by the definition of the utility function, (1), $V(\sigma, \bar{\alpha}_s)$ is strictly concave in $\bar{\alpha}_s$ and diverges to negative infinity as $\bar{\alpha}_s$ diverges. Therefore, a maximizer exists and is unique. It follows immediately that in the mutual information case, assuming it is optimal to choose a finite σ (gather some information), a solution exists and is also an equilibrium. Note that our normalization was without loss of generality, as

$$\chi^T \lambda \propto \chi^T \psi = 1 - \beta + \beta \bar{\alpha}^* > 0.$$

Now consider the Fisher information cost case. In any equilibrium $(\sigma', \bar{\alpha}'_s)$, we must satisfy mean-consistency. It follows that, assuming $(\sigma')^{-2} > 0$,

$$\frac{\partial}{\partial \bar{\alpha}_s}[V(\sigma', \bar{\alpha}_s) - C(\sigma, \bar{\alpha}_s)]|_{\bar{\alpha}_s = \bar{\alpha}'_s} = 2\theta(\sigma')^{-2}(1 + (\bar{\alpha}'_s)^2)^{-2}(\bar{\alpha}'_s).$$

Hence the equilibrium will be the solution to the planner's problem if and only if $\bar{\alpha}'_s = 0$, but for θ sufficiently small, $\bar{\alpha}_s > 0$.

11.3 Proof of Lemma 1

We begin with the following observation: suppose that for some $\nu, \nu' \in \mathcal{V}^A$ and aggregate action function $\bar{\alpha} \in \bar{\mathcal{A}}$,

$$\nu(\omega|s, r, \bar{\alpha}(s, r)) = \nu'(\omega|s, r, \bar{\alpha}(s, r)).$$

for all $(s, r) \in S \times R$, and $\omega \in \Omega$. It immediately follows by (20) and (21) that

$$\begin{aligned}
\pi\{\nu, \mu\}(\omega) &= \sum_{s \in S, r \in R} \int_{\bar{A}} \nu(\omega|s, r, \bar{a}) \mu_0(s, r) \delta(\bar{a} - \bar{\alpha}(s, r)) d\bar{a} \\
&= \sum_{s \in S, r \in R} \nu(\omega|s, r, \bar{\alpha}(s, r)) \mu_0(s, r) \\
&= \sum_{s \in S, r \in R} \nu'(\omega|s, r, \bar{\alpha}(s, r)) \mu_0(s, r) \\
&= \pi\{\nu', \mu\}.
\end{aligned}$$

Similarly, by (20) and (22),

$$\begin{aligned}
\mu^\omega\{\nu, \mu\}(s, r, \bar{a}) &= \frac{\nu(\omega|s, r, \bar{a}) \mu_0(s, r) \delta(\bar{a} - \bar{\alpha}(s, r))}{\pi\{\nu, \mu\}(\omega)} \\
&= \frac{\nu(\omega|s, r, \bar{\alpha}(s, r)) \mu_0(s, r) \delta(\bar{a} - \bar{\alpha}(s, r))}{\pi\{\nu, \mu\}(\omega)} \\
&= \mu^\omega\{\nu', \mu\}(s, r, \bar{a}).
\end{aligned}$$

11.4 Proof of Proposition 3

Invoking the results of Lemma 1, it is without loss of generality to assume that $\nu^A \in \mathcal{V}_0^A$.

By the finiteness of $S \times R$ and by the compactness for A (Assumption 1), the set $\mathcal{V}_0^A$ is a finite set of measures on the compact subsets of $\mathbb{R}^L$ (i.e. A). Consequently, by Prokhorov's theorem, using the topology of weak convergence, $\mathcal{V}_0^A$ is compact. Therefore, maxima exist; note also that $\mathcal{V}_0^A$ is non-empty and does not depend on $\bar{\alpha}$.

Individual optimality requires that

$$\nu^A \in \max_{\nu \in \mathcal{V}_0^A} \sum_{s \in S, r \in R} \int_{\bar{A}} \left[\int_{A^i} u^i \left(a^j, \bar{a}, s \right) \nu \left(a^j | s, r, \bar{a} \right) da^j \right] \mu \left(s, r, \bar{a} \right) d\bar{a}$$
$$- C^A \left(\nu, \mu \right),$$

where $\mu = \phi_{\bar{A}} \{\mu_0, \bar{\alpha}\}$ as defined in (20), and mean consistency requires that

$$\int_{A^i} a^j \nu^A (a^j | s, r, \bar{\alpha}^i (s, r)) da^j = \bar{\alpha}^i (s, r) \quad \forall s \in S, r \in R \text{ s.t. } \mu_0 (s, r) > 0.$$

We apply the Theorem of the Maximum and Kakutani's fixed point theorem in the usual fashion. Observe by continuity of u (Assumption 1), and by continuity of C (Assumption 3), that the objective function of the individual agent's problem is continuous in $(\nu, \bar{\alpha})$. Consequently, we can invoke the theorem of the maximum. It follows that the agent's optimal policy correspondence $\mathcal{A}^* : \bar{A} \Rightarrow \mathcal{V}_0^A$ is non-empty, upper semi-continuous, and compact-valued. By the concavity of the objective function (due the convexity of the cost function, Assumption 2), the optimal policy correspondence is convex.

Recall that $\bar{A}$ is the set of possible aggregate actions $\bar{\alpha} : S \times R \to \bar{A}$. Let $\bar{a}_0 \in \bar{A}$ be an arbitrary value in $\bar{A}$. Define the function $f : \mathcal{V}_0^A \to \bar{A}$ by

$$f(\nu) = \int_A a^i \nu(a^j | s, r, \bar{a}_0) da^j.$$

Observe that f is continuous and linear in ν, and does not in fact depend on $\bar{a}_0$ for $\nu \in \mathcal{V}_0^A$.

Define the correspondence $F : \bar{A} \Rightarrow \bar{A}$ by composing the correspondences $\mathcal{A}^*$ and the function f,

$$F(\bar{\alpha}) = f(\mathcal{A}^*(\bar{\alpha})).$$

By the upper semi-continuity of $\mathcal{A}^*$ and continuity of f, F is upper semi-continuous. By the non-emptiness of $\mathcal{A}^*$, F is non-empty. By the convexity of $\mathcal{A}^*$ and the linearity of f, F is convex.

By the finiteness of $S \times R$ and the fact that $\bar{A} \subseteq \mathbb{R}^L$, $\bar{A}$ is isomorphic to a subset of $\mathbb{R}^{|S| \times |R| \times L}$. Consequently, by Kakutani's fixed point theorem, there exists a fixed point of the correspondence F. This fixed point $\bar{\alpha}^*$, along with the best response $\mathcal{A}^*(\bar{\alpha}^*) \in \mathcal{V}_0^A$, constitute a symmetric BNE.

11.5 Proof of Lemma 2

Part (i). By definition,

$$\nu_R\{\nu,\mu\}(\omega|s,r,\bar{a}) = \frac{\sum_{r'\in R}\left(\nu\left(\omega|s,r',\bar{a}\right)\mu\left(s,r',\bar{a}\right)\right)}{\gamma_R\{\mu\}\left(s,\bar{a}\right)}, \quad \forall s\in S, r\in R, \bar{a}\in\bar{A}, \omega\in\Omega$$

and observe that this signal structure does not condition on r.

Given this signal structure, the unconditional probability of observing $\omega\in\Omega$ satisfies

$$\pi\{\nu_R\{\nu,\mu\},\mu\}\left(\omega\right) = \sum_{s\in S, r\in R}\int_{\bar{A}}\nu_R\{\nu,\mu\}(s,r,\bar{a})\mu(s,r,\bar{a})d\bar{a}$$

$$= \sum_{s\in S}\int_{\bar{A}}\nu_R\{\nu,\mu\}\{s,\bar{a}\}\gamma_R\{\mu\}(s,\bar{a})d\bar{a}$$

$$= \sum_{s\in S, r\in R}\int_{\bar{A}}\nu\left(\omega|s,r',\bar{a}\right)\mu\left(s,r',\bar{a}\right)d\bar{a}$$

and therefore

$$\pi\{\nu,\mu\} = \pi\{\nu_R\{\nu,\mu\},\mu\}, \qquad \forall\nu\in\mathcal{V}^\Omega, \mu\in\mathcal{U}. \tag{43}$$

Let $\mu^{\omega'}\equiv\mu^\omega\{\nu_R\{\nu,\mu\},\mu\}$, which is

$$\mu^{\omega'}(s,r,\bar{a}) = \frac{\nu_R\{\nu,\mu\}\left(\omega|s,r,\bar{a}\right)\mu\left(s,r,\bar{a}\right)}{\pi\{\nu_R\{\nu,\mu\},\mu\}\left(\omega\right)},$$

and note that for all $\omega\in\Omega$,

$$\gamma_R\{\mu^{\omega'}\}\left(s,\bar{a}\right) = \frac{\sum_{r\in R}\nu_R\{\nu,\mu\}\left(\omega|s,r,\bar{a}\right)\mu\left(s,r,\bar{a}\right)}{\pi\{\nu_R\{\nu,\mu\},\mu\}\left(\omega\right)}$$

$$= \frac{\sum_{r\in R}\nu\left(\omega|s,r,\bar{a}\right)\mu\left(s,r',\bar{a}\right)}{\pi\{\nu,\mu\}\left(\omega\right)}$$

$$= \gamma_R\{\mu^\omega\}\left(s,\bar{a}\right)$$

$$= \frac{\nu_R\{\nu,\mu\}\left(\omega|s,r',\bar{a}\right)\gamma_R\{\mu\}\left(s,\bar{a}\right)}{\pi\{\nu_R\{\nu,\mu\},\mu\}\left(\omega\right)}\ \forall r'\in R,$$

where the last equality follows from the fact that $\nu_R\{\nu,\mu\}\left(\omega|s,r',\bar{a}\right)$ does not condition on r.

