Inattentive Economies*

George-Marios Angeletos† **Karthik A. Sastry**‡

February 28, 2023

Abstract

We study the efficiency of inattentive but otherwise frictionless economies by augmenting the Arrow-Debreu framework with a flexible form of rational inattention. A version of the First Welfare Theorem holds if attention costs follow an invariance condition that is satisfied by mutual information but not more generally. Conversely, when this condition is violated, a novel form of inefficiency becomes possible even if markets are complete and competitive: welfare may be improved by "simplifying" the price system or by replacing it with other means of communication and coordination. We discuss how these results qualify Hayek's (1945) argument about the informational optimality of the price system and how they link the normative question of interest to an emerging decision-theoretic literature.

*An earlier version circulated in December 2018 under the title "Welfare Theorems for Inattentive Economies." We have benefited from the detailed feedback of the editor and four anonymous referees. We also thank Jakub Steiner and John Leahy for discussing our paper at the 2019 and 2021 ASSA meetings, respectively; and Daron Acemoglu, Basil Halperin, Benjamin Hébert, Jennifer La'O, Stephen Morris, Alessandro Pavan, Doron Ravid, and Harald Uhlig for comments and stimulating discussions. Angeletos acknowledges the support of the National Science Foundation under Grant Number SES-1757198.

†Northwestern University and NBER; angeletos@northwestern.edu

‡Harvard University; ksastry@fas.harvard.edu

1 Introduction

People are inattentive, forgetful, and cognitively constrained. In such circumstances, it is natural to question the functioning of the invisible hand. For instance, Sims (2010, p.170) worries that prices "cannot play their usual market-clearing role" in economies populated by inattentive agents; Gabaix (2019, p.309) argues that the welfare theorems fail in the presence of inattention; and Farhi and Gabaix (2020) identify Pigouvian taxes that correct the effects of (their form of) inattention on consumption and production. But if inattention is *rational*, i.e., the product of optimal choice given cognitive costs, what needs to be corrected?

A related question concerns Hayek's (1945) idea about the "economy of knowledge" afforded by markets:

> We must look at the price system as a mechanism for communicating information if we want to understand its real function. [...] The most significant fact about this system is the economy of knowledge with which it operates, or how little the individual participants need to know in order to be able to take the right action.

The formalization of this idea by Grossman (1981) presumes not only that markets are complete and competitive, but also that prices are observed perfectly and costlessly. But if attention to prices is costly, what exactly is the "economy of knowledge" achieved by markets? And if people make mistakes when reacting to price changes, do markets still provide the best means of communication and coordination?

We study these questions by augmenting the classic Arrow-Debreu framework with a general form of rational inattention and showing two results: a version of the First Welfare Theorem and its converse. The first helps clarify that rational inattention *per se* need not cause inefficiency. The second identifies a novel form of inefficiency, one contradictory to Hayek's conjecture. We show that which of the two scenarios obtains hinges on an invariance condition tightly connected to Sims's (2003) mutual information specification. This links the *normative* questions of interest to the *positive* question of what is the right model of inattention.

Framework. An *inattentive economy* is defined by the familiar objects (commodities, states of nature, preferences, etc.) plus a specification of feasible signals and attention costs. In equilibrium, agents make consumption and production decisions on the basis of imperfect signals of exogenous fundamentals and endogenous prices, and they optimally design these signals subject to a cost. As standard in the literature, the choice of a signal is a representation of costly attention or costly cognition.

In this context, agents may or may not be able to economize costs by paying attention to prices as opposed to the fundamentals underlying them. This casts Hayek's "economy of knowledge" under the umbrella of rational inattention. Furthermore, the equilibrium can be compared to a fictitious social planner that not only regulates agents' consumption, production, and attention choices but also sends flexibly designed messages in place of prices. This formalizes the question of whether markets are an optimal means for communication and coordination. Last but not least, because our planner is merely a representation of Pareto optimality, she fully internalizes not only agents' preferences over commodities but also their attention costs. This non-paternalistic stance, which contrasts with Gabaix (2019) and Farhi and Gabaix (2020), helps explain why we do not *a priori* equate inattention with inefficiency.

Efficiency in Inattentive Economies. Our first main result, Theorem 1, establishes the relevant version of the First Welfare Theorem under the assumption that attention costs satisfy an invariance condition that, as already mentioned, is tightly connected to the mutual information model of Sims (2003). The *positive* content of this condition and its axiomatization is the subject of Caplin et al. (2022) and a growing decision-theoretic literature. Its *normative* content, which is what matters here, becomes clear via our results.

We prove Theorem 1 in two steps. We first show that there is no way to improve upon equilibrium if we restrict the planner's message to correspond with equilibrium prices. To prove this intermediate result, we map the inattentive economy with fixed messages to a "twin" attentive economy that can be readily nested in the standard Arrow-Debreu framework. Under this translation, attention choices and attention costs are subsumed behind appropriately modified preferences and technologies. It follows that, although rational inattention may influence behavior in interesting and complex ways, it does not *by itself* upset efficiency.

We next relax the aforementioned restriction on messages. The planner is still subject to the primitive cognitive friction: agents can only hear a noisy signal of the messages sent in place of prices, and the planner internalizes the cost of any such signal. This step therefore allows us to check if prices are the best means of communication and coordination. The answer is no under invariance, because this condition guarantees that agents' cognition is unaffected by any alternative "repackaging" of the state of nature. Put differently, while agents with invariant costs may optimally choose to perceive the world in a coarse or noisy format, they do not benefit from "social engineering" of simplified market communication in order to do so.

Inefficiency from Cognitive Externalities. The above helps clarify that rational inattention *per se* does not disturb the invisible hand. But there is an important catch, formalized in Corollary 2: under the required invariance condition, there is no welfare loss from sending empty messages and forcing agents to learn directly about the underlying state of nature. This suggests that the most relevant scenario for a meaningful "economy of knowledge" is actually the one where invariance fails.

Our second main result shows that, in this scenario, a novel form of inefficiency becomes possible even if markets are frictionless. This inefficiency is rooted in what we call *cognitive externalities*: the bite of rational inattention on people's behavior and welfare depends on the stochastic properties of prices (e.g., their volatility, coarseness, or "complexity"), which in turn depends on the behavior of others.

We establish this result, which is basically the converse of Theorem 1, in two complementary forms. The first, Proposition 6, shows that, starting from an efficient economy with invariant costs, an appropriate but small perturbation of costs that removes invariance leads to inefficiency. The second, Proposition 7, shows that, if all agents hold the same globally non-invariant cost functional, then the economy is necessarily inefficient, except for two trivial cases: (i) agents collect no information, which would mute externalities by assumption, or (ii) prices do not vary across states, in which case they do not need to be learned.

Hayek Meets Sims. Together, Theorem 1 and its converse offer the following angle on Hayek (1945). When invariance holds, efficiency obtains but the "economy of knowledge" afforded by markets is weak instead of strict. When instead invariance fails, markets have the potential to strictly economize attention costs, but a benevolent social planner may do even better by appropriately simplifying the price system or by introducing other messages in addition to, or in place of, market prices. Hayek's (1945) argument is thus turned on its head: allowing markets to achieve a strict economy of knowledge opens the door to inefficiency, even when all other familiar conditions for efficiency (e.g., complete markets) are satisfied.[1]

Our analysis also links the normative question of interest to recent advances in decision theory and experimental economics. Our invariance condition, spelled out in Definition 6, is tightly connected to Sims's (2003) mutual information cost. Mutual information satisfies invariance by construction, because its foundation in Information Theory and optimal coding ignores the "units" of the state space. Conversely, within the class uniformly posterior separable cost functionals, invariance is uniquely satisfied by mutual information (Caplin et al., 2022). In this sense, invariance and mutual information are "one and the same." But despite its popularity in macroeconomics (Sims, 2010; Maćkowiak et al., 2023), this specification is at odds with various phenomena in the lab (Dean and Neligh, 2022; Woodford, 2012, 2020) as well as in the field (Chetty et al., 2009). This suggests that the source of inefficiency identified here is realistic.

Additional Points and Larger Context. It is worth emphasizing three elements of our analysis, the combination of which is instrumental to our main contribution.

First, we let attention costs take a flexible, essentially arbitrary, functional form. We can thus move beyond mutual information costs, to nest the larger class of posterior separable costs (Caplin and Dean, 2015; Denti, 2022) and even more. In more substantive terms, our formulation can embed a preference for coarseness or sparsity, for more discernible contingencies as in Hébert and Woodford (2021), and for essentially any other notion of what situations are easier for agents to learn or understand. Our analysis shows that only the subset of these models satisfying invariance shuts down cognitive externalities.

Second, we employ an expanded notion of complete markets: consumers are insured against both the usual suspects (e.g., idiosyncratic endowment and taste shocks) and the noise arising from their imperfect attention. While unrealistic, this assumption serves two valuable roles. First, it allows our formal analysis to stay as close as possible to the classic Arrow-Debreu framework. This, in turn, is instrumental to the proof not only of our First Welfare Theorem, but also of additional results regarding equilibrium existence, the Second Welfare Theorem, and sufficient conditions to rule out sunspot volatility. Second, and more substantially, this assumption shuts down the familiar kind of pecuniary externalities that emerge whenever markets are incomplete (Geanakoplos and Polemarchakis, 1986). It therefore not only gives Hayek's argument its best chance but also allows us to isolate a novel source of inefficiency—one originating exclusively from what we call cognitive externalities.[2]

Finally, we discuss how our methodology and insights extend to richer behavioral phenomena, like narrow bracketing, bounded recall, general stochastic choice, and sparsity, provided two key conditions: that these phenomena are rationalized as the consequence of costly cognition; and that these costs are properly accounted for in welfare calculations.

[1] As in the standard Welfare Theorems, our fictitious planner's problem is not meant to be a representation of real-world central planning. Thus, our results do not speak to this particular, more practical, motivation behind Hayek's (1945) argument.

[2] A third implication of our expanded complete-markets notion is that it allows consumers to meet their budget constraints *at equality* despite inattention. See the discussion in Section 7.1.

All in all, our framework and results help delineate three different perspectives on when and how cognitive frictions can justify government intervention. That favored by some behavioral literature (e.g., Gabaix, 2014, 2017; Farhi and Gabaix, 2020) requires one to move outside the rational-inattention framework, neglect cognition costs in welfare calculations, and take a paternalistic stance. We instead treat inattention as the product of rational choice; we give the market mechanism its best chance by imposing that the planner internalizes attention costs as well as that markets are complete; and we nevertheless show that inefficiency can emerge in the form of cognitive externalities once the invariance condition fails. A third approach, left for future work, could retain our non-paternalistic stance but shift the focus to the idea that rational inattention introduces uninsurable idiosyncratic risk.

Related Literature. The literature on rational inattention spurred by Sims (1998, 2003) is voluminous and has delivered valuable insights in a wide range of applications. Despite these advances, our paper is the first to show how to add rational inattention to the general Arrow-Debreu framework and to revisit the Fundamental Welfare Theorems in the presence of inattention. In so doing, we identify a novel form of inefficiency, offer a new angle to Hayek's "economy of knowledge," and connect the normative questions of interest to a modern literature axiomatizing and experimentally testing models of inattention (e.g., Caplin and Dean, 2015; Caplin et al., 2022; Dean and Neligh, 2022; Hébert and Woodford, 2021; Pomatto et al., 2023).

Letting consumers and firms choose their attention to prices in a market environment is akin to letting players obtain information about others' actions in a large game. This link is explored in a recent, complementary paper by Hébert and La'O (2020). These authors establish efficiency of equilibria in their class of games under two conditions: a close cousin of our invariance condition; and a restriction on payoffs akin to our complete-markets assumption (or the netting-out of pecuniary externalities). This shows how our paper's main ideas extend outside the Arrow-Debreu context.[3]

Angeletos and La'O (2020), Colombo, Femminis, and Pavan (2014), Gul, Pesendorfer, and Strzalecki (2017), and Tirole (2015) study efficiency in settings that trivially satisfy our invariance condition, because attention costs are specified as a function of the joint distribution of an agent's signal and the exogenous state of nature. It follows that the conditions for (in)efficiency found in these works relate exclusively to pecuniary or payoff externalities, as opposed to the kind of cognitive externalities identified here.

A similar point applies to Ravid (2020), who studies a bargaining game in which a buyer can flexibly but costly pay attention to a good's quality and a seller's take-it-or-leave-it offer. By including inattention to the terms of trade, that paper shares our theme of how inattentive markets work. However, by assuming mutual information costs, that paper rules out our cognitive externalities. Instead, it produces inefficiency from the interaction of inattention with market power.

Our notion of cognitive externalities nests the "information externalities" in the literature on Noisy Rational Expectations Equilibria (e.g., Grossman and Stiglitz, 1980; Laffont, 1985; Vives, 2017). This literature focuses on settings in which agents are uncertain about a fundamental (e.g., an asset's dividend) and can learn

[3] Hébert and La'O (2020) push the frontier further in a few additional ways. Most notably, they provide a result on non-fundamental volatility which, unlike our Proposition 9, extends to inefficient equilibria.

about it from perfectly observed market prices. That literature's assumption that prices can be observed perfectly, while fundamentals cannot, amounts to a particular violation of our paper's invariance condition. Notwithstanding this connection, the substance is quite different. We are concerned with inefficiencies that obtain when agents are inattentive to (or confused about) prices themselves. In this situation, the cognitive externality concerns exclusively the question of whether allocations could be improved by making prices easier to learn or understand, or perhaps even by shutting down some markets. A recent paper that bridges the two perspectives is Mondria et al. (2022); this studies an asset market in which traders incur a cost for "interpreting" prices (i.e., extracting information about the asset's return).[4]

Finally, a prior literature studies Hayek's hypothesis by showing that market mechanisms can be implemented with the lowest-dimensional message space among the class of Pareto efficient mechanisms (Hurwicz, 1972; Mount and Reiter, 1974; Jordan, 1982). Our approach, by contrast, focuses on what messages economize attention or cognitive costs. This requires that the notion of Pareto efficiency itself be modified to account for such costs, which as already highlighted is an essential and distinctive feature of our approach.

Outline. Section 2 illustrates the main ideas with an example. Section 3 sets up our general framework. Section 4 presents our main result regarding efficiency. Section 5 uses our results to discuss the link between the price system's "economy of knowledge" and the mutual information framework. Section 6 provides additional results. Section 7 discusses possible extensions and open questions. And Section 8 concludes.

2 Example: A Linear-Quadratic-Gaussian Economy

In this section, we study a simple, two-good, exchange economy, corresponding to the linear-quadratic-Gaussian framework used in Sims (2003) and other works. We show that efficiency is guaranteed when costs are proportional to the mutual information of signals with all variables determined exogenously to the consumers (i.e., fundamentals and prices). We then construct three different cost specifications under which efficiency fails and provide economic intuition for the *cognitive externalities* that explain these results.

2.1 Set-up

There is a continuum of agents, indexed by $i \in [0,1]$, and two goods, called "coconuts" and "money." Consumer i enjoys utility $x_i - \frac{1}{2}x_i^2 + y_i$ from coconut consumption $x_i \in \mathbb{R}$ and money consumption $y_i \in \mathbb{R}$. Each consumer's coconut endowment is equal to an aggregate exogenous shock, $\xi \sim N(0,1)$. Each consumer's endowment of money is zero and the price of coconuts in terms of money is $p \in \mathbb{R}$. The consumer's budget is therefore $px_i + y_i \leq p\xi$, for all realizations of these variables.

Agents make decisions under uncertainty about their endowment and the coconut price, which we define together as the *cognition state* $z := (\xi, p)$. We assume that coconut consumption x_i is measurable in a signal ω_i of the cognition state, while money consumption adjusts to meet the budget.

[4] We suspect their specific assumptions about signals and costs amount, under the lens of our analysis, to a joint violation of invariance and monotonicity. The first opens the door to inefficiency, and the second to non-fundamental volatility. Our example in Proposition 4 and Online Appendix C.4, although differently motivated, has a similar flavor.

Agents can pay a utility cost to observe signals of the form $\omega_i = a_1\xi + a_2 p + a_3\eta_i$, where $(a_1, a_2, a_3) \in \mathbb{R}^3$ are choice variables and $\eta_i \sim N(0,1)$ is idiosyncratic noise. Obtaining this signal has a utility cost represented by a function $C : \mathbb{R}^3 \to \mathbb{R}$. We write $\phi(\cdot \mid z; a_1, a_2, a_3)$ as the conditional density of ω_i. We let π denote the density function for the cognition state.[5]

Solving out for y_i in the budget constraint and suppressing the i index, we can write the consumer's problem as the following choice of a consumption rule $x : \mathbb{R} \to \mathbb{R}$ and a signal structure $(a_1, a_2, a_3) \in \mathbb{R}^3$:

$$\max_{x, a_1, a_2, a_3} \int\int \left(x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega)) \right) \phi(\omega | z; a_1, a_2, a_3) \pi(z)\, \mathrm{d}\omega\, \mathrm{d}z - C(a_1, a_2, a_3) \tag{1}$$

Because of the quadratic preferences, each agent's optimal demand for coconuts conditional on receiving state ω is affine in their expectation of the price: $x_i(\omega) = 1 - \mathbb{E}[p \mid \omega_i]$.

Apart from adding tractability, the quasilinear specification has two substantial implications. The first is that the optimal consumption of coconuts is insensitive to wealth, which in turn means that consumers do not care to know the endowment ξ *per se*; they only care to know the price p.[6]Thus, the key issue in our context is not how much agents know about a fundamental (e.g., as studied by Grossman and Stiglitz, 1980) but rather how attentive they are to prices. The second implication is that the marginal utility of wealth is equated across realizations of the idiosyncratic noise in ω. This is preserved in our general framework, thanks to a strong notion of complete markets. Our general framework will nevertheless allow for inequality due to fixed heterogeneity in utility functions and attention costs (cognitive capacities).

2.2 Equilibrium and Efficiency

We assume that a law of large numbers applies across realizations of signal noise in the cross-section of agents and that agents pick symmetric strategies. We can therefore write market clearing as $X(\xi, p) = \xi$ for each realization ξ, where $X(\xi, p) \equiv \int x(\omega)\, \phi(\omega \mid \xi, p)\, \mathrm{d}\omega$. We let $P : \mathbb{R} \to \mathbb{R}$ denote a price functional.

We define a competitive equilibrium as a tuple $(x, a_1, a_2, a_3, P, \pi)$ such that consumers optimize, the market for coconuts clears, and the consumer's prior about z is consistent with the equilibrium price function. Mathematically, the last requirement means that $\pi(\xi, p) = \pi_\Xi(\xi) \cdot \delta_{P(\xi)}(p)$ for all (ξ, p), where π_Ξ is the density of ξ and $\delta_{P(\xi)}$ is the Dirac delta function around $p = P(\xi)$. Note that the agents' prior about z, which enters attention costs, is generated by combining the exogenous prior about ξ with the equilibrium price function. This is where Rational Inattention (RI) meets Rational Expectations Equilibrium (REE).