We now apply the composition operator $\eta_R:\mathcal{U}\times\mathcal{U}\to\mathcal{U}$ defined in (31). This operation that coarsens its first argument in r, then embeds using the conditional distribution of its second argument. Applying this operator to μ^ω, we have

$$\eta_R\{\mu^\omega,\mu\}\left(s,r,\bar{a}\right) = \frac{\mu\left(s,r,\bar{a}\right)}{\gamma_R\{\mu\}\left(s,\bar{a}\right)}\gamma_R\{\mu^\omega\}\left(s,\bar{a}\right),$$

and therefore

$$\eta_R\{\mu^\omega,\mu\}\left(s,r,\bar{a}\right) = \frac{\mu\left(s,r,\bar{a}\right)}{\gamma_R\{\mu\}\left(s,\bar{a}\right)}\gamma_R\{\mu^{\omega'}\}\left(s,\bar{a}\right)$$

$$= \frac{\nu_R\{\nu,\mu\}\left(\omega|s,r,\bar{a}\right)\mu\left(s,r,\bar{a}\right)}{\pi\{\nu_R\{\nu,\mu\},\mu\}\left(\omega\right)}$$

$$= \mu^{\omega'}(s,r,\bar{a}).$$

By R-monotonicity,
$$D\left(\mu^{\omega}||\mu\right) \geq D\left(\eta_R\{\mu^{\omega},\mu\}||\mu\right) = D\left(\mu^{\omega\prime}||\mu\right),$$

and combining this with (43), we conclude that

$$C^{\Omega}\left(\nu',\mu\right) \geq C^{\Omega}\left(\nu_R\{\nu,\mu\},\mu\right),$$

as required.

Part (ii). We prove the "only if" by contradiction. Suppose there exists $\mu_0, \mu_1 \in \mathcal{U}$ with $\gamma_R\{\mu_1\} \ll \gamma_R\{\mu_0\}$ such that

$$D\left(\mu_1||\mu_0\right) < D\left(\eta_R\{\mu_1,\mu_0\}||\mu_0\right).$$

Define $\Omega = \{\omega_1, \omega_2\}$ and define, for $\epsilon > 0$ sufficiently small, the signal structure $\nu_\epsilon \in \mathcal{V}^{\Omega}$ as follows
$$\nu_\epsilon\left(\omega_1|s,r,\bar{a}\right) = \delta\left(\omega_1\right)\epsilon\frac{\gamma_R\{\mu_1\}\left(s,\bar{a}\right)}{\gamma_R\{\mu_0\}\left(s,\bar{a}\right)}$$

and

$$\nu_\epsilon\left(\omega_2|s,r,\bar{a}\right) = \delta\left(\omega_2\right)\left(1 - \epsilon\frac{\gamma_R\{\mu_1\}\left(s,\bar{a}\right)}{\gamma_R\{\mu_0\}\left(s,\bar{a}\right)}\right),$$

where $\delta(\cdot)$ is the Dirac delta function. Note that this signal structure does not condition on r, and that
$$\eta_R\{\mu_1,\mu_0\}\left(s,r,\bar{a}\right) = \mu_0\left(s,r,\bar{a}\right)\frac{\sum_{r'\in R}\mu_1\left(s,r',\bar{a}\right)}{\sum_{r'\in R}\mu_0\left(s,r',\bar{a}\right)}$$

Therefore

$$\nu_\epsilon\left(\omega_1|s,r,\bar{a}\right) = \delta\left(\omega_1\right)\epsilon\frac{\eta_R\{\mu_1,\mu_0\}\left(s,r,\bar{a}\right)}{\mu_0\left(s,r,\bar{a}\right)}$$

and

$$\nu_\epsilon\left(\omega_2|s,r,\bar{a}\right) = \delta\left(\omega_2\right)\left(1 - \epsilon\frac{\eta_R\{\mu_1,\mu_0\}\left(s,r,\bar{a}\right)}{\mu_0\left(s,r,\bar{a}\right)}\right).$$

Note also that $\pi(\nu_\epsilon,\mu_0)(\omega_1) = \epsilon$ and $\pi(\nu_\epsilon,\mu_0)(\omega_2) = 1-\epsilon$, by construction. Posteriors are therefore

$$\mu^{\omega_1}\{\nu_\epsilon,\mu_0\} = \eta_R\{\mu_1,\mu_0\},$$
$$\mu^{\omega_2}\{\nu_\epsilon,\mu_0\} = \frac{1}{1-\epsilon}\mu_0 - \frac{\epsilon}{1-\epsilon}\eta_R\{\mu_1,\mu_0\}.$$

Now consider instead an alternative signal structure $\nu'_\epsilon \in \mathcal{V}^{\Omega}$, defined by

$$\nu'_\epsilon\left(\omega_1|s,r,\bar{a}\right) = \delta\left(\omega_1\right)\epsilon\frac{\mu_1\left(s,r,\bar{a}\right)}{\mu_0\left(s,r,\bar{a}\right)},$$

$$\nu'_\epsilon\left(\omega_2|s,r,\bar{a}\right) = \delta\left(\omega_2\right)\left(1 - \epsilon\frac{\mu_1\left(s,r,\bar{a}\right)}{\mu_0\left(s,r,\bar{a}\right)}\right).$$

By construction,

$$\nu_R\{\nu'_\epsilon, \mu_0\}(\omega_1|s, r, \bar{a}) = \delta(\omega_1)\,\epsilon \frac{\gamma_R\{\mu_1\}(s, \bar{a})}{\gamma_R\{\mu_0\}(s, \bar{a})} = \nu_\epsilon(\omega_1|s, r, \bar{a}),$$

$$\nu_R\{\nu'_\epsilon, \mu_0\}(\omega_2|s, r, \bar{a}) = \delta(\omega_2)\,(1 - \epsilon \frac{\gamma_R\{\mu_1\}(s, \bar{a})}{\gamma_R\{\mu_0\}(s, \bar{a})}) = \nu_\epsilon(\omega_2|s, r, \bar{a}),$$

and therefore $\nu_R\{\nu'_\epsilon, \mu_0\} = \nu_\epsilon$. Posteriors given ν'_ϵ are given by

$$\mu^{\omega_1}\{\nu'_\epsilon, \mu_0\} = \mu_1$$

$$\mu^{\omega_2}\{\nu'_\epsilon, \mu_0\} = \frac{1}{1-\epsilon}\mu_0 - \frac{\epsilon}{1-\epsilon}\mu_1,$$

and again $\pi(\nu'_\epsilon, \mu_0)(\omega_1) = \epsilon$ and $\pi(\nu'_\epsilon, \mu_0)(\omega_2) = 1 - \epsilon$.