It may be useful to recast this definition in the familiar metaphor of a Walrasian auctioneer, to understand how, paraphrasing Sims (2010), prices *can* "play their usual market-clearing role" despite inattention to prices. Before ξ and p are realized, consumers use their knowledge of the price functional P to construct their prior over $z = (\xi, p)$ and to design an optimal signal. Then, nature chooses ξ and the auctioneer sets $p = P(\xi)$. Agents observe their noisy signals of p (and of ξ) and submit demands based on these signals. This

[5] In some abuse of terminology, we say "density" when, in equilibrium, the conditional distributions of π in p, given θ, will be delta distributions.

[6] We could have also introduced an attentive supplier, who owned the coconuts but valued only money. Then, ξ would be the random supply of coconuts and would have no intrinsic value for the inattentive consumer

yields an aggregate demand for coconuts that is a deterministic function of p given ξ, since idiosyncratic noise in the agents' signals washes out. The loop closes by requiring that the entire function P is such that this demand meets supply for on any possible realization of ξ.

We now describe our notion of feasibility and efficiency via a social planner's problem. The planner that cannot eliminate the cognitive friction but can otherwise regulate consumers' behavior. Moreover, the planner can send a "message" m, conditional on the state, that substitutes for prices in agents' cognition. For the present example, we assume that the message is affine in the fundamental: $m = \mu_0 + \mu_1 \xi$. Consumers form signals of the form $\omega_i = a_1 \xi + a_2 m + a_3 \eta_i$ and costs remain defined in terms of (a_1, a_2, a_3).

We say that a tuple $(x^*, a_1^*, a_2^*, a_3^*, \mu_0^*, \mu_1^*, \pi^*)$ is efficient if it solves the following program of maximizing payoffs subject to the resource constraint and consistency of the prior:[7]

$$\begin{aligned} \max_{x, a_1, a_2, a_3, \mu_0, \mu_1, \pi} \quad & \int\int \left(x(\omega) - \frac{x(\omega)^2}{2} \right) \phi(\omega | z; a_1, a_2, a_3)\, \pi(z)\, \mathrm{d}\omega\, \mathrm{d}z - C(a_1, a_2, a_3) \\ \text{s.t.} \quad & \int x(\omega)\, \phi(\omega \mid \xi, m; a_1, a_2, a_3)\, \mathrm{d}\omega = \xi \text{ for all } (\xi, m) \\ & \pi(\xi, m) = \pi_\Xi(\xi) \cdot \delta_{\mu_0 + \mu_1 \xi}(m) \text{ for all } (\xi, m) \end{aligned} \tag{2}$$

where $\delta_{\mu_0+\mu_1\xi}$ is the Dirac delta function indicating that the message is $m = \mu_0 + \mu_1 \xi$. We say that an equilibrium $(x, a_1, a_2, a_3, P, \pi)$ is efficient if it solves the above problem with a message equal to the equilibrium price, that is, with $\mu_0 + \mu_1 \xi = P(\xi)$ for all ξ.

The change of notation from p to m highlights the following point. In equilibrium, prices play the standard roles of affecting incentives and clearing markets, as well as the non-standard role of being a component of the state z that agents try to learn or understand. In the planner's solution, incentive compatibility is not an issue and market clearing is replaced by resource feasibility. It follows that prices remain relevant only because of the second role: they are merely "messages" (hence, m) that enter agents' cognition.

2.3 An Efficient Economy With Mutual Information Costs

We first study equilibrium and efficiency in a case of the model where attention costs can be expressed as a function K of the mutual information between ω and z, where $K : \mathbb{R}_+ \to \mathbb{R}$ is increasing, convex, and differentiable, with $K'(0) = 0$. Detailed derivations for all of this section's results are in Online Appendix C.

To solve for equilibrium, we first reexpress attention costs under the (soon to be verified) conjecture that prices are linear in fundamentals, or $p = \psi_0 - \psi_1 \xi$. Because p is a function of ξ, the mutual information between ω and $z = (\xi, p)$ is the same as that between ω and ξ. Thus, $C(a_1, a_2, a_3)$ can be simplified to

$$c_{MI}(\rho) \equiv K(-\log(1-\rho)) \qquad \rho = \frac{(a_1 - \psi_1 a_2)^2}{a_3^2 + (a_1 - \psi_1 a_2)^2} \tag{3}$$

where ρ is the squared correlation between ω_i and ξ. We henceforth refer to ρ as agents' "attention."

[7] When writing the planner's problem in (2) and the consumer's problem in (1), we *presume* that the respective maximums exist. The technicalities are taken care of in our general analysis. Also, our general analysis poses efficiency in terms of Pareto dominance. The planner interpretation is used only to ease the exposition.

We now solve for the equilibrium price. Fix an agent i and suppose that all other agents $j \neq i$ choose a signal described by (a_1^e, a_2^e, a_3^e) and corresponding attention $\rho^e = \frac{(a_1^e - \psi_1 a_2^e)^2}{(a_3^e)^2 + (a_1^e - \psi_1 a_2^e)^2}$. Agent j's coconut demand is $x_j(\omega) = 1 - \mathbb{E}[p \mid \omega_j]$. Substituting in for $\mathbb{E}[p \mid \omega_j]$ and averaging over agents' idiosyncratic signal noise, we can derive *aggregate demand* as $X(\xi, p) = \int_j x_j(\omega)\,\mathrm{d}j = 1 - (\rho^e p + (1-\rho^e)\psi_0)$. Applying market clearing, $X(\xi, p) = \xi$, we find that the coefficients for the price functional are $\psi_0 = 1$ and $\psi_1 = 1/\rho^e$.

Using this expression for prices as a function of others' attention, we can rewrite each agent's own optimal attention choice. First, the flow utility (ignoring information costs) from choosing attention ρ, when others choose attention ρ^e, can be written as the following "benefits function" $b : [0,1]^2 \to \mathbb{R}$:

$$b(\rho, \rho^e) \equiv \max_x \ \mathbb{E}\left[x(\omega) - \frac{x(\omega)^2}{2} + P(\xi)(\xi - x(\omega))\right] = \frac{\rho - 2\rho^e}{2(\rho^e)^2} \tag{4}$$

Combining this expression for attention's "benefits" with the earlier expression for its "costs," we express each consumer's choice as $\max_\rho\{b(\rho,\rho^e) - c_{MI}(\rho)\}$.

Equilibrium is therefore described by the following fixed point equation for attention ρ^e and its first-order condition:

$$\rho^e \in \arg\max_\rho \left\{b(\rho, \rho^e) - c_{MI}(\rho)\right\} \qquad \Rightarrow \qquad \frac{\partial}{\partial \rho} b(\rho^e, \rho^e) = c'(\rho^e) \tag{5}$$

Using almost identical logic, more formally described in Online Appendix C, we can also simplify the planner's problem. Conditional on any message, the planner chooses attention to solve

$$\rho^* \in \arg\max_\rho \left\{b(\rho, \rho) - c_{MI}(\rho)\right\} \qquad \Rightarrow \qquad \frac{\partial}{\partial \rho} b(\rho^*, \rho^*) + \frac{\partial}{\partial \rho^e} b(\rho^*, \rho^*) = c'_{MI}(\rho^*) \tag{6}$$

On the benefits side, compared to the equilibrium fixed-point, the planner internalizes the pecuniary externality whereby increasing one agent's attention affects others' budgets and utility. On the costs side, the planner's trade-off is identical to the agent's. Moreover, this trade-off for the planner is unaffected by the choice of message slope since, as observed earlier, the mutual information cost is invariant to arbitrary recoding of fundamentals in prices.

What are the pecuniary externalities? A direct calculation from (4) yields $\frac{\partial}{\partial \rho^e} b(\rho, \rho) = 0$ for any ρ. Intuitively, this muteness of pecuniary externalities is due to quasilinearity: additional price volatility is absorbed in the linear good, and marginal utility is unaffected. It follows that $\rho^e = \rho^*$, or that the economy's unique equilibrium coincides with the planner's solution. To summarize,

Proposition 1. *With mutual information costs, an equilibrium exists, is unique, and is efficient.*

2.4 Three Inefficient Economies

The economy studied above was efficient because externalities were muted on both the benefits and the costs of attention. We already explained why this was true on the benefit side. We now show that the absence of an externality on the cost side depended critically on the mutual information assumption. Away from this assumption, inefficiency is possible via what we call *cognitive externalities.*

The "Perceptual Distance" Economy. Consider first a case in which agents can only obtain signals of the form $\omega_i = p + a\eta_i$, for some $a > 0$, and pays costs which can be represented by $c(a) = \frac{1}{a^2}$. This cost embodies imperfect observation in the units of prices. For example, such a cost implies that it is more difficult to perceive price changes from \$1.99 to \$2.00 than price changes from \$2.00 to \$2.10. As discussed in Section 5.3, such scale-dependent cognition relates to the decision-theoretic work of Hébert and Woodford (2021) on how to embody "perceptual distance" in information costs.

Using the transformation $\rho = \psi_1^2/(a^2 + \psi_1^2)$ and the observation that, in equilibrium, $\psi_1 = 1/\rho^e$, we can reexpress these "Perceptual Distance" cognitive costs as a function of one's own and others' attention:

$$c_{PD}(\rho, \rho^e) \equiv \left(\frac{\rho}{1-\rho}\right)(\rho^e)^2 \tag{7}$$

The cost of achieving a given attention level ρ, which is the relevant notion of signal "quality" in equation 4, increases with *others'* signal-to-noise ratio. If others are more attentive, then their demands are more elastic and equilibrium prices are less dispersed across states of nature. With "Perceptual Distance" costs, these less dispersed prices are harder to tell apart from one another.

The cost-benefit representation of equations 5 and 6 continues to hold, but with $c_{PD}(\rho, \rho^e)$ replacing $c_{MI}(\rho)$. There is now an additional, cost-based externality, which does not net out across agents. This suggests that the planner prefers lower attention than the equilibrium delivers, as is formalized below:

Proposition 2. *With the costs c_{PD}, any equilibrium is inefficient and the planner prefers to send a message that is more volatile than the equilibrium price.*

The "Variance Penalty" Economy. Consider next an opposite case in which it is relatively *easier* for agents to track prices (or messages) when they are less dispersed. Agents are again restricted to obtain signals of the form $\omega_i = p + a\eta_i$ but now incur a cost in proportion to the signal's variance, or $c(a, \psi_1) = -(a^2 + \psi_1^2)$. One may interpret this as a continuous (full-support) analogue for the idea put forth in Gabaix (2014) that agents prefer to keep track of fewer possible realizations of ω. Following similar steps as above, these "Variance Penalty" costs can be reexpressed as

$$c_{VP}(\rho, \rho^e) \equiv -\left(1 + \sqrt{\frac{1-\rho}{\rho}}\right)(\rho^e)^2 \tag{8}$$

The cognitive externality is now the opposite as in the "Perceptual Distance" economy: higher attention from others lowers price volatility, which makes it *less* costly to obtain any given value for ρ. As in the previous example, this externality makes equilibrium inefficient; but unlike in the previous example, the planner would want to optimally "recode" their message to be *less* volatile than the equilibrium price:

Proposition 3. *With the costs c_{VP}, any equilibrium is inefficient and the planner prefers to send a message that is less volatile than the equilibrium price.*

The Sunspot Economy. As a final example, we construct an inefficient economy in which costs have a mutual information form but still violate the notion of invariance that our later analysis will formalize. This requires a small, but ultimately very consequential, enrichment of the setting. Whereas the exogenous state of nature was previously ξ, the coconut endowment, it is now $\theta = (\xi, \nu)$, where $\nu \sim N(0,1)$ is a new variable, independent of ξ, that plays no fundamental role in the economy. The assumption that ν is a "sunspot" is particularly stark, though it could just as well be a variable that is fundamental for another group of agents that do not interact with the coconut-consuming agents. Signals are restricted to be jointly Gaussian with the cognition state z, where $z = (\theta, p) = (\xi, \nu, p)$. In our context, this means that the signals can be written as

$$\omega_i = a_1\xi + a_2 p + a_3\eta_i + a_4\nu \tag{9}$$

where $(a_1, a_2, a_3, a_4) \in \mathbb{R}^4$ are coefficients and $\eta_i \sim N(0,1)$ is idiosyncratic noise.

We consider two variants of the model. In the first, cognitive costs are a function K of the mutual information between the signal ω_i and the entire state variable (ξ, ν, p); in the second, costs are a function K of the mutual information between the signal ω_i and only the price p. As in our original mutual information model, K is increasing, convex, differentiable, and satisfies $K'(0)$.

The model prediction depends starkly on which of the two cost formulations is used:

Proposition 4. *1. If costs are a function of the mutual information of ω_i and (ξ, ν, p), then there is a unique equilibrium. It is efficient and coincides with the one studied in Proposition 1. Moreover, the equilibrium price is independent from ν.*

2. If costs are a function of the mutual information of ω_i and p, there is a continuum of Pareto-ranked equilibria. One of these is the allocation from Part 1, and this equilibrium is inefficient. In all other equilibria, the equilibrium price is not independent from ν.

To build intuition for these results, consider trying to support the equilibrium allocation of our original mutual information case (Section 2.3 and Proposition 1). As shown in the proof of that result, we can represent each agent's optimal signal as $\omega_i = p + \sigma^e\eta_i$ for some $\sigma^e > 0$; moreover, p is affine in ξ. Now consider a small deviation that transfers some weight from the idiosyncratic noise η_i to the common noise ν, or generates a signal $\omega_i = p + (\sigma^e - \epsilon)\eta_i + \epsilon\nu$ for some small ϵ. Clearly, this does not affect an individual's inference about p, and hence their expected utility ignoring the information cost. Under specification 1 of Proposition 4, this deviation has a positive information cost: constructing the new signal requires "learning" about an additional component of the cognition state, ν. Therefore the deviation is not profitable. The remainder of our proof shows formally that there are no other equilibria which involve observation of ν. Under specification 2 of Proposition 4, by contrast, the ϵ-deviation has no information cost: the sunspot was effectively free to observe, since it is independent from the price. The remainder of the proof shows how to construct equilibria using this logic. We show also that there is a limiting case of these equilibria in which agents observe ξ perfectly at zero information cost, since the price is contaminated with infinite sunspot variance. To summarize, the difference between cases 1 and 2 in Proposition 4 comes from the fact that the second cognitive cost makes conditioning signals on the sunspot costless.

The argument underlying inefficiency in this example does not rely on the sensitivity of payoffs and behavior to *rescaling* prices, as was the case for Propositions 2 and 3. Nonetheless, the argument did rely on the sensitivity of payoffs and behavior to a different transformation of prices: reallocating variance across the fundamental variable ξ and the non-fundamental variable ν. Our general analysis will unify these ideas under one notion of "invariance" of cognitive costs to transformations of the price.

2.5 Summary

In the example of this section, efficiency hinged on whether attention costs were suitably invariant to the stochastic properties of prices and substitute messages. In the subsequent analysis, we will clarify the exact meaning of this invariance in general settings, with multi-dimensional goods, prices, and fundamentals and without parametric restrictions on information acquisition. We will also formally link our results with those in the decision-theoretic literature on costly information acquisition.

3 Model

We now introduce our general framework, which augments a standard Arrow-Debreu economy with rational inattention. We also define our paper's notion of invariance.

3.1 Goods, State of Nature, and Cognition State

There are N (non-contingent) goods, indexed as $n \in \{1, \ldots, N\}$. These goods have random prices $p \in \mathbb{R}_+^N$, the distribution of which is determined in equilibrium.

There is a stochastic state of nature $\theta \in \Theta$, with distribution $\pi_\Theta \in \Delta(\Theta)$. This can encode fundamental uncertainty in the economy (e.g., about endowments, preferences, technologies) or purely non-fundamental uncertainty (e.g., sunspots). We assume, for technical reasons, that Θ is finite and that $\pi_\Theta(\theta) > 0$ for all $\theta \in \Theta$.

To introduce rational inattention in this setting, we need to take a stand on what objects about which an agent can collect information. We denote the collection of such objects by z and refer to it as that agent's *cognition state*.[8] For our main analysis, we take $z = (\theta, p)$; in an extension in Section 7.2, we discuss how to accommodate more flexible choices. It may appear that z double counts θ, because in equilibrium p will be expressible as a function of θ. But this is an *equilibrium* property and, in the decision problem of the typical consumer or firm, we want to model the fact that θ and p appear as two conceptually distinct variables. As one example, our specification will allow the possibility that the probability of rain and the price of umbrellas loom differently in people's cognition despite their tight connection in equilibrium.

We let π be a probability mass function over realizations of $z = (\theta, p)$ and require that it belongs in the set $\mathscr{P}$ defined by the compositions of π_Θ, the prior about θ, with arbitrary functions from Θ to $\mathbb{R}_+^N$:

$$\mathscr{P} \equiv \left\{ \pi : \Theta \times \mathbb{R}_+^N \to [0,1] \ \text{ s.t. } \ \pi(\theta, p) = \pi_\Theta(\theta)\, 1_{f(\theta)}(p), \text{ for some } f : \Theta \to \mathbb{R}_+^N \right\} \tag{10}$$

[8] It is straightforward to let z differ across j to capture heterogeneity in what agents can possibly learn or think about. Alternatively, such heterogeneity can be embedded in the attention cost functional below, while preserving the symmetry in z.

where $1_{f(\theta)}$ is the indicator function for $f(\theta)$. We also define the π-indexed set $Z_\pi \equiv \{(\theta, p) : \pi(\theta, p) > 0\} \subset \Theta \times \mathbb{R}_+^N$, which returns the support of any $\pi \in \mathscr{P}$. We finally define $f_\pi : \Theta \to \Theta \times \mathbb{R}_+^N$ with $f_\pi(\theta) = \{(\theta, p) : \pi(\theta, p) > 0\}$ for any $\theta \in \Theta$, which reverse-engineers the mapping from (10) associated with π.

3.2 Inattentive Consumers

There is a unit-measure continuum of consumers, split into a finite number of types $j \in \{1, ..., J\}$, each with mass μ^j. Each consumer, regardless of type, can consume goods within a set $\mathscr{X}$. We assume $\mathscr{X}$ is a closed rectangle in $\mathbb{R}_+^N$, i.e., $\mathscr{X} = \prod_{n=1}^N [0, x_n^{\max}]$ for a vector $(x_n^{\max})_{n=1}^N \gg 0$ (where $\gg$ denotes strict inequality for each element). But we make these bounds arbitrarily large so as to make their value irrelevant.[9]

Consumers' payoffs over consumption are represented by the Bernoulli utility function $u^j : \mathscr{X} \times \Theta \to \mathbb{R}$. We assume that, conditional on each state θ, this Bernoulli utility function is (i) continuous in x and (ii) represents *weakly monotone* preferences over goods, meaning that it satisfies $u^j(x', \theta) > u^j(x, \theta)$ for each $x, x' \in \mathscr{X}$ such that $x' \gg x$ and each $\theta \in \Theta$. Consumption is described by a function $x : \Omega \to \mathscr{X}$, mapping the realizations of a signal $\omega \in \Omega$ to a consumption level. We assume, for technical simplicity, that Ω is finite.