The cost of signal structure ν_ϵ is given by

$$C(\nu_\epsilon, \mu_0) = \epsilon D(\eta_R\{\mu_1, \mu_0\}||\mu_0) + (1-\epsilon) D\left(\mu_0 - \frac{\epsilon}{1-\epsilon}(\eta_R\{\mu_1, \mu_0\} - \mu_0)\middle\|\mu_0\right),$$

and the cost of signal structure ν'_ϵ is likewise given by,

$$C(\nu'_\epsilon, \mu_0) = \epsilon D(\mu_1||\mu_0) + (1-\epsilon) D\left(\mu_0 - \frac{\epsilon}{1-\epsilon}(\mu_1 - \mu_0)\middle\|\mu_0\right).$$

Now consider the difference between these two cost structures:

$$f(\epsilon) \equiv C(\nu'_\epsilon, \mu_0) - C(\nu_\epsilon, \mu_0).$$

By assumption, $D(\mu'||\mu_0)$ is differentiable with respect to μ' at $\mu' = \mu_0$, and the gradient must be zero by the definition of the divergence. Therefore, $f(\epsilon)$ is differentiable with respect to ϵ at $\epsilon = 0^+$ and

$$f'(\epsilon)|_{\epsilon=0^+} = D(\mu_1||\mu_0) - D(\eta_R\{\mu_1, \mu_0\}||\mu_0) < 0.$$

But we must have that

$$C^\Omega(\nu'_\epsilon, \mu_0) - C^\Omega(\nu_R\{\nu'_\epsilon, \mu_0\}, \mu_0) \geq 0$$

for all feasible $\epsilon \geq 0$, and $f(0) = 0$ (both signal structures are uninformative in the limit), a contradiction. We conclude that no such (μ_0, μ_1) exists, and therefore that D satisfies R-monotonicity.

11.6 Proof of Proposition 4

The proof is essentially identical to our existence proof (Proposition 3), and we will refer to the existence proof (Proposition 3) rather than repeat most of the arguments. Let $\bar{\mathcal{A}}_S \subset \bar{\mathcal{A}}$ denote the subset of α functions that are s-measurable. Let $\mathcal{V}^A_{\bar{A}R} \subset \mathcal{V}^A$ denote the set of signal structures who distributions do not in fact depend on r or $\bar{a}$.

By Lemma 2 (characterizing R-monotonicity) and Lemma 1, the optimality policy correspondences $\mathcal{A}^*$ (defined in the proof of Proposition 3) are mappings from $\bar{\mathcal{A}}_S$ (inducing s-measurable priors) to $\mathcal{V}_{\bar{A}R}^A$. By construction, the functions f defined in the proof of Proposition 3 map conditional distributions in $\mathcal{V}_{\bar{A}R}^A$ to functions that are measurable on s. Consequently, the mapping F defined in the proof of Proposition 3 is a map from $\bar{\mathcal{A}}_S$ to $\bar{\mathcal{A}}_S$.

The arguments for the upper semi-continuity, non-emptiness, and convexity of F apply unchanged from the proof of Proposition 3. It follows that a fixed point in $\bar{\mathcal{A}}_S$ exists, and this fixed point constitutes an s-measurable equilibrium.

11.7 Proof of Proposition 5

The "if" part of the proposition is implied by Proposition 4. We therefore focus on the "only-if." To prove this result, we use the following lemma:

Lemma 4. *Suppose the divergence D is not monotone in R on the set of s-measurable priors* $\mathcal{U}^{s-meas}(\mu_0)$. *Then there exists an s-measurable, mean-consistent* $(\nu^A \in \mathcal{V}_0^A, \bar{\alpha} \in \bar{\mathcal{A}})$ *such that* ν^A *has finite support, and a* $\hat{\nu}^A \in \mathcal{V}_0^A$ *such that* $\nu^A = \nu_R\{\hat{\nu}^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\}$ *and*

$$C^A(\hat{\nu}^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}) < C^A(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}).$$

Proof. See the appendix, 11.8. □

The only-if direction of the proposition requires that if D is not monotone in R on the set of s-measurable priors, there exists a utility function satisfying Assumption 1 such that no s-measurable equilibrium exists. Armed with this lemma, we take as given the s-measurable, mean-consistent $(\nu^A, \bar{\alpha})$ described in the lemma, along with the alternative $\hat{\nu}^A$, such that

$$C^A(\hat{\nu}^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}) < C^A(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}).$$

Note that if such an inequality holds, it holds by continuity for some prior μ_0 with full support on $S \times R$.

We will construct a utility function with the property that, if an s-measurable equilibrium were to exist, it could only exist in the neighborhood of this $(\nu^A, \bar{\alpha})$. Let A^* be the finite set of actions that occur with positive probability under ν^A. Choose an arbitrary function $\kappa : S \to \mathbb{R}$, and define, for all $a \in A^*$,

$$\tilde{u}(a, \bar{\alpha}(s, r_0), s) = D(\mu^a || \mu) + \frac{\sum_{r \in R} \mu_0(s, r)\nabla_1\{D(\mu^a || \mu)\}(s, r, \bar{\alpha}(s, r))}{\sum_{r \in R} \mu_0(s, r)} + \kappa(s),$$

where $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$, $\mu^a = \mu^a\{\nu^A, \mu\}$, and ∇_1 denotes the gradient with respect to the first argument. This is the first-order condition associated with choosing ν^A among all s-measurable $\nu \in \mathcal{V}_0^A$, and by the convexity of the cost function it is sufficient for optimality.

We extend the utility function $\tilde{u}$ to other values of $\bar{a}$ using, for all $a \in A^*$,

$$\tilde{u}(a, \bar{a}, s) = \tilde{u}(a, \bar{\alpha}(s, r_0), s) - \chi(a - \bar{\alpha}(s, r_0))^T \cdot (\bar{a} - \bar{\alpha}(s, r_0)) + \frac{\chi}{2}|\bar{a} - \bar{\alpha}(s, r_0)|^2.$$

We extend this utility function to other actions $a \in A \setminus A^*$ by defining

$$\tilde{u}(a, \bar{a}, s) = \min_{a' \in A^*} \tilde{u}(a', \bar{a}, s) - |a' - a|^2.$$

By construction, all actions not in A^* are dominated by some action in A^*. Consequently, with this utility function, any optimal policy will have support only on A^*.

The utility function $\tilde{u}$ is not differentiable at certain points. However, we can define a mollified version of it,

$$u_\delta(a, \bar{a}, s) = \int_A \rho_\delta(a'' - a)\tilde{u}(a'', \bar{a}, s)da'',$$

where $\rho_\delta(z)$ is a smooth symmetric kernel with full support on $z < \delta$. Setting δ sufficiently small ensures that, for some constant c_δ that depends on the kernel,

$$u_\delta(a, \bar{\alpha}(s, r_0), s) = \tilde{u}(a, \bar{\alpha}(s, r_0), s) - c_\delta$$

for all $a \in A^*$, and that u_δ is continuously differentiable, while preserving the property that actions not in A^* are dominated.

It immediately follows that, given the utility function u_δ, $(\nu^A, \bar{\alpha})$ cannot be an equilibrium, and moreover there is no other s-measurable $(\tilde{\nu}^A, \bar{\alpha})$ that can be an equilibrium, because all such signal structures are dominated by $\hat{\nu}^A$.

Let us now rule out equilibrium for other s-measurable $\tilde{\alpha} \in \bar{A}$. We claim that, for sufficiently large values of χ, any equilibrium must lie in the neighborhood of $\bar{\alpha}(s, r_0)$. To see this, suppose that for some $\epsilon > 0$,

$$|\tilde{\alpha}(s, r_0) - \bar{\alpha}(s, r_0)| > \epsilon.$$

Observe by the optimality of the actions in A^* and mean-consistency that any equilibrium value of $\tilde{\alpha}(s, r_0) - \bar{\alpha}(s, r_0)$ must lie in the span of $\{a - \bar{\alpha}(s, r_0)\}_{a \in A^*}$. For sufficiently large χ, for any $a \in A^*$ such that $(a - \bar{\alpha}(s, r_0))^T \cdot (\bar{a} - \bar{\alpha}(s, r_0)) \leq 0$ will be dominated by an action $a' \in A^*$ with $(a' - \bar{\alpha}(s, r_0))^T \cdot (\bar{a} - \bar{\alpha}(s, r_0) > 0$. As a result, under any optimal policy, only the latter category of actions will be chosen, but this contradicts mean-consistency. We conclude that any equilibrium $\tilde{\alpha}$ must be close to $\bar{\alpha}$ for sufficiently large χ.

Now consider a sequence χ_n with $\lim_{n \to \infty} \chi_n = \infty$, and suppose that along this sequence of utility functions there exists an s-measurable equilibrium $(\nu_n^A, \bar{\alpha}_n)$. By construction, $\lim_{n \to \infty} \bar{\alpha}_n = \bar{\alpha}$. By the theorem of the maximum, the optimal policy correspondence for the agents is upper semi-continuous (see the proof of existence, Proposition 3), and consequently $\lim_{n \to \infty} \nu_n^A$ must be an s-measurable best response to $\bar{\alpha}$. But we have already shown that no such policy exists, and consequently for sufficient large χ no s-measurable equilibrium exists.