We model attention by allowing households to design their signals at a cost. Let $\phi(\cdot \mid z) \in \Delta(\Omega)$ denote the signal's likelihood distribution conditional on cognition state realization $z \in Z_\pi$. We define the *signal structure* $\phi = (\phi(\cdot|z))_{z \in Z_\pi}$ as the collection of likelihood functions characterizing the signal. Signal structures are chosen from the set $\Phi \equiv \Delta(\Omega)^{|\Theta|}$. We define the *information structure* as the concatenation of the signal structure and the prior, or $(\phi, \pi) \in \Phi \times \mathscr{P}$. We finally let $C^j : \Phi \times \mathscr{P} \to \mathbb{R}_+$ be, for each type, a cost functional which maps information structures to utility costs. We assume that C^j is continuous in its first argument conditional on the second, to help guarantee that the decision problem stated below is well-posed.[10]

We close the consumer's problem by describing their income and access to insurance. Each type-j consumer has a stochastic endowment represented by $e^j : \Theta \to \mathbb{R}_+^N$ and receives fraction a^j of stochastic aggregate firm profits, which are themselves represented by $\Pi : \Theta \to \mathbb{R}_+^N$. Moreover, consumers have access to complete markets over both the aggregate state of nature θ and the idiosyncratic signal ω. Thus, the consumer solves the following maximization problem:[11]

$$
\begin{aligned}
\max_{x, \phi} \quad & \sum_{\omega \in \Omega, z \in Z_\pi} u^j(x(\omega), \theta)\, \phi(\omega \mid z)\, \pi(z) - C^j[\phi, \pi] \\
\text{s.t.} \quad & \sum_{\omega \in \Omega, z \in Z_\pi} \left(p \cdot x(\omega) - p \cdot e^j(\theta) - a^j \Pi(\theta) \right) \phi(\omega \mid z)\, \pi(z) \leq 0 \\
& x : \Omega \to \mathscr{X}; \quad \phi(\cdot \mid z) \in \Delta(\Omega), \ \forall z \in Z_\pi
\end{aligned}
\tag{11}
$$

[9] These bounds help guarantee the existence of a solution to the consumer's problem. But since endowments are finite, the production set is compact, and $\mu^j > 0$ for all j, it is straightforward to let, for all n, $x_n^{\max}$ be greater than $1/\min_j \mu^j$ times the maximum conceivable amount of the respective good, so no consumer type can reach the bound in a feasible allocation.

[10] Since our probability distributions are finite, it is sufficient to think of continuity in the vector space $[0,1]^{|\Omega| \times |\Theta|}$. If the signal space were continuous, as in Online Appendix D, we need to define continuity with respect to the appropriate weak topology.

[11] In the sums appearing in the consumer's problem (11), z is of course the same variable as the pair (θ, p). The same applies to the firm's problem (12) below. Finally, note that we have let attention costs enter linearly in preferences. This is standard in the literature but is not essential for our main result: we could interpret $c^j = C^j[\phi, \pi]$ as a "bad" and let it enter u^j in a non-linear way alongside the other goods.

As noted above, the consumer has access to complete markets not only over θ, which is standard, but also over ω, which is specific to our analysis. This latter assumption plays a dual role. First, it provides insurance against the idiosyncratic noise in ω. Second, it ensures that agents can satisfy their budget constraint with equality even if none of their consumption choices are perfectly contingent upon the true state of nature. While both of these assumptions may be unrealistic, they are consistent with two principles of our approach: staying as close as possible to the original Arrow-Debreu model and isolating cognitive externalities as a novel source of inefficiency. We revisit these issues in Section 7.1.

3.3 Inattentive Firms

There is a unit-measure continuum of identical firms.[12] Firms, like consumers, choose two objects. The first is a signal parameterized again by a collection $\phi = (\phi(\cdot \mid z))_{z\in Z_\pi} \in \Phi$. The second is a signal-contingent production plan, $y:\Omega \to \mathscr{Y}$, where $\mathscr{Y} \subset \mathbb{R}^N$ is non-empty, is closed, and contains the zero vector (no production). Firms maximize expected profits subject to a technological constraint, which embeds attention costs. Formally, an output vector $y \in \mathscr{Y}$ is feasible for the firm in state θ if and only if it satisfies $H\left(y,c,\theta\right) \le 0$, where $H:\mathbb{R}^{N+1}\times\Theta\to\mathbb{R}$ is continuous and increasing in its first $N+1$ elements and c measures the cost of cognition. The latter is specified as $c = C^F[\phi,\pi]$, where $C^F:\Phi\times\mathscr{P}\to\mathbb{R}$ is defined analogously to the consumers' cost function. Finally, we introduce normalizations such that the firm can always "shut down." That is, we let $C^F(\phi,\pi) = 0$ whenever ϕ is such that $\phi(\omega|z) = \frac{1}{|\Omega|}$ for all ω and all $z \in Z_\pi$ and we assume that $H(0,0,\theta) < 0$.

Putting everything together, firm behavior is summarized in the following program:

$$\begin{aligned} \max_{y,\phi} \quad & \sum_{\omega\in\Omega, z\in Z_\pi} (p\cdot y(\omega))\, \phi(\omega \mid z)\, \pi(z) \\ \text{s.t. } \; & H\left(y(\omega), C^F[\phi,\pi],\theta\right) \le 0, \forall\,(\omega,\theta) : \phi(\omega \mid f_\pi(\theta)) > 0 \\ & y:\Omega\to\mathscr{Y}; \quad \phi(\cdot \mid z) \in \Delta(\Omega),\ \forall\, z \in Z_\pi \end{aligned} \tag{12}$$

This formulation treats firm cognition as a non-tradable, "in-house" production activity that diverts resources from production (insofar as it increases C^F and, thereby, reduces H), but also lets production plans respond more efficiently to demand and supply shocks (insofar as it allows ω to be more informative of z). Note that this readily nests the scenario in which attention costs emerge as a linear penalty in profits by letting $H(y,c,\theta) = \tilde{H}(y + \nu c,\theta)$ for some function $\tilde{H}$ and some constant $\nu \in \mathbb{R}_+$.

3.4 Equilibrium, Feasibility, and Efficiency

Throughout, we focus on equilibria in which strategies are symmetric *within* types. This is without serious loss of generality, because we can always partition types into sub-types with the opportunity to make different decisions. We thus define equilibrium as follows:

Definition 1. An equilibrium is a profile of consumption and production strategies, $\left((x^j)_{j=1}^J, y\right)$, attention choices, $\left((\phi^j)_{j=1}^J, \phi^F\right)$, a price function $P:\Theta\to\mathbb{R}_+^N$, and a prior $\pi\in\mathscr{P}$ such that:

[12] It would be straightforward to add types of firms as well, but we abstract from this for simplicity.

1. Consumers and firms optimize, respectively solving programs (11) and (12), taking as given π.

2. Markets clear, that is for all $\theta \in \Theta$,

$$\sum_{j=1}^{J} \mu^j \overline{x}^j(\theta) = \sum_{j=1}^{J} \mu^j e^j(\theta) + \overline{y}(\theta) \tag{13}$$

   where $\overline{x}^j(\theta) \equiv \sum_{\omega\in\Omega} x^j(\omega)\,\phi^j(\omega \mid f_\pi(\theta))$ and $\overline{y}(\theta) \equiv \sum_{\omega\in\Omega} y(\omega)\,\phi^F(\omega \mid f_\pi(\theta))$ are, respectively, the average demand of type-j consumers and the average supply of firms in state θ.

3. Profits are rebated to consumers, that is $\Pi(\theta) = p(\theta)\cdot\overline{y}(\theta)$ for all $\theta \in \Theta$.

4. The prior π about z is consistent with the price functional, that is $f_\pi(\theta) = (\theta, P(\theta))$ for all $\theta \in \Theta$.

The following two properties carry over from the equilibrium definition in the example. First, the prior about z, which enters each agent's cognitive cost, is an endogenous object required to be consistent with the equilibrium price functional—this is, again, where "RI meets REE." And second, because of a law of large numbers applied to idiosyncratic realizations of ω, all aggregate quantities and prices are functions of θ in equilibrium—but now θ may contain a multitude of fundamentals as well as sunspots.

Note also that Definition 1, in the spirit of the standard Arrow-Debreu framework, makes no material distinction between states of nature and "true," non-contingent goods. The importance of this distinction for attention choices (and, by extension, consumption and production choices) affects equilibrium aggregation only through the aggregate (θ-contingent) demand and supply functions. As one example, if all agents optimally chose to ignore the distinction between state realizations θ and θ', this would manifest at the aggregate level as $\bar{x}^j(\theta) = \bar{x}^j(\theta')$, for all types j, and $\bar{y}(\theta) = \bar{y}(\theta')$. In our proof of our main results (including Theorem 1), leveraging this "aggregate" representation of the economy, instead of the underlying disaggregate signal and action choices, will simplify our analysis.

We next turn to defining feasibility. As in the example of Section 2, our notion of feasibility allows for the replacement of the price variable p in the agents' cognition state with a message m. We require that this message, like the equilibrium price, be representable as a function $M : \Theta \to \mathbb{R}^N_+$. The consumer and firm problems are adjusted to redefine the cognition state as $z = (\theta, m)$, which has a prior $\pi \in \mathscr{P}$ which must now be consistent with the message function M.

Definition 2. A feasible arrangement is a profile of consumption and production strategies, $\left((x^j)_{j=1}^J, y\right)$, attention choices, $\left((\phi^j)_{j=1}^J, \phi^F\right)$, a message rule $M : \Theta \to \mathbb{R}^N_+$, and a prior $\pi \in \mathscr{P}$ such that:

1. Consumption and production are informationally feasible, that is $x^j : \Omega \to \mathscr{X}$ for all j and $y : \Omega \to \mathscr{Y}$.

2. Consumption and production are technologically feasible, that is equation 13 holds along with

$$H\left(y(\omega), C^F[\phi^F, \pi], \theta\right) \leq 0, \forall\,(\omega, \theta) : \phi^F(\omega \mid f_\pi(\theta)) > 0 \tag{14}$$

3. The prior π about z is consistent with the message rule, that is $f_\pi(\theta) = (\theta, M(\theta))$ for all $\theta \in \Theta$.

We define efficiency as Pareto optimality under the aforementioned notion of feasibility:

Definition 3. An equilibrium is efficient if there does not exist a feasible alternative such that (i) all agents are weakly better off and (ii) a positive measure of agents is strictly better off.

As discussed in Section 2.2, replacing prices with messages allows our question about the "economy of knowledge" to be well-posed. Such messages are herein restricted to live in the same space as prices. This simplifies the exposition by making sure that z itself remains in the same space as we move back and forth between equilibria and planning alternatives. In Section 7.2, however, we discuss how to accommodate two related extensions: to enrich the messages sent by a planner, and to let endogenous objects other than prices, such as aggregate trades or taxes, enter the cognitive process as distinct elements of z. Finally, note that our efficiency notion (Definition 3) fully internalizes attention costs, similarly to how its conventional counterpart fully internalizes agents' preferences. As mentioned in the Introduction, this is a crucial difference between our approach and the alternative taken in Gabaix (2014, 2016) and Farhi and Gabaix (2020).

3.5 Invariance of Attention Costs

We now define the invariance assumption that will underpin our version of the First Welfare Theorem (as well as its converse). Toward that goal, we first define "transformations of information structures" in terms of functions that alter the cognition state z and map back to the same domain:

Definition 4 (Transformations of Information Structures)**.** Consider two information structures (π, ϕ) and $(\tilde{\pi}, \tilde{\phi})$ and a function $g : (\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)$. We say that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under g if

$$\tilde{\pi}(z) = \sum_{z' \in Z_\pi} \pi(z') 1_z(g(z')) \qquad \forall z \in Z_{\tilde{\pi}} \tag{15}$$

$$\tilde{\phi}(\omega|z) = \frac{\sum_{z' \in Z_\pi} \phi(\omega|z')\pi(z') 1_z(g(z'))}{\tilde{\pi}(z)} \qquad \forall \omega \in \Omega, z \in Z_{\tilde{\pi}} \tag{16}$$

where 1_z is the indicator function that its argument equals z.

Such transformations amount to a change of variables in the following sense. Consider a random variable z and replace it with the random variable $\tilde{z} = g(z)$, for some $g \in \mathcal{G}$. If the former's distribution is $\pi \in \mathcal{P}$, then the latter's is $\tilde{\pi} \in \mathcal{P}$ constructed as in (15). Furthermore, if the agent had chosen a signal ϕ for the original random variable and wishes to preserve the informational content of that original signal with respect to the new random variable $\tilde{z}$, then the new signal $\tilde{\phi}$ would be constructed according to (16).

We next define a *sufficiency* relationship between information structures, recasting the familiar definition of a *sufficient statistic* in our language:

Definition 5 (Sufficiency)**.** Consider two information structures (ϕ, π) and $(\tilde{\phi}, \tilde{\pi})$ such that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under some $g : (\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)$. We say that $\tilde{\pi}$ is sufficient for π with respect to ϕ if $\phi(\omega \mid z) = \tilde{\phi}(\omega \mid g(z))$ for all ω and all z such that $\pi(z) > 0$.

This definition is equivalent to saying that the distribution of z conditional on $\tilde{z}$ does not depend on ω. This is trivial if g is bijective, or merely "relabels" states, and a meaningful restriction when g is surjective, or both "relabels and combines" states. In particular, if the original signal structure allowed one to learn the relative likelihood of two states $z \neq z'$ such that $g(z) = g(z')$, then the sufficiency property does not hold.

We now define invariance with respect to specific classes of transformations:

Definition 6 (Invariance)**.** Fix a set $G \subseteq \{g : (\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)\}$. Consider any function $g \in G$ and any two information structures (ϕ, π) and $(\tilde{\phi}, \tilde{\pi})$ such that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under g. A cost functional C is invariant with respect to G if $C[\phi, \pi] = C[\tilde{\phi}, \tilde{\pi}]$ whenever $\tilde{\pi}$ is sufficient for π with respect to ϕ.

Within the specified set of transformations, invariance requires that rescaling or relabeling the states does not affect attention costs, while merging particular states has no effect provided the signal did not originally distinguish between those states.

Relation to Literature. In Appendix B, we prove that the mutual information of ω with z is invariant by our definition (Lemma 3 and Corollary 6). We also connect our invariance notion to a related one from the economic literature on flexible information acquisition (Caplin et al., 2022) and the statistics literature on information geometry (Amari and Nagaoka, 2000; Amari, 2016). More specifically, we show that our notion of invariance is implied by a stronger property that requires both what we call invariance and what we call *monotonicity*, which we will introduce in Section 6.2 to show a supplemental result.

The tight link between mutual information and invariance is reinforced by various results in the literature establishing that, within different classes of studied cost functions, mutual information is implied by "invariance under compression" (Caplin et al., 2022), "weak compression invariance" (Bloedel and Zhong, 2021), and "informational separability" (Tian, 2021). Most notably, Caplin et al. (2022) show that mutual information is basically the same as invariance within the class of uniformly posterior separable costs.

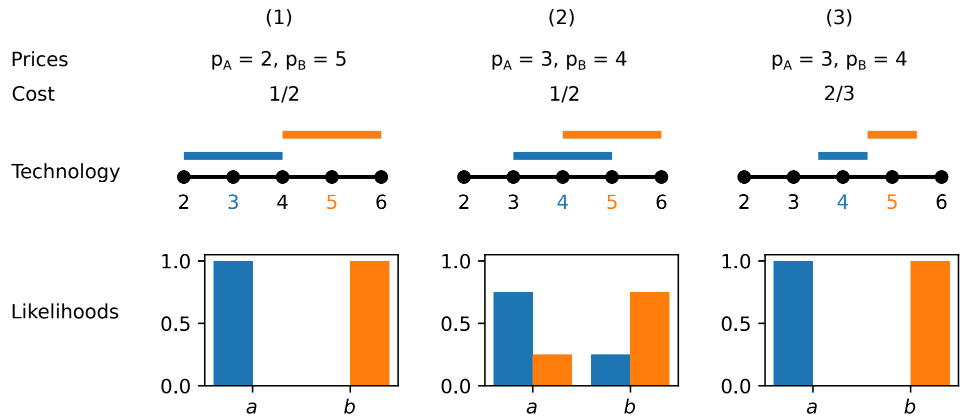
Outside the posterior separable context, the experimental costs of information defined by Denti et al. (2022), adapted to our setting to have representation $C[\phi, \pi] = C[\phi]$, are invariant because, by definition, they depend on the signal structure ("statistical experiment," in those authors' language) and not independently on prior beliefs.

By contrast, the neighborhood-based costs of Hébert and Woodford (2021), adapted to our setting with a fixed set of disjoint neighborhoods covering $\Theta \times \mathbb{R}_+^N$, are not invariant. Intuitively, when agents' costs are especially high for distinguishing prices in certain neighborhoods, then the costs of certain strategies will change if prices in different states are moved in and out of common neighborhoods. Similarly, the log-likelihood ratio cost of Pomatto et al. (2023) is invariant only under the trivial case in which the parameters controlling the distinguishability of pairs of states are identically equal to a constant.

We refer the reader to Hébert and La'O (2020) for a complementary, and more detailed, discussion of how the invariance condition in our paper and theirs alike connects to the literature.

An Example. Before proceeding to the main result, we provide a simple and concrete example of what invariance entails in a consumer decision problem. A consumer lives in a world with two states of nature with

Figure 1: Visualization of a Non-Invariant Cost Functional



Notes: Blue corresponds with state A and orange with state B. The first row of diagrams visualizes the agent's signal "technology" and the second plots, in a two-color bar graph, their likelihood distributions.

equal prior probabilities: $\theta \in \Theta = \{A, B\}$, with $\pi_\Theta(A) = \pi_\Theta(B) = \frac{1}{2}$. There is a single price, p, taking values in $\mathbb{R}_+$. The cognition state is $z = (\theta, p) \in \{A, B\} \times \mathbb{R}_+$. There are two possible price functionals, denoted by P and $\tilde{P}$. The first sets $P(A) = 3$ and $P(B) = 5$, and second sets $\tilde{P}(A) = 4$ and $\tilde{P}(B) = 5$. These mappings correspond to two possible distributions for z, denoted by π and $\tilde{\pi}$, with respective supports $Z_\pi = \{(A,3),(B,5)\}$ and $Z_{\tilde{\pi}} = \{(A,4),(B,5)\}$. Agents can construct signals of these variables, which lie in the space is $\Omega = \{a, b\}$.

We now construct a cost that does not satisfy invariance. Imagine that the agent has no way to directly learn θ; perhaps they do not even understand what "A" and "B" mean. But they can try to learn p. In particular, they can choose among a continuum of "observations," indexed by $\eta \geq 0$. Making an observation of type η costs $1/(\eta+1)$ and returns an outcome that is distributed uniformly on the interval $[p-\eta, p+\eta]$. The agent then applies the following algorithm to map observations to signals. If the outcome is uniquely consistent with the price in A (respectively, B), the agent receives signal $\omega = a$ (respectively, $\omega = b$) with probability 1. Otherwise, the agent receives either signal with probability $\frac{1}{2}$. This procedure defines a joint distribution ϕ over (ω, θ, p), and we define $C[\phi, \pi]$ as the lowest cost of observation generating a given ϕ.