11.8 Proof of Lemma 4

If D is not r-monotone on $\mathcal{U}^{s-meas}(\mu_0)$, there exists an s-measurable $\bar{\alpha} \in \bar{\mathcal{A}}$, and $\mu' \in \mathcal{U}$ with $\gamma_R\{\mu'\} \ll \gamma_R\{\mu\}$ such that

$$D(\mu'||\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}) < D(\eta_R\{\mu', \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\}||\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\})$$

Observe that if this inequality holds, it holds by continuity for some μ' that is mutually absolutely continuous with μ on S ($\gamma_R\{\mu'\} \sim \gamma_R\{\mu\}$). Similarly, such an inequality must hold for some $\bar{\alpha}$ function whose image is strictly in the interior of A and that is non-degenerate. To streamline the proof, we will assume without loss of generality that μ_0 has full support on $S \times R$.

Define $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$ and $\mu'' = \eta_R\{\mu', \mu\}$, and observe that we must have $\gamma_{\bar{A}}\{\mu''\} \neq \mu_0$. Define the actions

$$a_1 = \sum_{s \in S, r \in R} \bar{\alpha}(s, r)\gamma_{\bar{A}}\{\mu''\}(s, r),$$

$$a_0 = \sum_{s \in S, r \in R} \bar{\alpha}(s, r)\mu_0(s, r).$$

Note that, under the interiority assumption on $\bar{\alpha}$, a_1 and a_0 are interior, and by continuity it is without loss of generality to assume they are not equal to each other.

Now define a shifted $\bar{\alpha}$ function,

$$\tilde{\alpha}(s, r) = \bar{\alpha}(s, r) + \frac{\epsilon\rho\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)}}{1 - \epsilon\rho\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)} - \epsilon(1-\rho)}(\bar{\alpha}(s, r) - a_1) + \frac{\epsilon(1-\rho)}{1 - \epsilon\rho\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)} - \epsilon(1-\rho)}(\bar{\alpha}(s, r) - a_0)$$

for some $\epsilon > 0$ and $\rho \in (0, 1)$. By the interiority of the image of $\bar{\alpha}$, for sufficiently small ϵ, $\tilde{\alpha}(s, r)$ remains in A. By the fact that $\bar{\alpha}$ is s-measurable and $\mu'' = \eta_R\{\mu', \mu\}$, $\tilde{\alpha}(s, r)$ is s-measurable.

Define A^{ext} as the set of extreme points A, and define $A^* = A^{ext} \cup \{a_0, a_1\}$. Because a_0 and a_1 are interior, they are not extreme points and hence not in A^{ext}.

By definition, every value in the image of $\tilde{\alpha}$ is a unique convex combination of elements of A^{ext}. Consequently, there is a unique measure $\tilde{\nu}^A \in \mathcal{V}_0^A$ such that, for each $(s, r) \in S \times R$, and an arbitrary $\bar{a}_0 \in \bar{A}$, $\tilde{\nu}^A$ has support entirely on A^{ext} and that satisfies

$$\int_A a\tilde{\nu}^A(a|s, r, \bar{a}_0)da = \tilde{\alpha}(s, r).$$

Moreover, because $\tilde{\alpha}$ is s-measurable, $\tilde{\nu}^A$ does not condition on r.

Now define the signal structure

$$\nu^A(a|s, r, \bar{a}_0) = (1 - \epsilon\rho\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)} - \epsilon(1-\rho))\tilde{\nu}^A(a|s, r, \bar{a}_0)$$

$$+ \epsilon\rho\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)}\delta(a_1 - a) + \epsilon(1-\rho)\delta(a_0 - a)$$

61

By construction,

$$\int_A a\nu^A(a|s,r,\bar{a}_0)da = \bar{\alpha}(s,r),$$

and the posterior associated with the signal a_1 is (because $a_1 \notin A^{ext}$) equal to μ''. Similarly, the posterior associated with a_0 is the prior.

Moreover, because $\bar{\alpha}$ does not condition on r, and because by definition

$$\frac{\gamma_{\bar{A}}\{\mu''\}(s,r)}{\mu_0(s,r)} = \frac{\gamma_{\bar{A}}\{\eta_R\{\mu',\mu\}\}(s,r)}{\gamma_{\bar{A}}\{\mu\}(s,r)}$$

does not depend on r, ν^A does not condition on r.

Now observe by the convexity and differentiability of the divergence that

$$D(\mu'||\mu) \geq D(\mu''|\mu) + (\mu' - \mu'') \cdot \nabla_1 D(\mu''|\mu),$$

where ∇_1 is the gradient with respect to the first argument. It follows by assumption that

$$(\mu' - \mu'') \cdot \nabla_1 D(\mu''|\mu) < 0.$$

Let $\mu^a = \mu^a\{\nu^A, \mu\}$ be the posteriors associated with the signal structure ν^A, and let $\pi\{\nu^A, \mu\}$ be the unconditional probabilities. Now consider a new set of unconditional probabilities and posteriors, defined by $\hat{\pi}(a) = \pi\{\nu^A, \mu\}(a)$ for $a \in A^*$ and

$$\hat{\mu}^a = \begin{cases} \mu^a & a \notin \{a_0, a_1\} \\ \mu'' + \tau(\mu' - \mu'') & a = a_1 \\ \mu - \frac{\rho}{1-\rho}\tau(\mu' - \mu'') & a = a_0. \end{cases}$$

Observing that $\hat{\pi}(a_0) = \epsilon(1 - \rho)$ and $\hat{\pi}(a_1) = \epsilon\rho$, this set of unconditional probabilities and posteriors is Bayes-consistent with the prior μ. The resulting signal structure conditions on r, because $\mu' \neq \mu''$. Define this signal structure as $\hat{\nu}_\tau$. Considering the derivative, we have

$$\frac{\partial}{\partial\tau}C^A(\hat{\nu}_\tau, \mu) = \epsilon\rho(\mu' - \mu'') \cdot \nabla_1 D(\mu^{a_1}|\mu)$$
$$- \epsilon\frac{\rho}{1-\rho}(\mu' - \mu'') \cdot \nabla_1 D(\mu^{a_0}|\mu)$$

Evaluating at $\tau = 0^+$,

$$\frac{\partial}{\partial\tau}C^A(\hat{\nu}_\tau, \mu)|_{\tau=0^+} = \epsilon\rho(\mu' - \mu'') \cdot \nabla_1 D(\mu''|\mu)) < 0.$$

Hence it follows by convexity that for all $\tau > 0$,

$$C^A(\hat{\nu}_\tau, \mu) < C(\nu^A, \mu),$$

completing the proof.

11.9 Proof of Proposition 6

Suppose that in a neighborhood of some $\mu_0 \in \mathcal{U}_0$, there generically exists an s-measurable equilibrium. For each μ'_0 in this neighborhood, let $\bar{\alpha}(\mu'_0) \in \bar{\mathcal{A}}$ denote the corresponding s-measurable aggregate strategy. By the definition of generic R-non-monotonicity, agents must respond to almost all $(\mu'_0, \bar{\alpha}(\mu'_0))$ by choosing some $\nu^A \in \mathcal{V}^A$ such that either the distribution of $a^i \in A$ depends on r or the signal structure is uninformative. But in the former case the equilibrium cannot be s-measurable. Consequently, generically in all such neighborhoods, either equilibria are not s-measurable or equilibria are deterministic.

11.10 Proof of Proposition 7

Our proof is almost identical to that of Proposition 3. By equation (25), $\mathcal{B}^A(\nu^\Omega)$ is convex and compact (in the topology of weak convergence) subset of $\mathcal{V}_0^A$ (i.e. like $\mathcal{V}_0^A$ in the proof of Proposition 3). To adapt the proof of Proposition 3 to the exogenous information case, suppose that the information cost function is $C^A(\nu_0, \mu) = 0$, which is continuous and convex. The remainder of the proof applies unchanged.

11.11 Proof of Proposition 8

Note that $\nu^A(a^i|s, r, \bar{a}) = \nu^A(a^i|s, r, \bar{a}_0)$ for all $a^i \in A, s \in S, r \in R$, and $\bar{a}, \bar{a}_0 \in \bar{A}$, by the assumption that $\nu^\Omega \in \mathcal{V}_0^\Omega$. We prove that

$$u(a^i, \bar{a}, s) = g(a^i; s) + G(\bar{a}; s) + (a^i - \bar{a}) \cdot \nabla G(\bar{a}; s)$$

with G convex and twice-differentiable and u concave in $\bar{a}$, is sufficient.

Consider the Lagrangean version of the planner's problem (Definition 11), substituting in the mean-consistency constraint,

$$\sup_{\nu^A \in \mathcal{B}^A(\nu^\Omega)} \sum_{s \in S, r \in R} \left[\int_A g(a^i; s) \nu^A\left(a^i|s, r, \bar{a}_0\right) da^i \right] \mu_0(s, r)$$
$$+ \sum_{s \in S, r \in R} \mu_0(s, r) G(\int_A a^j \nu^A\left(a^j|s, r, \bar{a}_0\right) da^j; s),$$

where $\bar{a}_0 \in \bar{A}$ is arbitrary.

Observe that any solution to the problem remains optimal if we replace $g(a^i; s)$ with its concavification (see, e.g., Kamenica and Gentzkow [2011]), $\hat{g}(a^i; s)$. Any action a in the support of the optimal policy ν^A must satisfy, if it is interior,

$$\sum_{s \in S, r \in R} \hat{g}_a(a^j; s) \nu^A\left(a^j|s, r, \bar{a}_0\right) \mu_0(s, r) + \mu_0(s, r) \nabla G(\bar{\alpha}^*(s, r); s) = 0,$$

where $\bar{\alpha}^*(s,r)$ is the aggregate strategy associated with the solution to the planner's problem. If a is on the boundary of A defined by the hyperplane $a \cdot x \leq b$, the necessary condition is

$$z \cdot \left(\sum_{s \in S, r \in R} \hat{g}_a(a^j; s) \nu^A \left(a^j | s, r, \bar{a}_0 \right) \mu_0(s,r) + \mu_0(s,r) \nabla G(\bar{\alpha}^*(s,r); s) \right) \geq 0$$

for any $z \in \mathbb{R}^L$ with $z \cdot x > 0$.

These necessary conditions are exactly the necessary and sufficient conditions in the concavified version of the agent's problem, taking $\bar{\alpha}^*(s,r)$ as given. It follows immediately that there exists a solution $(\nu^{A*}, \bar{\alpha}^*)$ that is both optimal in the planner's problem and for which ν^{A*} is a best-response to $\bar{\alpha}^*$. Therefore, there exists a constrained efficient symmetric BNE.

11.12 Proof of Proposition 9

We will use rely on the following lemma, which characterizes two useful properties of the set A^{ce}.

Lemma 5. *The set A^{ce} is a convex subset of A. For any $a_1, a_2 \in A^{ce}$ and $\rho \in [0,1]$, there exists a finite set R, prior $\mu_0 \in \mathcal{U}_0$, and signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$ such that there exists a solution to the planner's problem $(\nu^A, \bar{\alpha})$ such that, for some $s \in S$ and $r \in R$, ν^A is a two-point distribution on a_1 and a_2,*

$$\nu^A(a^i | s, r, \bar{\alpha}(s,r)) = \rho \delta(a^i - a_1) + (1 - \rho) \delta(a^i - a_2).$$

Proof. See the appendix, 11.13. □

Armed with this lemma, we prove necessity. Let us suppose that for all (μ_0, ν^Ω), the solution to the planner's problem coincides with the competitive equilibrium. By Assumption 5, in any symmetric BNE, agents will take interior actions for all realizations of their signals. Consequently, this must be true in the solution to the planner's problem as well. Moreover, by the strict concavity part of Assumption 5, in any symmetric BNE the best response to any signal realization is unique. Consequently, this also applies in the planner's problem. Recall from above that $\nu^\Omega \in \mathcal{V}_0^\Omega$ does not condition on $\bar{a} \in \bar{A}$.

We can therefore write the planner's problem as

$$\sup_{\alpha \in \mathcal{A}, \bar{\alpha} \in \bar{A}} \inf_{\psi \in \mathbb{R}^{|S| \times |R| \times L}} \sum_{s \in S, r \in R} \left[\int_\Omega u \left(\alpha(\omega^j), \bar{\alpha}(s,r), s \right) \nu^\Omega \left(\omega^j | s, r, \bar{a}_0 \right) d\omega^j \right] \mu_0(s,r)$$

$$+ \sum_{l=1}^{L} \sum_{s \in S, r \in R} \mu_0(s,r) \psi_l(s,r) [\bar{\alpha}_l(s,r) - \int_\Omega \alpha(\omega^j) \nu^\Omega \left(\omega^j | s, r, \bar{a}_0 \right) d\omega^j],$$

where $\mathcal{A} \equiv \Omega \rightarrow A$ is the set of action strategies. Note that we have scaled the multiplier by $\mu_0(s,r)$ to denote that mean consistency need not hold for (s,r) not in the support of μ_0.

The first-order condition for $\alpha_l(\omega^j)$ is

$$\sum_{s \in S, r \in R} \frac{\partial u\left(a, \bar{\alpha}(s, r), s\right)}{\partial a_l}|_{a=\alpha(\omega^j)} \nu^\Omega\left(\omega^j|s, r, \bar{a}_0\right) \mu_0\left(s, r\right)$$

$$- \sum_{s \in S, r \in R} \mu_0\left(s, r\right) \psi_l(s, r) \nu^\Omega\left(\omega^j|s, r, \bar{a}_0\right) = 0.$$

Note that this must hold by the result that the optimal $\alpha(\omega^j)$ is interior. The first-order condition for $\bar{\alpha}_l(s, r)$ is

$$\left(\int_\Omega \frac{\partial u(\alpha(\omega^j), \bar{a}, s)}{\partial \bar{a}_l}|_{\bar{a}=\bar{\alpha}(s,r)} \nu^\Omega\left(\omega^j|s, r, \bar{a}_0\right) d\omega^j\right) \mu_0(s, r)$$

$$+\mu_0\left(s, r\right) \psi_l(s, r) = 0,$$

and note that this also must be interior by mean consistency. In contrast, the private FOC for $a_l(\omega^j)$ is

$$\sum_{s \in S, r \in R} \frac{\partial u\left(a, \bar{\alpha}(s, r), s\right)}{\partial a_l}|_{a=\alpha(\omega^j)} \nu^\Omega\left(\omega^j|s, r, \bar{a}_0\right) \mu_0\left(s, r\right) = 0,$$

which again must hold because the optimal policy is interior. Consequently, because there exists a symmetric BNE that coincides with the solution to the planning problem, we must have, in any solution to the planner's problem,

$$\sum_{s \in S, r \in R} \left(\int_\Omega \frac{\partial u(\alpha(\omega^j), \bar{a}, s)}{\partial \bar{a}_l}|_{\bar{a}=\bar{\alpha}(s,r)} \nu^\Omega\left(\omega^j|s, r, \bar{a}_0\right) d\omega^j\right) \mu_0(s, r) \nu^\Omega\left(\omega^i|s, r, \bar{a}_0\right) = 0 \qquad (44)$$

for all $l \in \{1, \ldots, L\}$, $\omega^i \in \Omega$, and $s \in S, r \in R$ such that $\mu_0(s, r) > 0$.

Suppose that, for some (μ_0, ν^Ω) there exists a solution to the planner's problem $(\nu^A, \bar{\alpha})$ such that, for some $s_0 \in S$ and $r_0 \in R$ with $\mu_0(s_0, r_0) > 0$ and $l \in \{1, \ldots, L\}$,

$$\int_A \frac{\partial u(\alpha(\omega^j), \bar{a}, s)}{\partial \bar{a}_l}|_{\bar{a}=\bar{\alpha}(s_0,r_0)} \nu^A\left(a^j|s_0, r_0, \bar{a}_0\right) da^j \neq 0.$$

By our assumption on the cardinality of Ω, it is without loss of generality to suppose there are elements of Ω that does not occur with positive probability under ν^Ω. Consider a new signal structure defined by

$$\nu_\epsilon^\Omega(\omega|s, r, \bar{a}_0) = \begin{cases} \nu^\Omega(\omega|s, r, \bar{a}_0) & (s, r) \neq (s_0, r_0) \\ (1 - \epsilon)\nu^\Omega(\omega|s, r, \bar{a}_0) + \epsilon\delta(\omega - \omega_0) & (s, r) = (s_0, r_0). \end{cases}$$

Applying (44) to the model with this signal structure and the prior μ_0, in the case of the signal ω_0 we must have

$$\int_\Omega \frac{\partial u(\alpha_\epsilon(\omega^j), \bar{a}, s_0)}{\partial \bar{a}_l}|_{\bar{a}=\bar{\alpha}_\epsilon(s_0,r_0)} \nu_\epsilon^\Omega\left(\omega^j|s_0, r_0, \bar{a}_0\right) d\omega^j = 0,$$

where α_ϵ and $\bar{\alpha}_\epsilon$ satisfy

$$\sum_{s\in S, r\in R} \frac{\partial u\left(a, \bar{\alpha}_\epsilon(s,r), s\right)}{\partial a_l}\big|_{a=\alpha_\epsilon(\omega^i)}\nu_\epsilon^\Omega\left(\omega^i|s,r,\bar{a}_0\right)\mu_0\left(s,r\right) = 0 \,\forall\, l \in \{1,\ldots,L\}, \omega^i \in \Omega,$$

and

$$\int_\Omega \alpha_\epsilon(\omega^i)\nu_\epsilon^\Omega(\omega^i|s,r,\bar{a}_0)d\omega^i = \bar{\alpha}_\epsilon\left(s,r\right) \,\forall s \in S, r \in R\ s.t.\ \mu_0(s,r) > 0.$$

Here, these should be understood as policies characterizing the solution to the planner's problem given ϵ, and by assumption these solutions coincide with a symmetric BNE.