Why does this not satisfy invariance? Imagine that, when the price mapping is P, or $(p_A, p_B) = (3,5)$, the consumer picks $\eta = 1$ at cost $C = 1/2$ and perfectly distinguishes the two states at the lowest possible cost. This is demonstrated in Column 1 of Figure 1, which visualizes the signal technology and plots the likelihood distributions. Next, see that when the price mapping is $\tilde{P}$, or $(p_A, p_B) = (3,4)$, the same experiment of $\eta = 1$ yields a different and strictly less informative signal (Column 2). Finally, see that the cheapest method of preserving the informational content of the original signal under the new price functional requires a lower η and therefore incurs a strictly higher cost (Column 3). The cost functional is not invariant, because the costs of creating the same "information" (i.e., perfectly distinguishing the states) changed when we applied a transformation mapping P to $\tilde{P}$.[13]

[13]The fact that the cost increased, rather than decreased, is irrelevant for invariance. But it is a violation of the *monotonicity* property that we introduce in Section 6.2.

On the other hand, any cost functional which depended only on the signal distributions visualized in the second row of Figure 1, without reference to the "names" or values of the states, would take an equal value in columns (1) and (3) and therefore be invariant with respect to this transformation. The mutual information of ω and (θ, p) is one such cost functional.[14]

4 (In)Efficiency of Inattentive Economies

In this section, we first state and prove our version of the First Welfare Theorem. This in turn leads to our second main result, which illustrates how a failure of invariance opens the door to inefficiency via cognitive externalities.

4.1 First Welfare Theorem1

Let G^p be the subset of functions $(\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)$ that transform the price component of the cognition state, i.e., they can be written as $g(\theta, p) = (\theta, h(p))$ for some $h : \mathbb{R}_+^N \to \mathbb{R}_+^N$. We now state our first main result:

Theorem 1 (First Welfare Theorem)**.** *If all agents have attention costs that are invariant with respect to G^p, then all equilibria are efficient.*

Below, we show how Theorem 1 builds on two intermediate results, which together make clear both the logic behind this theorem and the separate roles played by complete markets and invariance.[15]

Preliminaries: An Equivalent Economy. To set up our argument, we first define a notion of preferences and technology at the level of type-specific aggregate quantities, $\overline{x}^j(\theta)$ and $\overline{y}(\theta)$. This representation subsumes the attention choice to the preferences and technology of a "twin economy," and clarifies what is and is not standard about our problem.

Let $\overline{\boldsymbol{x}} \in \mathscr{X}^{|\Theta|}$ be a shorthand for the vectorization of state-specific consumption, $(\overline{x}(\theta))_{\theta\in\Theta}$. We define the following value function for each consumer type j that optimizes over consumption and signal structures subject the constraint that the average consumption across realizations of signals should not exceed an available basket $\overline{\boldsymbol{x}}$:

$$
\begin{aligned}
\overline{u}^j(\overline{\boldsymbol{x}}, \pi) \equiv \max_{x,\phi} \sum_{\omega,\theta} u^j(x(\omega), \theta)\, \phi(\omega \mid f_\pi(\theta))\, \pi_\Theta(\theta) - C^j[\phi, \pi] \\
\text{s.t. } \sum_{\omega} x(\omega)\, \phi(\omega \mid f_\pi(\theta)) \le \overline{x}(\theta),\ \forall \theta \in \Theta \qquad (17) \\
x : \Omega \to \mathscr{X}; \quad \phi(\cdot \mid f_\pi(\theta)) \in \Delta(\Omega),\ \forall \theta \in \Theta
\end{aligned}
$$

Under the maintained assumptions that (i) Ω is finite, (ii) $\mathscr{X}$ is compact, and (iii) $u^j(\cdot, \theta)$ and $C^j[\cdot, \pi]$ are continuous, Weierstrauss' theorem guarantees that this program has a solution and therefore $\overline{u}^j : \mathscr{X}^{|\Theta|} \times \mathscr{P} \to$

[14]The mutual information in each case is 0.69, 0.13, and 0.69.

[15]The statement and proof of the standard First Welfare Theorem requires neither the existence of an equilibrium nor its comparison to the solution of a planning problem. Instead, it presumes the existence of an equilibrium and proceeds to rule out Pareto improvements. The same applies to Theorem 1 here and its proof below. The study of equilibrium existence and Pareto optima is therefore *not* needed for our main result and is postponed to Section 6.1.

$\mathbb{R}$ is well-defined. We refer to $\overline{u}^j(\overline{\boldsymbol{x}},\pi)$ as a *reduced preference* for the basket of state-contingent commodities $\overline{\boldsymbol{x}}$, which depends on the prior π as an auxiliary parameter.

The consumer program (11) can now be rewritten as a more standard consumer optimization with the altered preferences defined above:

$$\begin{aligned} &\max_{\overline{\boldsymbol{x}}} \ \overline{u}^j(\overline{\boldsymbol{x}},\pi) \\ &\text{s.t. } \sum_{\theta}(P(\theta)\cdot\overline{x}(\theta) - P(\theta)\cdot e^j(\theta) - a^j\Pi(\theta))\,\pi_{\Theta}(\theta) \le 0 \end{aligned} \tag{18}$$

Two specific features of this representation are important. First reduced preferences inherit monotonicity in the goods space from the primitive utility function u^j.[16] This is formalized as Lemma 2 stated and proven in the Appendix. Second, prices affect preferences directly via the prior π (and, more fundamentally, their influence on the information-acquisition problem in Program 17). This feature embodies the cognitive externality on the consumer side of the economy

For firms, we analogously let $\overline{\boldsymbol{y}} = (\overline{y}(\theta))_{\theta\in\Theta} \in \mathscr{Y}^{|\Theta|}$ be the vectorization of state-specific consumption and define, for each prior π, the following *reduced production set*:

$$\begin{aligned} \overline{\boldsymbol{F}}(\pi) \equiv \Big\{\overline{\boldsymbol{y}} \in \mathscr{Y}^{|\Theta|} :\ \exists\,(y,\phi) \text{ s.t.: } \quad & \sum_{\omega} y(\omega)\,\phi(\omega \mid f_\pi(\theta)) \le \overline{y}(\theta), \forall \theta\in\Theta \\ & H\big(y(\omega), C^F[\phi,\pi],\theta\big) \le 0, \forall\,(\omega,\theta) : \phi(\omega\mid f_\pi(\theta)) > 0 \\ & y:\Omega\to\mathscr{Y}; \quad \phi(\cdot\mid f_\pi(\theta))\in\Delta(\Omega),\ \forall\,\theta\in\Theta\Big\} \end{aligned} \tag{19}$$

This set describes the combinations, across goods and states of nature, of the aggregate production levels that are both technologically and cognitively feasible: that is, they can be disaggregated with at least one attention strategy $\phi = (\phi(\cdot\mid z))_{z\in Z_\pi}$ and a production plan $y:\Omega\to\mathscr{Y}$. The representation further extends the metaphor from Section 3.3 about attention as an "in-house" productive activity.

The firms' profit maximization problem then reduces to the following program:

$$\begin{aligned} &\max_{\overline{\boldsymbol{y}}} \ \sum (P(\theta)\cdot\overline{y}(\theta))\,\pi(\theta) \\ &\text{s.t. } (\overline{y}(\theta))_{\theta\in\Theta} \in \overline{\boldsymbol{F}}(\pi) \end{aligned} \tag{20}$$

This is like the profit-maximization problem of a standard, attentive firm, except that the production set is allowed to depended on the prior. This embodies the cognitive externality on the firm side of the economy.

It is trivial to show that any equilibrium of the original economy corresponds to an equilibrium of the reduced economy, in terms of implying the same aggregate outcomes and payoffs. Moreover, if there exists a Pareto dominating feasible allocation in the original economy, there must also be one in the twin economy. Thus, it is sufficient to prove the First Welfare Theorem in the twin economy.

[16] The use of such monotonicity in the proof of Proposition 5 could be relaxed for an appropriate form of local non-satiation, as in our earlier draft (Angeletos and Sastry, 2019).

Step One: Efficiency Ignoring Messages. We now state a restricted version of the First Welfare Theorem that removes the planner's flexibility to replace prices with other messages, but also holds without any additional assumption on attention costs:

Proposition 5. *For any equilibrium with price functional $p = P(\theta)$, there does not exist a Pareto dominating allocation that is feasible with message $m = P(\theta)$.*

By restricting the messages to replicate the equilibrium prices, we fix π in the reduced preferences and production sets. The rest of the proof then reads much like the familiar, textbook argument for why competitive equilibria can not be Pareto dominated (e.g., section 16.C of Mas-Colell et al., 1995). We provide the proof in Appendix A.1.

It is worth emphasizing that this intermediate result hinges on our expanded notion of complete markets. In particular, although the step from our twin-economy representation to Proposition 5 depends solely on complete markets over θ and on the aforementioned restriction on messages, the assumption of complete markets over ω is crucial for obtaining our twin-economy representation in the first place. We revisit this point in Section 7.1. But as already mentioned, the key rationale for this assumption is that it allows us to accomplish two goals at once: to illustrate that rational inattention *per se* need not cause inefficiency and to draw attention to the question of whether the price system is the best way to economize attention costs. Proposition 5 sets the stage by abstracting from this question; the rest of the paper zeroes in on it.

Step Two: Efficiency With Messages. The translation from Proposition 5 to Theorem 1 uses informational invariance. We first establish the following result: if attention costs are invariant to transformations of the price component of the economic state, then preferences are production sets do not depend directly on the stochastic properties of prices. By the same token, the cognitive externality, or the dependence of preferences and technologies on the *endogenous* part of the prior π, is muted.

Lemma 1. *If consumers' attention costs are invariant with respect to G^p, there exist functions $\hat{u}^j : \mathscr{X}^{|\Theta|} \times \Delta(\Theta) \to \mathbb{R}$ such that the reduced preferences satisfy $\overline{u}^j(\overline{\boldsymbol{x}}, \pi) = \hat{u}^j(\overline{\boldsymbol{x}}, \pi_\Theta)$ for all $\overline{\boldsymbol{x}} \in \mathscr{X}^{|\Theta|}$ and all $\pi \in \mathscr{P}$. Similarly, when firms' attention costs are invariant with respect to G^p, there exists a correspondence $\hat{\boldsymbol{F}} : \Delta(\Theta) \rightrightarrows \mathscr{Y}^\Theta$ such that the reduced production set satisfies $\overline{\boldsymbol{F}}(\pi) = \hat{\boldsymbol{F}}(\pi_\Theta)$ for all $\pi \in \mathscr{P}$.*

To prove this Lemma, we observe that each agent's cognitive cost, C^x for $x \in \{1, \ldots, J\}$, can be written independent of the price functional. Select an arbitrary element $\overline{p} \in \mathbb{R}^N_+$ and define $g_{\overline{p}} = \{(\theta, p) \mapsto (\theta, \overline{p})\}$ as the transformation that "eliminates" the variation of prices from the cognition state. For any π, let $\pi_{\overline{p}}$ be the transformed prior under $g_{\overline{p}}$. Since $g_{\overline{p}} \in G^p$, invariance within G^p guarantees that $C^x[\phi, \pi] = C^x[\phi, \pi_{\overline{p}}]$ for any ϕ and π. It follows that we can uniquely define a new cost functional $\hat{C}^x : \Phi \times \Delta(\Theta)$ by taking $\hat{C}^x[\phi, \pi_\Theta] = C^x[\phi, \pi_{\overline{p}}]$ for each ϕ and any arbitrary choice of $\overline{p}$. As one example of how this Lemma applies to an invariant cost functional, we can observe that the mutual information between the signal ω and the cognition state $z = (\theta, p)$ ("C") can be rewritten as the mutual information between the signal ω and only the fundamental state θ ("$\hat{C}$"). The definitions of reduced preferences and production sets can then be applied with these new cost functionals to construct representations $\hat{u}^j$ and $\hat{\boldsymbol{F}}$.

This result has an even stronger implication in the context of general equilibrium. If all agents' cost functionals satisfy invariance with respect to G^p, as assumed in Theorem 1, then there exists an equivalent twin economy (i.e., with the same equilibria, feasible outcomes, and Pareto ranking) in which cost functionals are written only in reference to the fundamental state and therefore the cognitive externality, which as established earlier manifested as dependence of preferences and production sets on the endogenous component of the prior π, is absent by assumption.

The proof of Theorem 1 is then completed by using this Lemma to establish that there cannot exist a Pareto dominating allocation even with the use of *arbitrary* messages. Imagine there were. Lemma 1 implies that each consumer's payoff as well as the feasibility constraints would be identical under the message $m = P(\theta)$. Therefore, the allocation must also be implementable with that message; but Proposition 5 implies that such an allocation cannot exist. This concludes the proof of Theorem 1.

Finally, before proceeding, we note an implication of Theorem 1 that is especially relevant for the existing literature on equilibrium economies with rational inattention. A trivial case in which attention costs are invariant with respect to transformations of the price is the case when costs depend only on the joint distribution of signals and the exogenous state θ. We formalize below that these economies are efficient:

Corollary 1. *Define a* state-tracking economy *as one in which all agents' attention costs can be written as* $C(\phi,\pi) = \hat{C}(\phi,\pi_\Theta)$, *for all* $\phi \in \Phi$ *and* $\pi \in \mathscr{P}$, *and some* $\hat{C} : \Phi \times \Delta(\Theta) \to \mathbb{R}$. *In any state-tracking economy, all equilibria are efficient.*

The *state-tracking* simplification is common in some macroeconomic applications (e.g., Angeletos and La'O, 2020; Colombo et al., 2014; Maćkowiak and Wiederholt, 2015; Tirole, 2015; Gul et al., 2017), although it is inconsistent with our motivation, the decision-theoretic literature we relate to, and a large behavioral literature, all of which emphasize cognition about prices or other variables that may be exogenous to the individual decision maker but certainly not to the economy as a whole. Thus any results about efficiency or inefficiency in such prior works have abstracted from the kind of cognitive externality we have emphasized here and therefore did not hinge on informational invariance in any form.

4.2 Non-Invariant Costs, Cognitive Externalities, and Inefficiency

By emphasizing how rational inattention opens the door to cognitive externalities, and by using invariance to mute these externalities, our analysis suggests that invariance is not only sufficient for efficiency but also necessary. Theorem 1 established only sufficiency. We now argue that necessity is also true, subject to some qualifications: we must rule out cases where invariance fails off equilibrium but not on equilibrium, or where cognitive externalities cancel out across agents by mere coincidence.

Accordingly, we prove two partial converse results to Theorem 1. The first says that efficient economies satisfying the conditions of Theorem 1 can be made inefficient by adding "small" deviations from invariance. The second says that economies in which all agents' cost functionals fail invariance in a global and non-trivial sense are inefficient.

Small Changes to the Cost Functional. Consider an efficient economy that satisfies Theorem 1 and has at least one equilibrium. Let $P^*: \Theta \to \mathbb{R}_+$ be the price functional in one of these equilibria and $\boldsymbol{p}^* = (P^*(\theta))_{\theta\in\Theta}$ be its vectorization. Consider now a variant "ϵ-invariant" economy, indexed by $\epsilon > 0$ and P^*, in which the cost functional of each agent $i \in \{1,\ldots,J\} \cup \{F\}$ is

$$C^i_{\epsilon,\boldsymbol{p}^*}[\phi,\pi] = C^i[\phi,\pi] - \epsilon||\boldsymbol{p}(\pi) - \boldsymbol{p}^*|| \tag{21}$$

where $\boldsymbol{p}(\pi)$ is the vectorization of the price functional encoded in the prior π and $||\cdot||$ denotes the Euclidean norm. Observe that all agents' cost functionals are now non-invariant: changing the price clearly affects the second term, while the first is invariant by assumption.

We now prove that, fixing any $\epsilon > 0$ and performing this construction for any of the equilibria, the new economy is inefficient. This follows directly from the fact that the planner can simply send any message $M \neq P^*$ along with the original, equilibrium allocation to obtain a Pareto improvement.

Proposition 6. *For any efficient economy satisfying the conditions of Theorem 1 and in which at least one equilibrium exists, there is an inefficient ϵ-invariant economy.*

In this sense, the efficient economies covered by Theorem 1 are arbitrarily "close" in the space of cost functionals to invariant, inefficient economies.

Inefficiency With Global Non-Invariance. We now develop a second result which shows how specific non-invariant costs can lead to inefficiency. We first build intuition with an illustrative calculation. Consider an endowment economy (for simplicity) with a competitive equilibrium described by its twin-economy consumption vectors $(\overline{\boldsymbol{x}}^{j*})_{j=1}^J$ and price functional P^*, with vectorization $\boldsymbol{p}^*$. Assume further, for illustration, that the reduced payoffs $\overline{u}^j(\overline{\boldsymbol{x}}, \boldsymbol{p}, \pi_\Theta)$, written in terms of the vectorized consumption and prices, are differentiable in the first two arguments. Under these assumptions, the competitive equilibrium is efficient if and only if, for some non-negative Pareto weights χ^j, the allocations and prices solve

$$\begin{aligned} \max_{\overline{\boldsymbol{x}}^j \in \mathcal{X}^{|\Theta|}, \boldsymbol{p}\in\mathbb{R}_+^{N|\Theta|}} \quad & \sum_{j=1}^J \chi^j \overline{u}^j(\overline{\boldsymbol{x}}, \boldsymbol{p}, \pi_\Theta) \\ \text{s.t.} \quad & \sum_{j=1}^J \mu^j \overline{\boldsymbol{x}}^{\boldsymbol{j}} \leq \sum_{j=1}^J \mu^j \boldsymbol{e}^{\boldsymbol{j}} \end{aligned} \tag{22}$$

where the $\boldsymbol{e}^{\boldsymbol{j}}$ are the vectorized endowments. In this problem, the planner can be thought to choose both allocations and "prices," without any constraints related to market implementability. The prices are therefore "messages," as we have interpreted throughout. The first-order condition with respect to $\boldsymbol{p}$ is

$$\sum_{j=1}^J \chi^j \, \nabla_{\boldsymbol{p}} \overline{u}^j(\overline{\boldsymbol{x}}^j, \boldsymbol{p}, \pi_\Theta) = 0 \tag{23}$$

where $\nabla_{\boldsymbol{p}}$ denotes the gradient over $\boldsymbol{p}$. This condition, which is necessary for efficiency, asserts that there are no first-order effects of changing prices. When all agents' costs are invariant, each term of the sum in equation 23 is zero. When at least one agent's cost is non-invariant, at least one term may be nonzero at *some* $(\overline{\boldsymbol{x}}^j)_{j=1}^J$ and $\boldsymbol{p}$. What we do not generally know is this happens at the *equilibrium* point and whether the entire sum is nonzero (i.e., whether cognitive externalities do not accidentally cancel out across agents).

Resolving these questions requires more structure on cost functionals and the precise ways in which invariance fails. Toward stating one formal converse, we introduce the following definitions.