Taking the limit as $\epsilon \to 0^+$, we have by strict concavity and continuity of derivatives that

$$\lim_{\epsilon\to 0^+} \alpha_\epsilon(\omega^i) = \alpha(\omega^i) \,\forall \omega \neq \omega_0$$

and $\lim_{\epsilon\to 0^+} \bar{\alpha}_\epsilon(s,r) = \bar{\alpha}(s,r)$. But this implies

$$\lim_{\epsilon\to 0^+} \int_\Omega \frac{\partial u(\alpha_\epsilon(\omega^j), \bar{a}, s_0)}{\partial \bar{a}_l}\big|_{\bar{a}=\bar{\alpha}_\epsilon(s_0, r_0)}\nu_\epsilon^\Omega\left(\omega^j|s_0, r_0, \bar{a}_0\right)d\omega^j \neq 0,$$

a contradiction. It follows that, for all (μ_0, ν^Ω) and associated solutions to the planner's problem $(\nu^A, \bar{\alpha})$, we have

$$\int_A \frac{\partial u(a^j, \bar{a}, s)}{\partial \bar{a}_l}\big|_{\bar{a}=\bar{\alpha}(s,r)}\nu^A\left(a^j|s,r,\bar{a}_0\right)da^j = 0 \,\forall s \in S, r \in R\ s.t.\ \mu_0(s,r) > 0. \tag{45}$$

Let us now invoke Lemma 5, and observe that one-point (degenerate) ν^A also exist, placing all support on one $a \in A^{ce}$. It follows that we must have

$$\frac{\partial u(a, \bar{a}, s)}{\partial \bar{a}_l}\big|_{\bar{a}=a} = 0,$$

implying by concavity that for all $s \in S$ and $a \in A^{ce}$,

$$a \in \arg\max_{\bar{a}\in A^{ce}} u(a, \bar{a}, s).$$

Define the function

$$f(a, \bar{a}, s) = u(a, a, s) - u(a, \bar{a}, s).$$

It follows immediately that

$$a \in \arg\min_{\bar{a}\in A^{ce}} f(a, \bar{a}, s),$$

and $f(a, a, s) = 0$ by construction. Consequently, on the domain $A^{ce} \times A^{ce} \times S$, $f(a, \bar{a}, s)$ is a weakly positive, continuously twice-differentiable function. Moreover, by mean-consistency, (45), and Lemma 5, for all two-point measures $\kappa \in \Delta(A^{ce})$,

$$\int_{A^{ce}} a\kappa(a)da \in \arg\min_{\bar{a}\in A^{ce}} \int_{A^{ce}} f(a, \bar{a}; s)\kappa(a)da.$$

66

By theorem 4 of Banerjee et al. [2005] (which is proven using only two-point measures; see also the discussion in that paper on restrictions to subspaces of $\mathbb{R}^L$), it follows that

$$f(a, \bar{a}; s) = G(a; s) - G(\bar{a}; s) - (a - \bar{a}) \cdot \nabla G(\bar{a}; s).$$

Defining

$$g(a; s) = u(a, a, s) - G(a; s)$$

proves the result.

Note that by strict concavity and twice-differentiability of u, we have, for arbitrary vectors z,

$$\frac{\partial^2 u(a, \bar{a} + \epsilon z, s)}{\partial^2 \epsilon}\Big|_{\epsilon=0} = (a - \bar{a}) \cdot \frac{\partial^2 \nabla G(\bar{a} + \epsilon z; s)}{\partial^2 \epsilon}\Big|_{\epsilon=0} - \frac{\partial^2 G(\bar{a} + \epsilon z; s)}{\partial^2 \epsilon}\Big|_{\epsilon=0},$$

and hence the gradient of G is twice-differentiable. Moreover, evaluating at $\bar{a} = a$, we have

$$\frac{\partial^2 u(\bar{a}, \bar{a} + \epsilon z, s)}{\partial^2 \epsilon}\Big|_{\epsilon=0} = -\frac{\partial^2 G(\bar{a} + \epsilon z; s)}{\partial^2 \epsilon}\Big|_{\epsilon=0}$$

and therefore by strict concavity G is strictly convex.

11.13   Proof of Lemma 5

Let $a_1 \in A^{ce}$ be an action that occurs in a constrained efficient symmetric BNE given prior and signals (μ_1, ν_1^Ω) and non-fundamental states R_1, after agents observe signal ω_1. Let a_2, (μ_2, ν_2^Ω), R_2, and ω_2 be another such set of objects. Suppose without loss of generality that R_2 does not intersect R_1, and that ν_1^Ω and ν_2^Ω use disjoint parts of the signal alphabet Ω (w.l.o.g. due to our cardinality assumption on Ω). The actions a_1 and a_2 are associated with posteriors $\mu_1^{\omega_1} = \mu^{\omega_1}\{\nu_1^\Omega, \mu_1\}$ and $\mu_2^{\omega_2} = \mu^{\omega_2}\{\nu_2^\Omega, \mu_2\}$.

Define a new set of states, $R = R_1 \cup R_2 \cup \{r_0\}$, and extend the distributions $\mu_1, \mu_2, \mu_1^{\omega_1}, \mu_2^{\omega_2}$ to $\Delta(S \times R)$ by assigning zero probability any state not defined in the original distribution. Define a new prior

$$\mu_{0,\epsilon} = (\frac{1}{2} - \epsilon)\mu_1 + (\frac{1}{2} - \epsilon)\mu_2 + 2\epsilon \mathbf{1}(r = r_0, s = s_0)$$

for some arbitrary $s_0 \in S$. Define a new signal structure $\nu^\Omega \in \mathcal{V}_0^\Omega$ by, for arbitrary $\rho \in [0, 1]$,

$$\nu^\Omega(\omega | s, r, \bar{a}_0) = \begin{cases} \nu_1^\Omega(\omega | s, r, \bar{a}_0) & r \in R_1 \\ \nu_2^\Omega(\omega | s, r, \bar{a}_0) & r \in R_2 \\ \rho\delta(\omega - \omega_1) + (1 - \rho)\delta(\omega - \omega_2) & r = r_0. \end{cases}$$

From this signal structure, it is immediately apparent that the posteriors $\mu^\omega\{\nu^\Omega, \mu_{0,\epsilon}\}$ converge in the limit as ϵ approaches zero to the posteriors associated with either (ν_1^Ω, μ_1) or (ν_2^Ω, μ_2). It follows that in this limit a solution to the social planner's problem conditional on $r \in R_1$ features actions that are identical to the solution associated with (ν_1^Ω, μ_1), and likewise for $r \in R_2$ with

the solution associated with (ν_2^Ω, μ_2). Consequently, in this limit the action given ω_1 is a_1, and likewise with ω_2 and a_2. It follows immediately that in the planner's solution to the merged problem,

$$\lim_{\epsilon \to 0^+} \nu^A\left(a^j|s_0, r_0, \bar{a}_0\right) = \rho\delta(a^j - a_1) + (1 - \rho)\delta(a^j - a_2).$$

We conclude that given any pair (a_1, a_2) from the set of actions that occur in a solution to the planner's problem, any two-point distribution on these actions can occur in some equilibrium.