Definition 7. Cost functional C is globally non-invariant if, for every (ϕ, π) such that ϕ embodies the collection of some information and π represents prices that differ across states,[17] there exists some $g \in G^p$ and associated transformation $(\tilde{\phi}, \tilde{\pi})$ such that $C[\phi, \pi] > C[\tilde{\phi}, \tilde{\pi}]$.

Definition 8. An economy is a C-economy if $C^x = C$ for all $x \in \{1, \dots, J\} \cup \{F\}$.

The first definition describes cost functionals that always display meaningful deviations from invariance, except in null cases with no information collection or non-stochastic prices. This property is satisfied by the graphical example in Section 3.5—moving the prices closer together always economized on costs under any information acquisition strategy. It is also satisfied by cost functionals that continuously specify the role of perceptual distance or difficulty in distinguishing nearby states, like a version of the log likelihood ratio cost in Pomatto et al. (2023) with a fixed mapping between state values and the parameters controlling the distinguishability of states.[18] In these cases, there may always exists some transformation (e.g., reducing the distance between different states' spot prices in $\mathbb{R}_+^N$) that economizes on cognitive costs. The second definition requires that all agents have the same cost functional.[19] This may be a natural assumption in applications. Formally, it ensures that all agents' payoffs are increased by perturbing prices in a specific way.

We can now state the result:

Proposition 7. *Let C be a globally non-invariant cost functional. There does not exist an efficient C-economy in which (i) some consumers collect information and (ii) prices vary across states.*

The proof of this result, spelled out in Appendix A.2, relies on the fact that the social planner can always improve upon equilibrium allocations by sending a message that benefits consumers by making information acquisition easier. The result does not show that an economy in which agents have non-invariant costs is *necessarily* inefficient, which is false.[20] Nonetheless, Proposition 7 may offer a useful perspective on the scope for inefficiency when a researcher conjectures that a specific, globally non-invariant cost functional best characterizes macroeconomic decisions. We expand on this interpretation in the next section.

[17] That is, there does not exist any element $\omega \in \Omega$ such that $\phi(\omega \mid \theta) = 1$ for all $\theta \in \Theta$; and there does not exist a vector $p \in \mathbb{R}_+^N$ such that $\pi(\theta, p) = \pi_\Theta(\theta)$ for all $\theta \in \Theta$.

[18] In Pomatto et al. (2023), the "distinguishability" parameters are the β_{ij} corresponding to pairs of states $i, j \in \Theta$. If $\beta_{ij} \propto (1-j)^{-2}$, as derived in that paper's Proposition 2 under additional restrictions, the LLR cost is globally non-invariant.

[19] Note that definition of the utility functions and production possibilities set can accommodate different "scales" for these costs for different agent types.

[20] A trivial example is one in which attention costs happen to be zero, for any signal structure ϕ, when $z = (\theta, P^*(\theta))$ and P^* is an equilibrium price function in the underlying *attentive* economy, and the costs are positive, constant, and sufficiently large whenever $z = (\theta, g(\theta))$ for any $g \neq P^*$. This economy is inefficient, because there must be at least one equilibrium in which $P \neq P^*$ (e.g., $P = 2P^*$). This cost functional is not uniformly non-invariant, because at the prior associated with $P^*(\theta)$, there is no transformation that reduces cognitive costs—only transformations that increase them.

5 Interpretation and Discussion

We now circle back to the two motivating questions at the beginning of the Introduction.

5.1 Non-Paternalistic Behavioral Economics

The first question raised in the Introduction was whether inattention *per se* causes a failure of the First Welfare Theorem. Our first main result, Theorem 1, gave a negative answer under three key assumptions: that our invariance condition holds; that markets are complete in the properly expanded sense; and that inattention is rational, i.e., the product of optimizing behavior subject to cognitive limitations.

Needless to say, none of these assumptions have to hold in practice. But realism is not the point of either the standard First Welfare Theorem or our variant. Rather, by spelling out the conditions that suffice for efficiency, the goal is to give guidance on the precise assumptions that could justify government intervention.

Seen from this perspective, a key contribution of our paper is to clarify that rational inattention does *not* share the normative implications of the "behavioral" alternatives considered in Gabaix (2014, 2017) and (Farhi and Gabaix, 2020). The reason is simple but important: by recasting what looks like an "optimizing friction" as the product of rational behavior subject to attention costs, and by including these costs in welfare calculations, one finds that the invisible hand continues to do its magic despite people's inattention. More succinctly, what drives the normative conclusions of Gabaix (2014, 2017) and Farhi and Gabaix (2020) is not that people are inattentive but rather that their planner is paternalistic: she does not internalize agents' attention costs and she treats their inattention as an irrational mistake that ought to be corrected.

At the same time, our analysis shows how exactly inattention can open the door to inefficiency *without* such a paternalistic stance. One possibility, which we have abstracted from but return to in Section 7.1, is that inattention introduces uninsurable idiosyncratic risk. The other possibility, which we choose to emphasize in this paper, is that inattention introduces cognitive externalities. We next discuss how this particular channel of inefficiency speaks to Hayek's (1945) argument, as well as how it relates to recent decision-theoretic and experimental advances.

5.2 Revisiting the Economy of Knowledge

Consider now our formalization of Hayek (1945): does the price system optimally economize attention costs relative to other means of communication and coordination? Theorem 1 says "yes," under a specific assumption about the model of inattention, namely our invariance condition.

But this lesson comes with the following important caveat: under the invariance condition used to guarantee efficiency, a social planner could implement the *same* outcomes by merely announcing the state of nature and a completely degenerate message:

Corollary 2. *Under the conditions of Theorem 1, any equilibrium can be replicated with a feasible mechanism that uses the same consumption, production, and attention strategies but replaces prices with an "uninformative" message rule, namely a rule M such that $M(\theta) = \overline{m}$ for all θ and for arbitrary $\overline{m} \in \mathbb{R}_+^N$.*

This result is immediate from the last step of our proof of Theorem 1. In the context of Hayek's economy of knowledge, the result says that the same conditions that guarantee the efficiency of markets also imply that there is no welfare loss from scrapping market signals altogether and, instead, having agents redirect all their attention to learning the underlying state of nature alone. The "economization of knowledge" in Theorem 1 is weak. It does not rely on prices' coarsely representing the state of nature, or repackaging the state of nature in a cognitively friendly manner, but instead on the agents' ability to costlessly generate an equivalent transformation of the state of nature in their minds.

It may be useful to return to the illustration of Section 3.5. Imagine that the state $\theta \in \{A, B\}$ represented the primitive demand shifter for another agent in the economy. Our example cost was motivated by the idea that learning about θ was prohibitively difficult, while learning about prices was feasible but scale-dependent. This mapped to a failure of our invariance condition. Mutual information costs, on the other hand, implied that there was no difference between learning "directly" about this taste (e.g., via direct research) versus learning about prices. These costs were, in light of Theorem 1, compatible with efficiency.

More generally, this presents the following paradox: the idea that markets *strictly* economize knowledge seems most meaningful in cases that open the door to inefficiency. Away from invariance, the invisible hand may naturally do better than the centralized mechanism with degenerate messages from Corollary 2. But an optimal mechanism might do even better.

Policies for a Better Economy of Knowledge. How exactly can a planner improve welfare in economies in which failures of invariance lead to inefficiency, and how does this relate to Hayek's logic about the price system? While our notion of feasibility for the social planner allows for manipulation of both allocations and prices (messages), our argument centered around cognitive externalities suggested that only manipulation of prices was necessary to achieve welfare improvements. This was hinted at in the proof of Proposition 7, which showed that, in globally non-invariant economies, the social planner could achieve a Pareto improvement merely by announcing a message that differed from the equilibrium price. Provided that the planner could induce the "correct" prices via taxes and transfers, such intervention would be sufficient to achieve a Pareto improvement.

This argument can be made more concrete in the example of Section 2. The proof of Proposition 1 showed that a small tax would only have a second-order welfare loss in terms of distorting consumption and attention choices, *regardless* of the cost functional. But if we were in a case in which invariance failed for the cost functional (e.g., the examples of Section 2.4), the implied change in the stochastic properties of prices could have a first-order welfare gain via attention costs.

What would such welfare-improving taxation look like in the example? Assume that the planner can impose a flat, non-contingent tax τ on coconut expenditure. The demand x of a consumer with signal ω is given by $x = 1 - (1+\tau)\mathbb{E}[p|\omega]$ and, as a result, the equilibrium price satisfies $p = 1 - \frac{1}{(1+\tau)\rho}\xi$, where ξ is the unknown endowment of coconuts and ρ is the squared correlation of the agent's signal with ξ. For a given ρ, a positive tax makes demands more elastic and, hence, prices less dispersed across states of nature. In the efficient, mutual information benchmark, Proposition 1 implies that the optimal tax is $\tau = 0$. In the "perceptual distance" economy, the arguments of Proposition 2 show that small positive taxes improve welfare because they increase price variance and thus make inference about correct choice easier. In the "variance penalty" economy, the arguments of Proposition 3 show that small subsidies improve welfare because they decrease price variance and make inference about correct choice easier.

5.3 Invariant Costs: Characterization and Experimental Evidence

Our analysis has equated the non-existence of cognitive externalities and the informational optimality of the market mechanism to the invariance assumption. We now discuss (i) how restrictive is invariance among the class of sensible information costs and (ii) the available evidence on it.

Characterization. As mentioned in Section 5.3, our notion of invariance is closely connected to related notions from the statistics literature on information geometry (Amari, 2016), which have recently been applied in information economics by Hébert and Woodford (2021), Hébert and La'O (2020), and Caplin et al. (2022). In Appendix B, we show, via arguments familiar from other decision-theoretic applications, that mutual information costs satisfy invariance. Thus:

Corollary 3. *If all agents' attention costs are given by (any transformation of) the mutual information between ω and z, then all equilibria are efficient.*

Our notion of invariance is closely related to a property of state-dependent stochastic choice data defined by Caplin et al. (2022), *invariance under compression*. Loosely speaking, this property requires that observed stochastic choice patterns be invariant to relabeling states of nature and/or merging those that correspond with identical payoffs. Theorem 3 of Caplin et al. (2022) establishes that mutual information is the only cost that is consistent with invariance under compression among uniformly posterior separable costs, a subset of the posterior separable class.[21] This confines attention to a narrower class than the one allowed in our analysis, but suggests that one can loosely think of the provided condition for efficiency as synonymous with mutual information costs. In this sense, we justify the following "if and only if" result:

Corollary 4. *Consider the class of uniformly posterior separable costs, as defined in Caplin et al. (2022). Within this class, Hayek's (1945) argument about the informational optimality of the price system, as formulated herein, is valid if and only if Sims's (2003) proposal for how to measure attention costs is also valid.*

As for the broader class of posterior separable costs, Theorem 2 of Caplin et al. (2022) says that consistency of stochastic choice data with invariance under compression holds if and only if the cost functional is invariant in the sense of Amari and Nagaoka (2000). In light of the previous discussion of how this notion in turn relates to ours, it seems a safe guess that Theorem 2 of Caplin et al. (2022) extends outside the posterior separable class with our expanded definition of invariance. We conclude that the combination of our results with those of Caplin et al. (2022) provides a pathway to test the key assumption on cost functionals that underpins efficiency, and Hayek's (1945) economy of knowledge, in the environment that we study.

[21] See Section 4.1 and Definition 3 in Caplin et al. (2022).

Experimental Evidence. The best available evidence on the failure of invariance comes from perceptual experiments. In these experiments, there is an objective "state of the world" (e.g., "51 of 100 balls are red"), participants in the laboratory observe some representation of that state (e.g., a picture of the 100 red or blue balls), and participants then make a decision whose payoff depends on the state (e.g., electing to receive a payment if more balls are red). Dean and Neligh (2022) design such an experiment, with the aforementioned set-up, to test numerous axioms of state-dependent stochastic choice including invariance under compression. Their experimental data reject invariance under compression, and the authors propose a variant of the Shannon mutual information model that can rationalize the results.

In an earlier perceptual experiment, Shaw and Shaw (1977) had subjects try to recall the identity of a symbol (a letter E, T, or V) which was briefly displayed, and varied the location of the letter in the display. In their data, subjects are more able to distinguish letters if those letters consistently appear in the same locations. Woodford (2012) interprets this as a rejection of the mutual information model's implication that the probabilities of a decision-irrelevant state, the location of the symbol, are irrelevant for decisions.

A number of recent theoretical works (e.g., Pomatto et al., 2023; Morris and Yang, 2022; Hébert and Woodford, 2021) have explored the formal underpinnings for cost functionals that embody various dependencies on the physical attributes of the state space. These latter two works, in particular, prioritize the idea that distinguishing "closer" states in some metric may be more difficult, like in the example of Section 2.4.

But with the exception of suggestive evidence on nominal illusion, there is comparatively less research on the dependence of consumer or firm decisions on scale or other stochastic properties of prices, which is the exact setting of interest. While complete invariance seems unlikely, there are plausible economic arguments for departures in both directions as illustrated in Section 2.4. Moreover, as noted by the existing literature, it is not obvious how externally valid perceptual experiments are for these economic settings. Our results thus provide a new, macroeconomic motivation for testing invariance in laboratory settings that even more precisely approximate consumer and firm choices in markets.

6 Additional Results

In this section, we show several additional results which address other equilibrium properties of inattentive economies: equilibrium existence, the implementability of Pareto optima, and the absence or presence of non-fundamental volatility in efficient equilibria.

6.1 Equilibrium Existence and the Second Welfare Theorem

Theorem 1 presumed existence of equilibria and was silent on when Pareto optima exist and can be implemented as equilibria with transfers.[22] In this subsection, we fill both gaps. Our main observation is that, in the twin-economy representation developed in Section 4.1, the key ingredients for proving equilibrium existence and, by extension, proving a Second Welfare Theorem for environments in which Theorem 1 holds are

[22] An equilibrium with transfers in our setting is a direct extension of Definition 1 in which each consumer has a state-dependent wealth $p(\theta)e^j(\theta) + T^j(\theta)$, where $T^j(\theta) \in \mathbb{R}$ is a transfer that nets out across agents, that is, $\sum_{j=1}^J T^j(\theta) = 0$ for all $\theta \in \Theta$.

convexity and continuity of preferences and convexity and closedness of production sets. We then have the following result, the proof of which formally states the exact required notions of convexity and continuity:[23]

Proposition 8 (Equilibrium Existence and Second Welfare Theorem)**.** *Assume that attention costs are invariant with respect to G^p; that reduced preferences are convex and continuous; and that reduced production sets are convex and closed. Then, an equilibrium exists and any Pareto optimum can be implemented as an equilibrium with transfers.*

By mapping the inattentive economy to a twin attentive economy we make clear that rational inattention does not pose any *new* difficulty for existence than that familiar from standard general equilibrium theory, namely, the possibility of discontinuity in aggregate excess demands due to non-convexities in preferences and technologies. In fact, one may conjecture that, starting from an attentive but non-convex economy, the introduction of rational inattention may *aid* equilibrium existence—and thereby the Second Welfare Theorem, too—by smoothing out the associated discontinuities in aggregate excess demands.

In Online Appendix D, we show how an inattentive economy with a continuum signal space and posterior separable (Caplin and Dean, 2015) attention costs can satisfy the required continuity and convexity assumptions. In a nutshell, we observe that posterior separable costs allow agents to replicate lotteries over allocations at weakly negative cost.

6.2 Fundamental Equilibria and Optimal Attention

Theorem 1 and its proof ignored the question of how inefficiency relates to the presence, or lack, of "non-fundamental volatility." Here, we show that these two issues are related but distinct—efficient equilibria can feature non-fundamental volatility, while a more stringent condition on cognitive costs can rule this out.

We first define notions of the *economy-wide fundamental* and a *group-specific fundamental* for each consumer type or firm. Heuristically, these concepts coarsen the state at the level at which all or some agents' preferences, endowments, and technologies are identical. Our specific definition describes one such coarsening, which applies under the maintained, simplifying assumption that Θ is totally ordered. We next use these concepts to define when equilibria are *fundamental* and *price-tracking*.

Definition 9. The *economy-wide fundamental* is a random variable $\theta^* \in \Theta$ that can be expressed as

$$\begin{aligned}\theta^* \equiv \min\Big\{ & t \in \Theta \text{ s.t. } : e^j(t) = e^j(\theta) \text{ and } u^j(x,t) = u^j(x,\theta),\ \forall j \in \{1,\dots,J\}, x \in \mathscr{X}; \\ & \text{and } H(y,c,t) = H(y,c,\theta),\ \forall y \in \mathscr{Y}, c \in \mathbb{R}_+ \Big\}\end{aligned} \tag{24}$$

The *group-specific fundamental* for consumers of type j is a random variable $\theta^j \in \Theta$ such that $\theta^j \equiv \min\{t \in \Theta : u^j(x,\theta) = u^j(x,\theta'),\ \forall x \in \mathscr{X}; e^j(\theta) = e^j(t)\}$, and the corresponding object for the firm is a random variable $\theta^F \in \Theta$ such that $\theta^F \equiv \min\{t \in \Theta : H(y,c,\theta) = H(y,c,t),\ \forall y \in \mathscr{Y}, c \in \mathbb{R}_+\}$.

[23] Naturally, invariance is as essential for our version of the Second Welfare Theorem as it is for our version of the First Welfare Theorem. It may be possible, though, to prove existence without invariance (and without efficiency) by appropriately adapting techniques from the general equilibrium literature on analysis with price-dependent preferences. See, for instance, Section III in Sonnenschein (2017) for an overview of these techniques.

Definition 10. An equilibrium is *fundamental* if $\overline{x}^j(\theta) = \overline{x}^{j*}(\theta^*)$ and $\overline{y}(\theta) = \overline{y}^*(\theta^*)$, for some functions $(\overline{x}^{j*})_{j=1}^J$ and $\overline{y}^*$. An equilibrium is *price-tracking* if (θ^j, p) is a sufficient statistic for (θ, p) with respect to (ϕ^j, π), for all j, and similarly (θ^F, p) is a sufficient statistic for (θ, p) with respect to to (ϕ^F, π).[24]

In a fundamental equilibrium, allocations can not depend on sunspots; and in a price-tracking equilibrium, agents only pay attention to the objects that enter their payoffs.[25]

We finally define a *monotonicity* property of cost functionals:

Definition 11. Fix a set $G \subseteq \{g : (\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)\}$. Consider any function $g \in G$ and any two information structures (ϕ, π) and $(\tilde{\phi}, \tilde{\pi})$ such that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under g. A cost functional C is monotone with respect to G if $C[\phi, \pi] > C[\tilde{\phi}, \tilde{\pi}]$ whenever $\tilde{\pi}$ is not sufficient for π with respect to ϕ.

Monotonicity, in words, says that any strict coarsening of the state leads to cognitive savings. Monotonicity is combined with invariance in the characterization of Appendix B, and satisfied by mutual information. Monotonicity was not satisfied by the graphically illustrated example in Section 3.5.

Our result is that full invariance plus monotonicity guarantees that every equilibrium is both fundamental and price-tracking:

Proposition 9. *Assume that all agents' attention costs are invariant and monotone with respect to the set of all possible transformations,* $\{g : (\Theta \times \mathbb{R}_+^N) \to (\Theta \times \mathbb{R}_+^N)\}$. *Then, all equilibria are efficient, fundamental, and price-tracking.*

To prove this result, we first show that monotonicity implies that any *efficient* allocation is fundamental. The intuition is that, given monotonicity, eliminating contingency on non-fundamental contingencies would economize on cognitive costs without harming feasibility. Along with the fact any equilibrium is efficient under invariance, this proves that any equilibrium is fundamental under both invariance and monotonicity. To show the price-tracking property, we finally show that contingency of information strategies on extraneous states would contradict individual, rather than social, optimality under monotonicity.

This argument invites the following intuition for why the invisible hand may *efficiently* produce non-fundamental volatility when attention costs are invariant but not monotone. In such circumstances, consumers can enjoy lower attention costs by making their signals, and thereby also their consumption choices, co-vary with sunspots. This willingness to pay for sunspots manifests in the twin economy's demand functions for state-contingent goods; prices then adjust to ensure an efficient allocation.

What if *both* invariance and monotonicity are violated? This is precisely the scenario captured by the third example of an inefficient economy in Section 2.4. In such circumstances, the invisible hand may fail to produce the efficient amount of non-fundamental volatility, and may even yield multiple, Pareto-ranked equilibria. This possibility represents a form of "cognitive trap" that, unlike the form articulated in Tirole (2015), does not depend on payoff (or pecuniary) externalities and originates instead from attention costs.

[24]Here, we are using the language from the remarks immediately after Definition 5. This can be translated into the original definition of sufficiency by applying the function $(\theta, p) \mapsto (\theta^j, p)$ for each agent type and describing the associated transformed prior.

[25]In fact, it is straightforward to strengthen Proposition 9 below to show that, when an agent enjoys utility from only a subset of the available goods, she only pays attention to the prices of this particular subset as opposed to the entire price vector.

7 Extensions and Open Questions

In this Section, we expand on the meaning of complete markets in our context and on an alternative assumption that lets rational inattention become the source of uninsurable idiosyncratic risk. We next sketch how one could extend the analysis in two other directions: richer specifications of the cognition state and of the planner's options; a more general class of behavioral frictions. We finally comment how our insights translate from Arrow-Debreu to games.

7.1 The Role of Complete Markets

Our analysis assumed the following "expanded" notion of complete markets: consumers had access to transfers contingent on not only the primitive state of nature θ, but also idiosyncratic signals ω. This assumption kept our arguments close to classical GE theory and allowed us to reach our version of the First Welfare Theorem (Theorem 1), which linked efficiency to informational invariance. This, in turn, paved the way to our take-home message, which was basically the converse of that result: we showed how realistic departures from Sims's (2003) mutual information specification map to a novel form inefficiency—what we called "cognitive externalities"—and we used this lesson to shed new light on the price system's "economy of knowledge" (Hayek, 1945).

It should be clear that, while Theorem 1 itself relied on the complete-markets assumption, our take-home message did not: the form of inefficiency that we identified in this paper, and the new angle that we offered on Hayek's classic point, naturally extend from complete to incomplete markets.

Moreover, the complete-markets assumption does not imply that inattentive agents make sophisticated portfolio choices: compared to the case with no attention costs, consumers may suboptimally allocate their spending not only across goods within a state θ, but also across different realizations of θ. For instance, an inattentive agent may fail to transfer resources from a state in which consumption is expensive to a state in which consumption is cheap, precisely because she has difficulty learning, understanding, or responding to the relevant asset prices (as embedded in the equilibrium variation of p across θ). Put differently, inattentive consumption choice in our setting captures both "grocery-choice mistakes" and "portfolio-choice mistakes."

Notwithstanding this point, one may naturally feel uncomfortable with our assumption that markets are complete with respect to both θ and ω. This amounts to letting consumers enjoy full insurance against the idiosyncratic noise (i.e., against the variation in ω conditional on θ), which is unrealistic. In the rest of this subsection, we therefore zero in on this particular feature of our setting. First, to gauge when our assumption may or may not be a useful approximation of reality, we provide conditions under which the aforementioned insurance can be vanishingly small. Next, we discuss how the absence of such insurance opens the door to pecuniary externalities à la Geanakoplos and Polemarchakis (1986), without however upsetting our own lessons about cognitive externalities. Finally, we address the subtlety of how inattentive agents meet their budgets in the absence of the aforementioned insurance.

How Much Insurance is Needed Against Idiosyncratic Mistakes?. To shed some light on this issue, in Online Appendix E we write a dynamic variant of our consumer problem which explicitly models how agents achieve insurance over idiosyncratic mistakes. The consumer faces a repeated, T-period decision problem under uncertainty about a time-invariant price vector p and state of nature θ. In each period t, a consumer i chooses their consumption bundle x_{it} contingent on a concurrent signal ω_{it}. In addition, the consumer has access to a "bank account" that helps smooth spending over time. In particular, the consumer's period-t expenditure $p(\theta)x_{it}(\omega_{it})$ is debit on the bank account and their income $P(\theta)e(\theta)$ is a credit. It follows that the final-period balance of the account is given by $b_{iT} = b\left(\{\omega_{it}\}_{t=1}^{T}, \theta\right) = \sum_{t=1}^{T} [P(\theta)(e(\theta) - x_{it}(\omega_{it})]]$.

In this context, complete markets amounts to assuming that the bank bails out the consumer whenever $b_{iT} < 0$, in return of retaining the consumer's end-of-life savings whenever $b_{iT} > 0$.[26] The required "bail-out" is non-zero in general, and it could indeed be very large if inattention results in big mistakes.

We show that, if the idiosyncratic noise is uncorrelated over time, then b_{iT} converges almost surely to a function that depends at most on θ as $T \to \infty$. That is, the debits and credits caused by idiosyncratic noise net out over time, and the required insurance over idiosyncratic mistakes converges to zero with probability one. Our interpretation is that the complete-markets assumption can be viewed as useful approximation of reality in the context of small, repeated consumption choices (e.g., weekly grocery shopping), but of course not in the context of large, once-in-a-lifetime, choices (e.g., education or home purchase).

Allowing for Incomplete Markets. Let us now take seriously the idea that there is no insurance over idiosyncratic mistakes. Toward this goal, let us first note that our original, complete-markets consumer problem in (11) can been rewritten as follows:

$$\begin{aligned}
&\max_{x,\phi,b} \sum_{\omega\in\Omega, z\in Z_\pi} u^j(x(\omega),\theta)\,\phi(\omega \mid z)\,\pi(z) - C^j[\phi,\pi] \\
&\text{s.t. } P(\theta)\cdot x(\omega) - P(\theta)\cdot e^j(\theta) - a^j\Pi(\theta) \leq b(\omega,\theta),\ \forall \omega \in \Omega, \theta \in \Theta \\
&\qquad \sum_{\omega,\theta} b(\omega,\theta)\phi(\omega|\theta)\pi(\theta) = 0
\end{aligned} \tag{25}$$

This makes explicit the (ω,θ)-contingent transfers, represented above by the function $b : \Omega \times \Theta \to \mathbb{R}$, that underlie the "consolidated" budget in (11). Removing the "exotic" half of our complete-markets assumption (i.e., the insurance over mistakes) maps to adding the following constraints to (25): $b(\omega,\theta) = b(\omega',\theta)$ for all (ω,ω',θ). Finally, if we wanted to remove insurance over θ as well, we could further restrict $b(\omega,\theta) = 0$ for all (ω,θ).

A planner that can provide the missing insurance at no cost can trivially improve welfare. More interestingly, suppose that the planner is as constrained as the market, in the sense made precise in Geanakoplos and Polemarchakis (1986). The planner can still achieve a Pareto improvement by leveraging pecuniary externalities: because such externalities do not net out under incomplete markets, and because the planner

[26] Note that such a contract does *not* require the consumer to observe θ at any point. It suffices for the bank to observe θ along with $\{\omega_{it}\}_{t=1}^{T}$. Although the observability and contractibility of $\{\omega_{it}\}_{t=1}^{T}$ is unrealistic, it is analogous to the standard assumption in Arrow-Debreu settings that agents can write contracts on idiosyncratic income and taste shocks.

can regulate the agents' exposure to them, the planner can use these externalities to substitute for the missing insurance. This logic is the same regardless of whether the missing insurance is about θ or ω.

An incomplete markets extension of our analysis would therefore naturally blend the lessons of Geanakoplos and Polemarchakis (1986) about pecuniary externalities with our own lessons about cognitive externalities and Hayek's economy of knowledge. But it is unclear what additional lesson could obtain at the present level of generality. Instead, a more fruitful direction for future research might be the exploration of how incomplete markets and rational inattention interact within specific applications.

Inattention and Budget Constraints. We close this section by discussing how inattentive agents satisfy their budget constraints despite their perfectly perceiving prices and incomes. This was trivially guaranteed in our main analysis thanks to complete markets. But whereas complete markets are *sufficient* for that purpose, they are clearly not *necessary*.

As we remove insurance possibilities in equation 25, we effectively introduce more constraints over the consumer's problem. The addition of these constraints naturally reduces payoffs, and also changes optimal behavior. But the consumer problem remains well-posed: as long as $p(\theta) > 0$ and $p(\theta) \cdot e(\theta) > 0$ for all θ, the choice set for x remains non-empty and compact, so the problem continues to admit a solution. In other words, there is no logical inconsistency between the following three statements: agents are inattentive; markets are incomplete; and agents meet their budgets.

The following, however, is also true: while complete markets permit inattentive consumers to exhaust their resources (i.e., meet their consolidated budget with equality), incomplete markets may force them to "burn money." To illustrate, suppose that $\Theta = \{\theta_L, \theta_H\}$; that $p(\theta_L) = p(\theta_H) = \bar{p} > 0$ and $\bar{p} \cdot e(\theta_L) < \bar{p} \cdot e(\theta_H)$, meaning that prices are the same across the two states and income is lowest in the L state; and that it is prohibitively costly for the consumer to learn the true state perfectly. Then, the consumer can meet their budget in the L state only by under-consuming enough in the H state. (Starvation in the L state is inevitable.)[27]

Analogous scenarios emerge in more standard incomplete markets models. Consider an (attentive) agent who dynamically chooses consumption and savings, may die with probability $q \in (0,1)$ at any age, does not have access to annuities, is not allowed to default, and does not have any bequest motives. Similarly to our inattentive agent above, this attentive agent may optimally choose to "burn money," now in the form of dying with positive wealth; and once again, the rationale is to avoid a violation of their budget in certain states. In this sense, the effect of removing complete markets on agents' ability to meet their budget constraints is symmetric in our rational-inattention setting as in a conventional setting.[28]

[27] As long as $\bar{p} \cdot e(\theta_L) > 0$, such under-consumption won't amount to "starvation". But what if $\bar{p} \cdot e(\theta_L) = 0$? If the disutility of starvation is sufficiently small enough, the consumer may optimally choose starvation along with no information acquisition. Otherwise, the consumer may opt to learn the state H perfectly, so as to avoid starvation in this state.

[28] A similar point applies to the question of default, ruled out in our model as in standard Arrow-Debreu. The incorporation of default in GE theory is a challenging task even in the absence of inattention (see, e.g., Dubey et al., 2005), so it is risky to speculate about the normative properties of settings that combine default with rational inattention. But it stands to logic that our cognitive externalities will survive in such setting, and may even take an appealing new flavor: to the extent that default is the product of inattention, and to the extent that our invariance condition is violated, a regulation of the price system may help minimize inattentive default in the same way it helps minimize consumption mistakes in our main analysis.

Notwithstanding this point, we can also consider a model variant that lets inattentive agents meet their budgets with equality in all states. We briefly outline the logic here, and analyze the model more thoroughly in Online Appendix F. We continue to assume that the consumption of goods $i = \{1,...,N-1\}$ must be contingent on the signal ω, but now assume that the consumption of the last good can be contingent on both θ and ω. The existence of such an "adjustment" good, common in the applied literature,[29] trivializes the question of how inattentive agents meet their budgets, but does not change the essence of how rational inattention and incomplete markets blend together. We show in Appendix F that, if utility is concave in x_N, then the absence of insurance translates to pecuniary externalities a la Geanakoplos and Polemarchakis (1986), in line with our earlier discussion. If, on the other hand, utility is linear in x_N, then pecuniary externalities net out and our cognitive externalities reemerge as the sole possible source of inefficiency. This second scenario maps directly to the example of Section 3 and explains why that example features *effectively* complete markets: pecuniary externalities are switched off.

To sum up, the presence or absence of an "adjustment good" does not change the essence of the issue. The key efficiency property of incomplete markets, with or without rational inattention, is that they open the door to pecuniary externalities; the key efficiency property of rational inattention, with or without complete markets, is that it opens the door to the cognitive externalities that are the focus of our main analysis.

7.2 Expanding the Planner's Options and the Cognition State z

Our notions of feasibility and efficiency in Definitions 2-3 allowed messages to replace prices but did not give the planner any of the following options: to "customize" messages, sending different messages to different agents; to relabel or merge the underlying states of nature; and to preclude all or some agents from learning directly about the state of nature. The last option could be interpreted quite literally as the power to restrict access to information.[30] The option to merge states, on the other hand, could represent the planner's choice to "uncomplete" the markets—by making agents unable to consume or produce differently in different state realizations, the planner effectively shuts down agents' ability to cross-insure between them. We can readily show an extension of our main result to such an expansion of the planner's powers, provided a commensurate enlargement in the assumed invariance of attention costs.

Toward this result, let the planner now manipulate the entire z via a collection of functions $Z^j : \Theta \to \Theta \times \mathbb{R}^N_+$, one for each type of consumer and firms. Let $\mathscr{A}^{\text{all}}$ denote the set of *all* such collections of functions, and let the planner choose this collection of functions from a subset $\mathscr{A} \subseteq \mathscr{A}^{\text{all}}$, which embeds the precise ways in which the planner may or may not manipulate cognition. Our benchmark model is nested as one such $\mathscr{A}$ that restricts the message space.

Our notions of equilibrium (Definition 1) and efficiency (Definition 3) are unchanged, modulo to the following two natural adjustments. First, different types of agents are now allowed to have different priors,

[29]For instance, this is second-period consumption in Sims (2006) and the y good in Lian (2021). In many other papers, such an adjustment good is implicit in the assumption that utility is linear in "money" or net wealth (e.g., Colombo et al., 2014; Kőszegi and Matějka, 2020; Ravid, 2020).

[30]Under this first interpretation, our upcoming result, showing conditions under which such restrictions are *not* optimal, relates also to a literature on information design with inattentive receivers (Lipnowski et al., 2020; Bloedel and Segal, 2021).

henceforth denoted by π^j; and second, consistency of the prior requires $\pi^j(z) = \sum_{\theta'} \pi_\Theta(\theta') \cdot 1_z(Z^j(\theta'))$ for all $z \in \text{Im}[Z^j]$ and all j.[31] The following version of Theorem 1 then applies provided attention costs are invariant to the full set of transformations:

Corollary 5. *Assume that all consumers and firms have attention costs that are invariant with respect to the full set of transformations* $\{g : (\Theta \times \mathbb{R}^N_+) \to (\Theta \times \mathbb{R}^N_+)\}$. *Then, equilibria are efficient with respect to the enlarged feasibility concept described above.*

The only difference from the proof of Theorem 1 is the translation from Proposition 5 to the desired result. Because the planner can now manage attention not only by replacing prices but also by directly relabeling the states of nature or compressing them in fewer contingencies, we must invoke invariance with respect to the entire set of transformations as opposed to the subset G^p.

One can push this argument in a slight different direction as follows. In our main analysis, we defined the cognition state z as the combination of the state of nature and the price vector. This was motivated by the sufficiency of (θ, p) to describe each agent's decision problems in "free markets." But if we think of more realistic market structures, augmented with taxation or regulation, policy tools may naturally enter as additional elements in the agent's decision problem. This invites a redefinition of the cognition state from $z = (\theta, p)$ to $z = (\theta, p, \tau)$, where τ is the tax, to formalize how agents may be differentially able to attend to prices and taxes.[32] Alternatively, one could imagine including in z other endogenous objects that may enter cognition even if they do not directly enter payoffs, like average trades $(\overline{x}_j)_{j=1}^J$ of each type. In Online Appendix G, we build the machinery to study economies with such a flexible definition of the cognition state z and even the potential for the social planner to send messages in a different space. We conjecture that our main result extends with the appropriately extended notion of invariance—as implied, for example, by measuring costs by the mutual information between ω and z, regardless of the space of z.

7.3 Alternative Behavioral Frictions

In this subsection, we comment on how our methodology and insights could extend to models of behavioral frictions other than unrestricted costly information acquisition. The key idea is to cast these frictions as the product of rational behavior under a general form of cognitive, or optimization, costs.

Narrow Bracketing and Bounded Recall. Let the signal variable have N sub-components, indexed as $\omega = (\omega_n)_{n=1}^N$, and require that x_n is measurable in ω_n for all n. This defines ω_n as the information set upon which the consumption of good n must be conditioned on, but does not by itself put any restriction on how correlated this information may be across n. Our main analysis can now be nested by assuming that there is

[31] It is also necessary to redefine the domain of attention costs when priors over z do not have a marginal distribution π_Θ on the first element. In particular, we define the set of mass functions

$$\mathscr{P}' \equiv \left\{\pi : \Theta \times \mathbb{R}^N_+ \to [0,1] \text{ s.t. } \pi(\theta, z) = \sum_{\theta' \in \Theta} \pi_\Theta(\theta') \cdot 1_z(f(\theta')) \text{ for some } f : \Theta \to \Theta \times \mathbb{R}^N_+\right\} \tag{26}$$

and require that attention costs are well-defined functions $C : \Delta(\Omega)^{|\Theta|} \times \mathscr{P}' \to \mathbb{R}_+$.

[32] The behavioral literature on "tax salience" (e.g., Chetty et al., 2009) provides direct evidence that agents behave as if they perceive the tax component of prices inaccurately.

neither a gain nor a loss, in terms of C, from making ω_n have the same information for all n. The scenario often considered in applied work, where budgets clear via an "adjustment good", is nested by letting $\omega_n = \omega$ for all $n < N$ and $\omega_N = (\omega, \theta)$. More generally, if we let information sets differ across goods, we can nest the model of "narrow thinking" proposed by Lian (2021) and, by extension, the type of narrow bracketing captured therein. If we interpret the index of goods, n, as different time periods, then we can capture learning and/or bounded recall.[33]

There is no obstacle to proving an extension of Theorem 1 that carries the restriction that x_n be measurable in ω_n.[34] Thus we can accommodate fairly broad notions of asymmetric cognitive constraints across different goods choices or different time periods without necessarily opening the door to government intervention. In fact, this is possible even in the "vanilla" model from our main analysis. This follows from combining the results of Kőszegi and Matějka (2020), which show how narrow bracketing can arise as the optimal solution to multi-dimensional tracking problems with mutual information costs, with our result that the invariance property of such costs guarantees efficiency.