Let us next demonstrate that A^{ce} is convex. Consider the same setup as in the previous argument, and suppose without loss of generality there is a signal $\omega_0 \in \Omega$ that does not occur under either ν_1^Ω or ν_2^Ω. Let $R = R_1 \cup R_2$ and

$$\mu_0 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2.$$

Consider a signal structure ν_ϵ^Ω such that

$$\mu^\omega\{\nu^\Omega, \mu_0\} = \begin{cases} \mu^\omega\{\nu_1^\Omega, \mu_1\} & \pi\{\nu_1^\Omega, \mu_1\}(\omega) > 0 \\ \mu^\omega\{\nu_2^\Omega, \mu_2\} & \pi\{\nu_2^\Omega, \mu_2\}(\omega) > 0 \\ \rho\mu^{\omega_1}\{\nu_1^\Omega, \mu_1\} + (1 - \rho)\mu^{\omega_1}\{\nu_1^\Omega, \mu_1\} & \omega = \omega_0. \end{cases}$$

By Bayes' rule, such a signal structure exists with

$$\pi\{\nu^\Omega, \mu_0\}(\omega) = \begin{cases} \pi\{\nu_1^\Omega, \mu_1\}(\omega) & \omega \neq \omega_1, \pi\{\nu_1^\Omega, \mu_1\}(\omega) > 0 \\ \pi\{\nu_2^\Omega, \mu_2\}(\omega) & \omega \neq \omega_2, \pi\{\nu_2^\Omega, \mu_2\}(\omega) > 0 \\ \pi\{\nu_1^\Omega, \mu_1\}(\omega) - \epsilon\rho & \omega = \omega_1 \\ \pi\{\nu_2^\Omega, \mu_2\}(\omega) - \epsilon(1 - \rho) & \omega = \omega_2 \\ \epsilon & \omega = \omega_0 \end{cases}$$

for any ϵ sufficiently small.

Now consider the socially optimal action after observing signal ω_0 in the limit as $\epsilon \to 0^+$. In this limit, the solution to the social planner's problem must converge to the aggregate action strategies that are the solutions to the separate problems. By the strict concavity of the utility function, the optimal action must be continuous in ρ, and therefore by the intermediate value theorem must traverse all points on the line segment between a_1 and a_2. It follows that the set of socially optimal actions is a convex set.

11.14   Proof of Lemma 3

By definition, for any $\mu_0 \in \mathcal{U}_0$, $\bar{\alpha} \in \bar{\mathcal{A}}$, $\nu^A \in \mathcal{V}_0^A$, and $\mu \in \mathcal{U}$ such that $\mu = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}$, and an arbitrary $\bar{a}_0 \in \bar{A}$, the posterior after observing $a \in A$ is

$$\mu^a\{\nu^A, \mu\}(s, r, \bar{a}) = \frac{\nu^A\left(a|s, r, \bar{a}_0\right)\mu_0\left(s, r\right)\delta(\bar{a} - \bar{\alpha}(s, r))}{\pi\{\nu^A, \mu\}\left(a\right)},$$

as described in the text. By the definition of $\eta_{\bar{A}}$, (36),

$$\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\} = \delta(\bar{a} - \bar{\alpha}'(s, r))\frac{\nu^A(a|s, r, \bar{a}_0)\mu_0(s, r)}{\pi\{\nu^A, \mu\}(a)}$$
$$= \mu^a\{\nu^A, \mu'\}.$$

Consequently, for all $\mu \in \mathcal{U}$ and $\nu^A \in \mathcal{V}_0^A$,

$$D(\mu^a\{\nu^A, \mu\}||\mu) = D(\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu\}||\eta_{\bar{A}}\{\mu, \mu\}).$$

Note also that for all $\mu_0 \in \mathcal{U}_0$ and $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,

$$\eta_{\bar{A}}\{\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\} = \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\},$$

which implies by the definition of $\mathcal{U}$ that for all $\mu, \mu' \in \mathcal{U}$,

$$\eta_{\bar{A}}\{\mu, \mu'\} = \mu'.$$

With these results, we first prove sufficiency, then necessity.

Part (i). If D is invariant in $\bar{A}$, then by definition

$$D(\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu\}||\eta_{\bar{A}}\{\mu, \mu\}) = D(\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu'\}||\eta_{\bar{A}}\{\mu, \mu'\})$$

for all $\mu, \mu' \in \mathcal{U}$ and $\nu^A \in \mathcal{V}_0^A$. It follows from the results above that

$$D(\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu'\}||\eta_{\bar{A}}\{\mu, \mu'\}) = D(\eta_{\bar{A}}\{\mu^a\{\nu^A, \mu\}, \mu'\}||\mu')$$
$$= D(\mu^a\{\nu^A, \mu'\}||\mu').$$

It immediately follows that for all $\mu_0 \in \mathcal{U}_0$, $\nu^A \in \mathcal{V}_0^A$, and $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,

$$D(\mu^a\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\}||\phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}) = D(\mu^a\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\}||\phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}),$$

and by the argument in the text

$$\pi\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\} = \pi\{\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\}.$$

Consequently, by the definition of posterior separability, (23),

$$C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right) = C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\right)$$

as required.

Part (ii). Suppose that all $\mu_0 \in \mathcal{U}_0$, $\nu^A \in \mathcal{V}_0^A$, and $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$,

$$C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right) = C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\right).$$

It follows by the definition of posterior separability and the argument in the text that we must have

$$D(\mu^a\{\nu^A,\mu\}||\mu) = D(\mu^a\{\nu^A,\mu'\}||\mu').$$

But by the arguments above,

$$D(\mu^a\{\nu^A,\mu\}||\mu) = D(\eta_{\bar{A}}\{\mu^a\{\nu^A,\mu\},\mu\}||\eta_{\bar{A}}\{\mu,\mu\})$$

and

$$D(\mu^a\{\nu^A,\mu'\}||\mu') = D(\eta_{\bar{A}}\{\mu^a\{\nu^A,\mu\},\mu'\}||\eta_{\bar{A}}\{\mu,\mu'\}),$$

and therefore

$$D(\eta_{\bar{A}}\{\mu^a\{\nu^A,\mu\},\mu\}||\eta_{\bar{A}}\{\mu,\mu\}) = D(\eta_{\bar{A}}\{\mu^a\{\nu^A,\mu\},\mu'\}||\eta_{\bar{A}}\{\mu,\mu'\}).$$

Since this must hold for all μ, μ', ν^A, D satisfies the definition of invariance.

11.15   Proof of Proposition 10

Recall the definition of the planner's problem, applying Lemma 1 and using any $\bar{a}_0 \in \bar{A}$,

$$\sup_{\nu^A \in \mathcal{V}_0^A, \bar{\alpha} \in \bar{\mathcal{A}}} \sum_{s \in S, r \in R} \left[\int_A u\left(a^i, \bar{\alpha}\left(s,r\right), s\right) \nu^A\left(a^i|s,r,\bar{a}_0\right) da^i \right] \mu_0\left(s,r\right) - C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right)$$

subject to mean consistency.

Now consider a relaxed version of the problem, without the mean consistency constraint. Let $(\nu^{A*}, \bar{\alpha}^*)$ denote a solution to the relaxed problem, which exists by our continuity assumptions that guarantee compactness (see the proof of Proposition 3).

Because D is invariant with respect to $\bar{A}$,

$$C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right) = C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}'\}\right)$$

for all $\bar{\alpha}, \bar{\alpha}' \in \bar{\mathcal{A}}$. Consequently,

$$\bar{\alpha}^* \in \arg\max_{\bar{\alpha} \in \bar{\mathcal{A}}} \sum_{s \in S, r \in R} \left[\int_A u\left(a^i, \bar{\alpha}\left(s,r\right), s\right) \nu^{A*}\left(a^i|s,r,\bar{a}_0\right) da^i \right] \mu_0\left(s,r\right).$$

As shown in the proof of Proposition 8, given our assumptions on the utility function, $\bar{\alpha}^*$ satisfies mean-consistency even though this was not imposed. We therefore conclude that the mean-consistent symmetric strategy profile $(\nu^{A*}, \bar{\alpha}^*)$ is a constrained-efficient (Definition 13) if it is an equilibrium.

By optimality in the relaxed problem,

$$\nu^{A*} \in \arg\max_{\nu^A \in \mathcal{V}_0^A} \sum_{s \in S, r \in R} \left[\int_A u\left(a^i, \bar{\alpha}^*\left(s,r\right), s\right) \nu^A\left(a^i|s,r,\bar{a}_0\right) da^i \right] \mu_0\left(s,r\right) - C^A\left(\nu^A, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}^*\}\right).$$

It immediately follows that $(\nu^{A*}, \bar{\alpha}^*)$ constitutes a symmetric BNE (Definition 3), proving the result.

By this argument, any symmetric BNE $(\nu^A, \bar{\alpha})$ is a critical point of the social planner's objective function. Consequently, if the social planner's problem is concave, all symmetric BNE are constrained efficient. Define the function

$$f(\bar{\alpha}') = \max_{\nu^A \in \mathcal{V}_0^A} \sum_{s \in S, r \in R} \left[\int_A g(a^i; s) \nu^A \left(a^i | s, r, \bar{a}_0 \right) da^i \right] \mu_0(s, r) - C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha}^* \} \right)$$

subject to $\mu(s, r) \bar{\alpha}'(s, r) = \mu(s, r) \int_A g(a^i; s) \nu^A \left(a^i | s, r, \bar{a}_0 \right) da^i, \ \forall s \in S, r \in R,$

where $g(a; s)$ is part of the functional form of the utility function. Observe that this constraint set is convex in $\bar{\alpha}'$, and the objective does not depend on $\bar{\alpha}'$, and consequently $f(\bar{\alpha}')$ is concave.