On the other hand, it may be reasonable to let narrow bracketing arise *because* of a failure of invariance. In particular, the core assumption in Lian (2021) is that it is primitively cheaper for ω_n to contain information about p_n than about p_r for $r \neq n$. This assumption captures the natural idea that "the price of apples is less salient than the price of bananas when choosing how many bananas to buy, and vice versa." Similarly, if bounded recall means it is less costly for ω_n to contain information about current prices than about the past ones, then it, too, amounts to a violation of our invariance condition. In such cases, our results suggest that it may be possible to improve welfare by manipulating the informational content of prices or other endogenous objects that attract people's attention. Put differently, if "What You See is All There Is" in the phrasing of Kahneman (2011), a social planner may be motivated to make "What You See" especially informative.

Stochastic Choice. Models of costly information acquisition are nested within the broader class of models of *state-dependent stochastic choice*. The converse is not true: there are models of state-dependent stochastic choice which cannot be micro-founded as models of information acquisition. Such models can directly be motivated as models of costly control or trembling hands (see, e.g., Morris and Yang, 2022; Flynn and Sastry, 2021). We now illustrate how our results can be extended to these contexts.

For the sake of simplifying the argument, focus on an exchange economy and let the consumption space $\mathcal{X}$ be discrete. [35] Next, let $\psi(x|z) \in \Delta(\mathcal{X})$ denote the probability of consuming x when the cognition state is z. Keeping with previous notation, we use the shorthand notation $\psi = (\psi(\cdot \mid z))_{z \in Z_\pi} \in (\Delta(\mathcal{X}))^{|\Theta|}$. Let attention

[33] Order goods such that lower n corresponds to earlier periods and consider an economy with one "actual" good, consumed over many periods. Then, learning corresponds to ω_n containing more information about $z = (\theta, p)$ than ω_m for $m < n$; and bounded recall corresponds to any scenario in which at least some of the information contained in ω_m is not contained in ω_n, for $m < n$.

[34] This was proved in an earlier draft (Angeletos and Sastry, 2019): the extension with choice-specific signals described above was actually that draft's main model.

[35] Otherwise, we could extend our notation and continuity notion to handle distributions on a continuous commodity space, similar to what we do in Appendix D for a continuous signal space.

costs be given by some function $K^j : (\Delta(\mathscr{X}))^{|\Theta|} \times \mathscr{P} \to \mathbb{R}$. The consumer's problem is as follows:

$$\begin{aligned} &\max_{\psi} \sum_{x,z} u^j(x,\theta)\, \psi(x \mid z)\, \pi(z) - K_X^j[\psi,\pi] \\ &\text{s.t.} \sum_{x,z} (p\cdot x - p\cdot e^j(\theta) - a^j \Pi(\theta))\, \psi(x \mid z)\, \pi(z) \leq 0 \end{aligned} \tag{27}$$

The definitions of equilibrium, efficiency and invariance are similarly adapted. The arguments in Lemma 1 and hence also Theorem 1 then follow from the same premises. What changes is only the interpretation of our invariance condition: invariance now refers to whether the costs of *random plans of action*, as opposed to signals, are sensitive or not to a specific labeling of the state space.

When cost functionals satisfy Infeasible Perfect Discrimination (IPD) as proposed by Morris and Yang (2022), which is loosely speaking a notion of "continuous stochastic choice," invariance is necessarily violated. Concretely, if the grocery store randomly switches which of two substitutable products are \$1.99 versus \$2.00, an IPD consumer struggles to shift consumption decisively from one to the other, and a planner would do better to move those prices further apart. Conversely, the likelihood-separable costs motivated by Flynn and Sastry (2021), to study trembling hands conditional on observing the state, are invariant in the required ways. Inattentive economies with such a friction would therefore be efficient as per our results.

Default Points and Sparsity. The control-cost formulation in equation 27 can also capture "default points," as studied in a long tradition of behavioral economics (e.g., Tversky and Kahneman, 1991). To illustrate, suppose that $K_X^j[\psi,\pi]$ were a penalty for picking x away from $x^{j,d}$, where $x^{j,d}$ is a type-specific default point. The latter could be a deterministic variable that depends on π or a random variable that depends on both z and p. If the default point depends on the stochastic properties of p, then costs are non-invariant.

Gabaix's (2014) model of "sparsity" naturally fits in this discussion: that model amounts to a set of assumptions about the default point and the aforementioned penalty. But there is an important difference: whereas Proposition 8 in Gabaix (2014) compares competitive equilibria to a paternalistic planner that does *not* internalize agents' cognitive costs, our "sparsity-meets-rational-inattention" approach invites one to take the opposite stance and to reconsider the policy implications of that paper (and of Farhi and Gabaix, 2020). In particular, our approach equates sparsity to a potential failure of invariance: the relevant default point is connected to the full-attention optimal consumption plan, which itself is endogenous to prices. From this perspective, the normative properties hinge on the possibility of cognitive externalities.

7.4 Markets vs. Games

Complete, competitive markets can be understood as a special class of games, in which players are infinitesimal and payoff externalities are muted on equilibrium. This suggests that our main insight, regarding the role of invariance for efficiency, may extend to games. Consistent with this guess, Hébert and La'O (2020) establish efficiency of equilibria in their class of games under two conditions: a close cousin of our invariance condition; and a restriction on payoffs that, in our market context, translates to netting-out of pecuniary externalities. Together with our discussion in Section 7.1, this hints at the following idea: an extension of our analysis to incomplete markets a la Geanakoplos and Polemarchakis (1986) is likely to contain similar insights as an extension of Hébert and La'O (2020) to games without the aforementioned payoff restriction.

It is also worth clarifying the following, subtler point. Our equilibrium and efficiency concepts specify the agents' prior about z and their associated attention choices on equilibrium but not off equilibrium. This is innocuous here because agents are infinitesimal and off-equilibrium beliefs are immaterial. But this is not true in settings with large players, where the threat not to pay attention to something off equilibrium could influence what happens on equilibrium (see, e.g., Ravid, 2020). The applicability of our paper's logic to such settings remains an open question.

8 Conclusion

Cognitive frictions cause individuals to make mistakes. These can propagate in markets, causing others to change their behavior and the economy as a whole to malfunction relative to the textbook scenario with fully rational and attentive agents. But, unless a social planner has the power to "cure" cognitive frictions, it is not obvious why the planner should try to regulate these mistakes or manipulate market outcomes.

Our main results formalized how rational inattention may or may not lead to inefficiency. If attention costs are invariant in the sense defined in this paper and markets are complete, then market outcomes are efficient. If attention costs are not invariant, then market outcomes may be inefficient due to unchecked *cognitive externalities*, whereby one's actions affect the stochastic properties of prices and, in turn, others' propensity to make mistakes. In such cases, there may be room for policies that manipulate or "simplify" the stochastic properties of prices or even shut down certain "confusing" markets. Additional results provided conditions for existence of equilibria, for implementation of Pareto optima, and for equilibrium attention to be concentrated on "fundamental" objects. We also discussed how to map these lessons to settings with different market structures, social planning concepts, and underlying cognitive frictions.

Our analysis drew a link between two seemingly disparate issues in information economics: the validity of Hayek's (1945) argument about the "economy of knowledge" afforded by the price system was shown to hinge on the appropriateness of Sims's (2003) mutual information specification for attention costs. This reinforces the value of an active decision-theoretic and experimental literature that departs from mutual information costs within the rational-inattention framework. Such departures offer the promise of understanding jointly individual behavior (the focus of this literature) and equilibrium properties including efficiency (the focus of our paper), without either a violation of individual rationality or the presumption of a policy maker that can cure, bypass, or ignore people's cognitive constraints.

Our analysis committed to the interpretation of internal signals as a representation of inattention, cognition, or stochastic choice. This interpretation was most suitable for the connections we built to the related advances in decision theory and experimental economics. A literal interpretation in terms of collecting and processing market data is also possible. Our results then apply to the extent that such information represents primarily a private good, as in the case of a monopolist learning about demand or costs. How reasonable invariance is in this context is an open question.

References

AMARI, S.-I. (2016): *Information geometry and its applications*, vol. 194, Springer.

AMARI, S.-I. AND H. NAGAOKA (2000): *Methods of Information Geometry*, vol. 191, American Mathematical Society and Oxford University Press.

ANGELETOS, G.-M. AND J. LA'O (2020): "Optimal monetary policy with informational frictions," *Journal of Political Economy*, 128, 1027–1064.

ANGELETOS, G.-M. AND K. SASTRY (2019): "Inattentive Economies," Working Paper 26413, National Bureau of Economic Research, December 2019 Revision.

ARROW, K. J. AND G. DEBREU (1954): "Existence of an equilibrium for a competitive economy," *Econometrica*, 265–290.

BLOEDEL, A. W. AND I. R. SEGAL (2021): "Persuading a Rationally Inattentive Agent," Working paper, UCLA.

BLOEDEL, A. W. AND W. ZHONG (2021): "The cost of optimally-acquired information," Working paper, UCLA.

CAPLIN, A. AND M. DEAN (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105, 2183–2203.

CAPLIN, A., M. DEAN, AND J. LEAHY (2022): "Rationally inattentive behavior: Characterizing and generalizing Shannon entropy," *Journal of Political Economy*, 130, 1676–1715.

CHETTY, R., A. LOONEY, AND K. KROFT (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–77.

COLOMBO, L., G. FEMMINIS, AND A. PAVAN (2014): "Information Acquisition and Welfare," *Review of Economic Studies*, 81, 1438–1483.

DEAN, M. AND N. L. NELIGH (2022): "Experimental Tests of Rational Inattention," Working paper, Columbia University.

DEBREU, G. (1954): "Valuation Equilibrium and Pareto Optimum," *Proceedings of the National Academy of Sciences of the United States of America*, 40, 588–592.

DENTI, T. (2022): "Posterior separable cost of information," *American Economic Review*, 112, 3215–59.

DENTI, T., M. MARINACCI, AND A. RUSTICHINI (2022): "Experimental cost of information," *American Economic Review*, 112, 3106–23.

DUBEY, P., J. GEANAKOPLOS, AND M. SHUBIK (2005): "Default and punishment in general equilibrium," *Econometrica*, 73, 1–37.

Farhi, E. and X. Gabaix (2020): "Optimal taxation with behavioral agents," *American Economic Review*, 110, 298–336.

Flynn, J. P. and K. A. Sastry (2021): "Strategic Mistakes," Working Paper 3663481, SSRN.

Gabaix, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 129, 1661–1710.

——— (2016): "A Behavioral New Keynesian Model," *NBER Working Paper No. 22954*.

——— (2017): "A Behavioral New Keynesian Model," *Harvard mimeo*.

——— (2019): "Behavioral Inattention," in *Handbook of Behavioral Economics*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier, vol. 2.

Geanakoplos, J. and H. M. Polemarchakis (1986): "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete," *Uncertainty, Information and Communication: Essays in Honor of K.J. Arrow*, 3, 65–96.

Grossman, S. J. (1981): "An Introduction to the Theory of Rational Expectations Under Asymmetric Information," *The Review of Economic Studies*, 48, 541–559.

Grossman, S. J. and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393–408.

Gul, F., W. Pesendorfer, and T. Strzalecki (2017): "Coarse Competitive Equilibrium and Extreme Prices," *American Economic Review*, 107, 109–37.

Hayek, F. A. (1945): "The Use of Knowledge in Society," *The American Economic Review*, 35, 519–530.

Hébert, B. and J. La'O (2020): "Information Acquisition, Efficiency, and Non-Fundamental Volatility," *Stanford GSB and Columbia University mimeo*.

Hébert, B. and M. Woodford (2021): "Neighborhood-based information costs," *American Economic Review*, 111, 3225–3255.

——— (2023): "Rational inattention when decisions take time," *Journal of Economic Theory*, 105612.

Hurwicz, L. (1972): "On the Dimensional Requirements of Informationally Decentralized Pareto-satisfactory Processes," Tech. rep., presented at the Conference Seminar on Decentralization at Northwestern University.

Jordan, J. S. (1982): "The competitive allocation process is informationally efficient uniquely," *Journal of Economic Theory*, 28, 1–18.

Kahneman, D. (2011): *Thinking, Fast and Slow*, Macmillan.

KŐSZEGI, B. AND F. MATĚJKA (2020): "Choice simplification: A theory of mental budgeting and naive diversification," *The Quarterly Journal of Economics*, 135, 1153–1207.

LAFFONT, J.-J. (1985): "On the Welfare Analysis of Rational Expectations Equilibria with Asymmetric Information," *Econometrica*, 53, 1–29.

LIAN, C. (2021): "A theory of narrow thinking," *The Review of Economic Studies*, 88, 2344–2374.

LIPNOWSKI, E., L. MATHEVET, AND D. WEI (2020): "Attention management," *American Economic Review: Insights*, 2, 17–32.

MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2023): "Rational Inattention: A Review," *Journal of Economic Litertaure*, forthcoming.

MAĆKOWIAK, B. AND M. WIEDERHOLT (2015): "Business Cycle Dynamics under Rational Inattention," *Review of Economic Studies*, 82, 1502–1532.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*, Oxford University Press.

MONDRIA, J., X. VIVES, AND L. YANG (2022): "Costly interpretation of asset prices," *Management Science*, 68, 52–74.

MORRIS, S. AND M. YANG (2022): "Coordination and continuous stochastic choice," *The Review of Economic Studies*, 89, 2687–2722.

MOUNT, K. AND S. REITER (1974): "The informational size of message spaces," *Journal of Economic Theory*, 8, 161–192.

POMATTO, L., P. STRACK, AND O. TAMUZ (2023): "The Cost of Information: The Case of Constant Marginal Costs," *American Economic Review*, Forthcoming.

RAVID, D. (2020): "Ultimatum Bargaining with Rational Inattention," *American Economic Review*, 110, 2948–63.

SHAW, M. L. AND P. SHAW (1977): "Optimal allocation of cognitive resources to spatial locations," *Journal of Experimental Psychology: Human Perception and Performance*, 3, 201–211.

SIMS, C. A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356.

——— (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665–690.

——— (2006): "Rational inattention: Beyond the linear-quadratic case," *The American Economic Review*, 96, 158–163.

——— (2010): "Rational Inattention and Monetary Economics," *Handbook of Monetary Economics*, 3, 155–181.

SONNENSCHEIN, H. F. (2017): "Chicago and the origins of modern general equilibrium," *Journal of Political Economy*, 125, 1728–1736.

TIAN, J. (2021): "Informational Separability and Entropy," *Available at SSRN 3873168*.

TIROLE, J. (2015): "Cognitive Games and Cognitive Traps," *Toulouse School of Economicss mimeo*.

TVERSKY, A. AND D. KAHNEMAN (1991): "Loss aversion in riskless choice: A reference-dependent model," *The Quarterly Journal of Economics*, 106, 1039–1061.

VIVES, X. (2017): "Endogenous Public Information and Welfare in Market Games," *Review of Economic Studies*, 84, 935–963.

WOODFORD, M. (2012): "Inattentive valuation and reference-dependent choice," Working paper, Columbia University.

——— (2020): "Modeling imprecision in perception, valuation, and choice," *Annual Review of Economics*, 12, 579–601.

A Omitted Proofs of Main Results

This section completes the proofs of the main results in Sections 4 and 5. Proofs for all other results are in Online Appendix A.

A.1 Proof of Proposition 5

We prove this by contradiction, following closely the textbook proof in Chapter 16 of Mas-Colell et al. (1995).

Let the competitive equilibrium, which we assume to exist, be denoted by $((x^j,\phi^j)_{j=1}^J,(y,\phi^F),P)$, and aggregate demands and production by $((\overline{x}^j(\theta))_{j=1}^J,\overline{y}(\theta))$. Since we have assumed equilibrium exists, it must be the case that for each consumer (x^j,ϕ^j) solves program (17). Let the proposed variant allocation be denoted by $((x^{j\prime},\phi^{j\prime})_{j=1}^J,(y',\phi^{F\prime}),P)$, and aggregate demands and production be denoted by $((\overline{x}^{j\prime}(\theta))_{j=1}^J,\overline{y}'(\theta))$. It is without loss to assume that $(x^{j\prime},\phi)$ solves (17) for parameters $(\overline{x}^{j\prime}(\theta))_{j=1}^J$ and π. Otherwise we could construct another feasible allocation with weakly higher payoffs for each agent, and then apply the proof by contradiction to this variant allocation. Note finally that the message and prior are unchanged from the price and prior in the equilibrium.

Assume that, under the variant allocation, all agents are weakly better off and some positive mass of agents are strictly better off. Under our restriction to symmetry in allocations within types, this implies that $\overline{u}^j((\overline{x}^{j\prime}(\theta))_{\theta\in\Theta},\pi)\geq\overline{u}^j((\overline{x}^j(\theta))_{\theta\in\Theta},\pi)$, for all types j, holding strictly for at least one type.

We now establish that $\sum_\theta P(\theta)\cdot\overline{x}^{j\prime}(\theta)\,\pi_\Theta(\theta) \geq \sum_\theta (P(\theta)\cdot e^j(\theta) + a^j\Pi(\theta))\,\pi_\Theta(\theta)$ for all agents j, with inequality for at least one type. Let us first establish the inequality. Assume instead that $\sum_\theta P(\theta)\overline{x}^{j\prime}(\theta)\,\pi_\Theta(\theta) < \sum_\theta (P(\theta)e^j(\theta) + a^j\Pi(\theta))\,\pi_\Theta(\theta)$. In this case, given that $P(\theta) \in \mathbb{R}^N_+$ for each θ, there exists an ϵ such that $\sum_\theta (P(\theta)\cdot\overline{x}^{j\prime}(\theta))\,\pi_\Theta(\theta) + \epsilon\sum_\theta (P(\theta)\cdot\boldsymbol{e})\,\pi_\Theta(\theta) < \sum_\theta (P(\theta)e^j(\theta) + a^j\Pi(\theta))\,\pi_\Theta(\theta)$, where $\boldsymbol{e}\in\mathbb{R}^N$ is a vector of ones. Moreover,

$$\overline{u}^j((\overline{x}^{j\prime}(\theta)+\epsilon\boldsymbol{e})_{\theta\in\Theta},\pi) > \overline{u}^j((\overline{x}^{j\prime}(\theta))_{\theta\in\Theta},\pi) \geq \overline{u}^j((\overline{x}^{j}(\theta))_{\theta\in\Theta},\pi) \tag{28}$$

where the first, strict inequality uses the monotonicity of preferences established in Lemma 2, stated and proven at the end of this proof.[36] Because this bundle is strictly preferred to (x^j,ϕ^j) and feasible given the same prices (and profits), its existence would contradict consumer optimality. We use a similar argument to establish that $\sum_\theta P(\theta)\overline{x}^{j\prime}(\theta)\,\pi_\Theta(\theta) > \sum_\theta (P(\theta)e^j(\theta) + a^j\Pi(\theta))\,\pi_\Theta(\theta)$ for agents experiencing a strict utility gain in the new allocation. If not, $\overline{x}^{j\prime}(\theta)$ would be feasible and preferred, contradicting consumer optimality.