By the invariance of D and the functional form of the utility function, the social planner's problem is

$$\max_{\bar{\alpha} \in \bar{\mathcal{A}}} f(\bar{\alpha}') + \sum_{s \in S, r \in R} \mu_0(s, r) G(\bar{\alpha}(s, r); s),$$

noting that the term involving the gradient of G has vanished due to mean-consistency. Defining $\bar{G}(\bar{\alpha}') = -f(\bar{\alpha}')$ proves the result.

11.16 Proof of Proposition 11

Let us assume that u satisfies the conditions of Proposition 8 and Assumptions 1 and 6, and let U be the set of such functions. Endow this set with the topology of pointwise convergence.

The social planner's problem is to solve

$$\sup_{\nu^A \in \mathcal{V}_0^A, \bar{\alpha} \in \bar{\mathcal{A}}} \sum_{s \in S, r \in R} \left[\int_A u \left(a^i, \bar{\alpha}(s, r), s \right) \nu^A \left(a^i | s, r, \bar{a}_0 \right) da^i \right] \mu_0(s, r) - C^A \left(\nu^A, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha} \} \right)$$

subject to mean consistency.

By the finiteness of $S \times R$ and by the compactness for A (Assumption 1), the set $\mathcal{V}_0^A$ is a finite set of measures on the compact subsets of $\mathbb{R}^L$ (i.e. A). Consequently, by Prokhorov's theorem, using the topology of weak convergence, $\mathcal{V}_0^A$ is compact. By the compactness of $\bar{A}$ and the finiteness of $S \times R$, $\bar{\mathcal{A}}$ is compact, and it follows from the definition of mean-consistency that the set of $(\nu^A, \bar{\alpha})$ satisfying mean-consistency is compact. Therefore, a maximizing solution to the social planner's problem exists for all $u \in U$.

We apply the Theorem of the Maximum. Observe by continuity of u (Assumption 1), and by continuity of C (Assumption 3), that the objective function of the planner's problem is continuous in $(\nu, \bar{\alpha})$, and is continuous on U. Consequently, we can invoke the theorem of the maximum. It follows that the planners optimal policy correspondence $\xi^* : U \Rightarrow \mathcal{V}_0^A \times \bar{\mathcal{A}}$ is non-empty, upper semi-continuous, and compact-valued.

Let us suppose there is some $u \in U$ such, generically in a neighborhood around u, a non-deterministic, constrained-efficient symmetric BNE exists. By Assumption 6, these equilibria must involve interior actions.

We next show that generic non-invariance with respect to $\bar{A}$ implies that, generically in this neighborhood, externalities exist. Define, for some arbitrary vector $z \in \mathbb{R}^L$, a continuously differentiable function $f_\epsilon : A \to A$ with the property that

$$f_\epsilon(a) = a + \epsilon z$$

for all $a \in A^*$, and $|f_\epsilon(a) - a| \leq \epsilon |z|$ for all A. Because ν^A places support only on A^*, such a function exists for sufficiently small ϵ.

Now define the utility function

$$
\begin{aligned}
u_\epsilon(a, \bar{a}, s) &= u(f_\epsilon(a), \bar{a}, s) \\
&= g(f_\epsilon(a); s) + G(\bar{a}; s) + (f_\epsilon(a) - \bar{a}) \nabla G(\bar{a}; s),
\end{aligned}
$$

By construction, $\mu_\epsilon \in U$, and $\lim_{\epsilon \to 0} u_\epsilon = u$.

Observe that in both the planner's problem and the competitive equilibrium, this perturbation simply shifts the meaning of all relevant actions in the direction z, holding fixed $\bar{a}$. Consequently, holding fixed $\bar{\alpha}$, the optimal choice of $\nu^\Omega \in \mathcal{V}_0^\Omega$ and $\nu_\epsilon^\Omega \in \mathcal{V}_0^\Omega$ are identical, but all actions are shifted in the direction z. However, this violates mean-consistency, and consequently no $(\nu^\Omega, \bar{\alpha}) \in \xi^*(u)$ can be an equilibrium with the utility function u_ϵ. Similarly, $\xi^*(u_\epsilon)$ and $\xi^*(u_{\epsilon'})$ must be disjoint for distinct ϵ, ϵ'.

By compactness and upper semi-continuity, there exists a convergent sequence of constrained efficient symmetric BNE. By generic non-invariance we must have, along this sequence we must have

$$\sum_{s \in S, r \in R} \nu^\Omega \left(\omega^j | s, r, \bar{a}_0 \right) \nabla_{\bar{\alpha}} \{ C^\Omega \left(\nu_\epsilon^\Omega, \phi_{\bar{A}} \{ \mu_0, \bar{\alpha}_\epsilon \} \right) \} (s, r) \neq \vec{0}$$

for except at isolated values of ϵ. We conclude that this property holds generically in the neighborhood of u.

Let us next show that this creates a wedge between private and social incentives that rules out constrained efficient equilibria. Consider the planner's problem, and observe also that it is without loss of generality to equate signals with actions, and hence to assume there is a signal action taken for each signal realization. We therefore rewrite the Lagrangean version of the planner's problem as

$$
\sup_{\nu^\Omega \in \mathcal{V}_0^\Omega, \alpha \in \mathcal{A}, \bar{\alpha} \in \bar{\mathcal{A}}} \inf_{\psi \in \mathbb{R}^{|S| \times |R| \times L}} \sum_{s \in S, r \in R} \left[\int_\Omega u\left(\alpha(\omega^j), \bar{\alpha}(s,r), s\right) \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) d\omega^j \right] \mu_0(s,r)
$$

$$
- C^\Omega\left(\nu^\Omega, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right)
$$

$$
+ \sum_{l=1}^L \sum_{s \in S, r \in R} \mu_0(s,r) \psi^l(s,r) [\bar{\alpha}_l(s,r) - \int_\Omega \alpha_l(\omega^j) \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) d\omega^j],
$$

where $\mathcal{A} \equiv \Omega \to A$ is the set of action strategies. Note that we have scaled the multiplier by $\mu_0(s,r)$ to denote that the policy need not hold for (s,r) not in the support of μ_0.

Recall that we have supposed that all utility functions in the neighborhood of u have constrained-efficient, non-deterministic symmetric BNE, and therefore involve interior actions. It follows that the first-order conditions are necessary.

The planner's first-order condition for $\alpha_l(\omega^j)$ is

$$
\sum_{s \in S, r \in R} \frac{\partial u\left(a, \bar{\alpha}(s,r), s\right)}{\partial a_l}\Big|_{a=\alpha(\omega^j)} \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) \mu_0(s,r)
$$

$$
- \sum_{s \in S, r \in R} \mu_0(s,r) \psi^l(s,r) \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) = 0.
$$

The first-order condition for $\bar{\alpha}_l(s,r)$ is

$$
\left(\int_\Omega \frac{\partial u\left(\alpha(\omega^j), \bar{a}, s\right)}{\partial \bar{a}_l}\Big|_{\bar{a}=\bar{\alpha}(s,r)} \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) d\omega^j \right) \mu_0(s,r)
$$

$$
- \nabla_{\bar{\alpha}_l}\{C^\Omega\left(\nu^\Omega, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right)\}(s,r)
$$

$$
+ \mu_0(s,r) \psi^l(s,r) = 0,
$$

why simplifies under the condition of Proposition 39 to

$$
\mu_0(s,r) \psi^l(s,r) = \nabla_{\bar{\alpha}_l}\{C^\Omega\left(\nu^\Omega, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right)\}(s,r).
$$

In contrast, the private FOC for $a_l(\omega^j)$ is

$$
\sum_{s \in S, r \in R} \frac{\partial u\left(a, \bar{\alpha}(s,r), s\right)}{\partial a_l}\Big|_{a=\alpha(\omega^j)} \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) \mu_0(s,r) = 0,
$$

Efficiency therefore requires

$$
0 = \sum_{s \in S, r \in R} \nu^\Omega\left(\omega^j | s, r, \bar{a}_0\right) \nabla_{\bar{\alpha}_l}\{C^\Omega\left(\nu^\Omega, \phi_{\bar{A}}\{\mu_0, \bar{\alpha}\}\right)\}(s,r)
$$

for all $\omega^j \in \Omega$. But this is non-zero generically in the neighborhood of u, and therefore generically all equilibria are either deterministic or inefficient.