Adding up the previously established conditions and simplifying gives

$$\begin{aligned}
\sum_j\sum_\theta (P(\theta)\cdot\overline{x}^{j\prime}(\theta))\,\pi_\Theta(\theta) &> \sum_j\sum_\theta (P(\theta)\cdot e^j(\theta) + a^j\Pi(\theta))\,\pi_\Theta(\theta)\\
&> \sum_j\sum_\theta (P(\theta)\cdot e^j(\theta)) + \sum_\theta (P(\theta)\cdot y(\theta))\\
&> \sum_j\sum_\theta (P(\theta)\cdot e^j(\theta)) + \sum_\theta (P(\theta)\cdot y'(\theta))
\end{aligned} \tag{29}$$

where the second line uses $\sum a^j = 1$ and $\Pi(\theta) = P(\theta)\cdot y(\theta)$, and the third line uses the guarantee from profit maximization that $\sum_\theta (P(\theta)\cdot y(\theta)) \geq \sum_\theta (P(\theta)\cdot y''(\theta))$ for any $(y''(\theta))_{\theta\in\Theta}$ in $\overline{\boldsymbol{F}}(\pi)$ and the assumed feasibility of $(y'(\theta))_{\theta\in\Theta}$. But if the last line is true, then it is a contradiction of feasibility. Therefore the proposed Pareto dominating allocation cannot exist.

We now complete the argument by stating and proving the required monotonicity of preferences:

Lemma 2. *For each type j, each prior π, and each pair $\overline{\boldsymbol{x}},\overline{\boldsymbol{x}}' \in \mathscr{X}^{|\Theta|}$, the following is true: if $\overline{\boldsymbol{x}}' \gg \overline{\boldsymbol{x}}$, then $\overline{u}^j(\overline{\boldsymbol{x}}',\pi) > \overline{u}^j(\overline{\boldsymbol{x}},\pi)$.*

Proof. This proof is constructive. Let $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}}' \gg \overline{\boldsymbol{x}}$ be two consumption vectors in $\mathscr{X}^{|\Theta|}$. Necessarily, $\overline{\boldsymbol{x}}$ is in the interior of $\mathscr{X}^{|\Theta|}$. Let $\overline{x}_n(\theta)$ denote consumption of the nth good. Define $d\in\mathbb{R}^N_+$ as the point-wise minimum increase across states in the aggregate consumption vector or $d \equiv (\min_\theta(\overline{x}'_n(\theta) - \overline{x}_n(\theta)))_{n=1}^N$. Now let us take the optimizing values (x,ϕ) which we assume to exist for program (17) with parameters $\overline{\boldsymbol{x}}$ and π. Let $d^\omega = (d_n^\omega)_{n=1}^N = (x_n^{\max} - x_n(\omega))_{n=1}^N$ be the distance to the boundary of $\mathscr{X}$ in each dimension; and note that for any interior $\overline{\boldsymbol{x}}$, that the "capped" vector $\tilde{d} \equiv (\min\{d_n, d_n^\omega\})_{n=1}^N$ also has all strictly positive elements.

Construct $x'(\omega) = x(\omega) + \tilde{d}(\omega)$ for each ω and note that

$$\sum_\omega x'(\omega)\,\phi(\omega\mid f_\pi(\theta)) \leq d + \sum_\omega x(\omega)\,\phi(\omega\mid f_\pi(\theta)) \leq d + \overline{x}(\theta),\ \forall\theta\in\Theta \tag{30}$$

[36] We also use the fact that the original equilibrium allocation was interior to $\mathscr{X}^\Theta$ to establish that the proposed deviation lies in $\mathscr{X}^\Theta$. This was without loss of generality from setting the boundaries of $\mathscr{X}$ sufficiently large (see footnote 9).

where the third statement uses the feasibility constraint in (17). Note that $d + \overline{x}(\theta) < \overline{x}'(\theta)$ for all θ by construction. Therefore $(x'(\omega), \phi)$ is feasible in the variant program with parameters $(\overline{x}'(\theta))_{\theta\in\Theta}$ and π.

Observe next that, for a positive measure of ω, $x'(\omega) > x(\omega)$.[37] Moreover, because $x'(\omega) \gg x(\omega)$ for a positive measure of ω and $u^j(\cdot,\theta)$ represents weakly monotone preferences for each θ, expected utility is also strictly ranked:

$$\sum_{\omega,\theta} u^j(x(\omega),\theta)\,\phi(\omega \mid f_\pi(\theta))\pi_\Theta(\theta) < \sum_{\omega,\theta} u^j(x'(\omega),\theta)\,\phi(\omega \mid f_\pi(\theta))\pi_\Theta(\theta) \tag{31}$$

Therefore, $\sum_{\omega,\theta} u^j(x'(\omega),\theta)\,\phi(\omega \mid f_\pi(\theta))\pi_\Theta(\theta) - C[\phi,\pi] > \overline{u}^j(\overline{\boldsymbol{x}},\pi)$. Since a maximum exists to program (17) we know also that

$$\overline{u}^j(\overline{\boldsymbol{x}}',\pi) \geq \sum_{\omega,\theta} u^j(x'(\omega),\theta)\,\phi(\omega \mid f_\pi(\theta))\pi_\Theta(\theta) - C[\phi,\pi] \tag{32}$$

Therefore, $\overline{u}^j(\overline{\boldsymbol{x}}',\pi) > \overline{u}^j(\overline{\boldsymbol{x}},\pi)$, which completes the proof. □

A.2 Proof of Proposition 7

If no equilibria exist, then the statement is true. Consider next the case in which at least one equilibrium exists. Denote the attention strategy of each agent type in this equilibrium as ϕ^{j*}, the price functional as P^*, and the associated prior as π^*. Definition 7 guarantees that there exists some transformation $g \in G^p$ such that $C[\phi^{j*},\pi^*] \geq C[\tilde{\phi}^{j*},\tilde{\pi}^*]$ for all j, where $(\tilde{\phi}^{j*},\tilde{\pi}^*)$ is the transformation under g. Moreover, the previous holds with strict inequality for at least one consumer type j, which collects information. Moreover, $C^x[\tilde{\phi}^{j*},\tilde{\pi}^*] \leq C^x[\phi^{j*},\pi^*]$ for each $x \in \{1,\ldots,J\} \cup \{F\}$, because $C^x = C$ for all agents (Definition 8).

Let this function g be represented as $g(\theta, p) = (\theta, h(p))$. The social planner can implement the original allocation of goods and attention with the new message $M(\theta) = h(P^*(\theta))$. This remains feasible because firms' production constraint is more slack, or $C^F[\tilde{\phi}^{F*},\tilde{\pi}^*] \leq C^F[\phi^{F*},\pi^*]$, and therefore $H(y(\omega), C^F(\phi^{F*},\tilde{\pi}^*),\theta) \leq 0$ for all (ω,θ) such that $\phi(\omega \mid f_{\tilde{\pi}^*}(\theta)) > 0$. This new allocation strictly improves payoffs for at least one agent type and leaves all other payoffs unaffected. Therefore the equilibrium is Pareto dominated. The economy thus has one Pareto-dominated equilibrium, and is therefore inefficient by our definition. This proves the original claim.

B Invariance in the Space of Posteriors

In this Appendix, we formalize the relationship of our invariance condition with notions of invariance in the literature on information geometry and the axiomatization of mutual information in Caplin et al. (2022). We refer the reader to Hébert and La'O (2020) for a complementary, and more detailed, discussion of how the

[37] If not, then for measure 1 of ω, we have $x(\omega) = (x_n^{\max})_{n=1}^N$. But this implies $\overline{x}(\theta) = (x_n^{\max})_{n=1}^N$ for all θ and there cannot exist an $\overline{x}'(\theta)$ that is larger in every dimension.

invariance condition in our paper and theirs alike connects to this literature, as well as to various information costs proposed in the literature.

We hereafter restrict attention to the class of *posterior separable costs*, which is studied by Caplin and Dean (2015) and Denti (2022) and encompasses many specifications used in the literature (e.g., Sims, 2003; Pomatto et al., 2023; Hébert and Woodford, 2021). With our notation, this class is defined as follows.

Definition 12. Attention costs are *posterior separable* if they admit the following representation:

$$C[\phi, \pi] = \sum_{\omega \in \Omega} \phi_\omega(\omega) \cdot T\left[\phi_{z|\omega}(\cdot|\omega); \pi\right] - T[\pi; \pi] \tag{33}$$

where $\phi_\omega(\omega) = \sum_z \phi(\omega \mid z)\, \pi(z)\ \forall \omega$, $\phi_{z|\omega}(z \mid \omega) = \frac{\phi(\omega|z)\,\pi(z)}{\phi_\omega(\omega)}\ \forall \omega : \phi_\omega(\omega) > 0$ (and $\phi_{z|\omega}(z \mid \omega) = \pi(z)$ otherwise), and $T[\cdot; \pi] : \mathscr{P} \to \mathbb{R}$ is strictly convex for each $\pi \in \mathscr{P}$.

Loosely speaking, a posterior separable cost functional corresponds the expected increase in a measure of the difference between the posterior to the prior. A signal structure, represented by its induced posterior distributions, costs more if it generally induces posteriors that differ from the prior. This makes the specification of T the key to understanding the economic properties of information costs. We therefore ask how our notions of invariance and monotonicity of C translate to properties of T.

We now define a notion of invariance for T adapted from the literature on information geometry, which has recently been applied in economics by Hébert and Woodford (2023, 2021), Hébert and La'O (2020), and Caplin et al. (2022).[38]

Definition 13 (Invariance, from Amari and Nagaoka (2000) and Amari (2016))**.** Let $\pi \in \mathscr{P}$ and $\pi' \in \mathscr{P}$ be two distributions on $\mathscr{Z}$, fix a $g : \mathscr{Z} \to \mathscr{Z}$, and construct $\tilde{\pi}$ and $\tilde{\pi}'$ as in (15). The functional T satisfies invariance if $T[\pi', \pi] \geq T[\tilde{\pi}', \tilde{\pi}]$ for any g, with equality if and only if the distributions can be written as $\pi(z) = \tilde{\pi}(g(z))r(z)$ and $\pi'(z) = \tilde{\pi}'(g(z))r(z)$ for the same $r : \mathscr{Z} \to [0, 1]$.

When applying this definition to T in the posterior separable cost (33), π and π' are the prior and posterior about a random variable z, and $\tilde{\pi}$ and $\tilde{\pi}'$ are the implied prior and posterior about the random variable $\tilde{z} = g(z)$, for some $g \in \mathscr{G}$. T is invariant if the "difference" between prior and posterior weakly decreases with any such transformation, and remains the same if and only if the prior and posterior completely agree about the realizations of the state that have been relabeled and/or combined.

The next Lemma verifies that invariant and monotone T in a posterior separable model implies a cost functional that is invariant and monotone per our definition:

Lemma 3. *Suppose that the cost functional is posterior separable, that T is invariant in the sense of Definition 13, and that $T[\pi; \pi] = 0$ for any $\pi \in \mathscr{P}$. Then, the cost functional is invariant and monotone with respect to $\mathscr{G}$ in the sense of Definition 6.*

[38] Most precisely, this is a specialization of the definition from Amari and Nagaoka (2000) and Amari (2016) that imposes that transformations preserve the state space.

We give the proof in Online Appendix C.8. Most of the argument is algebraic manipulation to show that the notions of "sufficiency" coincide.

We can use this method to construct invariant cost functionals. Theorem 3.1 in Amari (2016) (p. 54) demonstrates that the class of divergences defined by

$$T[\pi;\pi'] = \sum_z \pi(z) \cdot f\left(\frac{\pi'(z)}{\pi(z)}\right) \tag{34}$$

for a differentiable convex $f(\cdot)$ satisfying $f(1) = 0$ are invariant. When $f(u) = -\log u$, T is the Kullback-Leibler divergence, which verifies the following:

Corollary 6. *Mutual information costs are invariant and monotone in the sense Definition 6.*

Online Appendices for "Inattentive Economies" by Angeletos and Sastry

C Additional Proofs

Here, we prove the results from Sections 2 (Example: A Linear-Quadratic-Guassian Economy), 6 (Additional Results), and 7 (Extensions and Open Questions).

C.1 Proof of Proposition 1

The main logic of the proof is given in the main text. Here, we fill in the calculations.

Consumer Problem. We first derive the optimal coconut demand of an agent conditional on receiving a signal ω. The following program to maximize over coconut demand, conditional on beliefs for (ω, z)

$$\max_{x:\mathbb{R}\to\mathbb{R}} \int\int \left(x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega)) \right) \phi(\omega|z)\, \pi(z)\, \mathrm{d}\omega\, \mathrm{d}z \tag{35}$$

is globally concave. The first-order condition for each $x(\omega)$ is

$$\int (1 - x(\omega) - p)\, \phi(\omega \mid z)\, \pi(z)\, \mathrm{d}z = 0 \tag{36}$$

Rearranging gives

$$x(\omega) = 1 - \int p \frac{\phi(\omega \mid z)\, \pi(z)}{\int \phi(\omega \mid z')\, \pi(z')\, \mathrm{d}z'} \mathrm{d}z = 1 - \mathbb{E}[p \mid \omega] \tag{37}$$

where the second equality rewrites the right-hand-side as a conditional expectation.

Next, we specialize this demand function to the relevant case in which: p is Gaussian with mean $\mathbb{E}[p]$ and agents obtain a Gaussian signal of the price with mean p and signal-to-noise ratio ρ. In this case, $\mathbb{E}[p \mid \omega] = \rho\omega + (1-\rho)\mathbb{E}[p]$ and

$$x(\omega) = 1 - \rho\omega - (1-\rho)\mathbb{E}[p] \tag{38}$$

We now solve for the price functional. Applying the law of large numbers over realizations of ω, for the continuum of agents $i \in [0,1]$, aggregate demand in state (ξ, p) is the conditional expectation of $x(\omega)$ given (ξ, p). Simplifying,

$$\begin{aligned}
\bar{x}(\xi, p) &= \int x(\omega)\, \phi(\omega \mid z)\, \mathrm{d}\omega \\
&= \int (1 - \rho\omega - (1-\rho)\mathbb{E}[p])\, \phi(\omega \mid z)\, \mathrm{d}\omega \\
&= 1 - \rho p - (1-\rho)\mathbb{E}[p]
\end{aligned} \tag{39}$$

Market clearing requires that $\bar{x}(\xi, p) = \xi$ for all ξ. Substituting in the maintained conjecture that $P = \psi_0 - \psi_1\xi$, we obtain

$$1 - \rho(\psi_0 - \psi_1\xi) - (1-\rho)\psi_0 = \xi\ ,\ \forall\xi \tag{40}$$

Taking expectations of both sides, we get $1-\rho\psi_0-(1-\rho)\psi_0=0$, the unique solution to which is ψ_0. Substituting this back in and conditioning on $\xi \neq 0$, we get $\psi_1\rho\xi=\xi$, the unique solution to which is $\psi_1=1/\rho$. Thus the price functional that clears the market is $P(\xi)=1-\frac{\xi}{\rho}$. Under this observation, we can further simplify individual demand to $x(\omega)=1-\rho\omega-(1-\rho)=\rho(1-\omega)$.

We next formulate agents' reduced-form benefits (equation 4) as a function of the choice variable ρ and the conjectured price $P(\xi)=1-\frac{\xi}{\rho^e}$. As preliminary calculations, we observe that $\mathbb{E}[P(\xi)]=1$, $\mathbb{E}[P(\xi)^2]=\left(\frac{1}{\rho^e}\right)^2+1$, and $\mathbb{E}[x(\omega)]=0$. We calculate

$$
\begin{aligned}
\mathbb{E}[x(\omega)^2] &= \mathbb{E}[\rho^2(1-\omega)^2] \\
&= \rho^2\mathbb{E}[1-2(P(\xi)+a_3\eta_i)+(P(\xi)+a_3\eta_i)^2] \\
&= \rho^2(1-2\mathbb{E}[P(\xi)]+\mathbb{E}[P(\xi)^2]+a_3^2) \\
&= \rho^2\left(\left(\frac{1}{\rho^e}\right)^2+a_3^2\right)=\rho^2\frac{1}{\rho(\rho^e)^2}=\frac{\rho}{(\rho^e)^2}
\end{aligned}
\tag{41}
$$

where the last line uses the definition $\rho=\frac{(\rho^e)^{-2}}{(\rho^e)^{-2}+a_3^2}$. Finally, we calculate covariance terms $\mathbb{E}[x(\omega)P(\xi)]=\rho\mathbb{E}[P(\xi)]-\rho\mathbb{E}[P(\xi)^2]=-\frac{\rho}{(\rho^e)^2}$ and $\mathbb{E}[P(\xi)\xi]=-\frac{1}{\rho^e}$. Using these expressions, we can simplify

$$
\begin{aligned}
b(\rho,\rho^e) &= \max_{\tilde{x}}\ \mathbb{E}\left[\tilde{x}(\omega)-\frac{\tilde{x}(\omega)^2}{2}+P(\xi)(\xi-\tilde{x}(\omega))\right] \\
&= \mathbb{E}[x(\omega)]-\mathbb{E}\left[\frac{x(\omega)^2}{2}\right]+\mathbb{E}[P(\xi)\xi]-\mathbb{E}[x(\omega)P(\xi)] \\
&= 0-\frac{1}{2}\frac{\rho}{(\rho^e)^2}-\frac{1}{\rho^e}+\frac{\rho}{(\rho^e)^2} \\
&= \frac{\rho-2\rho^e}{2(\rho^e)^2}
\end{aligned}
\tag{42}
$$

as claimed in equation 4.

We now return to the signal acquisition problem. Due to the joint Gaussianity of all variables, it is without loss to focus on signals of the form $\omega_i=p+a_3\eta_i$. The signal to noise ratio of this signal, as a function of ρ and $\psi_1=-1/\rho^e$, is

$$
\rho \equiv \frac{a_3^{-2}}{a_3^{-2}+(\rho^e)^2}
\tag{43}
$$

See that, for any $\rho^e \in (0,1]$, we can set $a_3=\left(\frac{\rho}{1-\rho}\right)^{-\frac{1}{2}}\frac{1}{\rho^e}$ to achieve any $\rho\in(0,1)$ as claimed. This establishes the claim that ρ^e is irrelevant for attention costs written as a function of (ρ,ρ^e).

We finally provide a sharper characterization of the equilibrium fixed-point equation

$$
\rho^e \in \arg\max_{\rho}\left\{b(\rho,\rho^e)-c(\rho)\right\}.
\tag{44}
$$

Note that the first-order condition

$$
\frac{\partial}{\partial\rho}b(\rho^e,\rho^e)=c'(\rho^e)
\tag{45}
$$

is necessary and sufficient for equilibrium because $b-c$ is differentiable and strictly concave in $(0,1)$, and the corners $\rho = 0$ or $\rho = 1$ are ruled out by, respectively, $c'(0) = 0$ and $\lim_{\rho \to 1} c'(\rho) = \infty$. Existence is guaranteed by the continuities of $x \mapsto \frac{\partial}{\partial \rho} b(x,x)$ and c', and uniqueness by their monotonicity.

Social Planner's Problem. Let us now consider the social planner's problem, reprinted here:

$$\begin{aligned} \max_{x,a_1,a_2,a_3,M,\pi} \; & \iint \left(x(\omega) - \frac{x(\omega)^2}{2} \right) \phi(\omega|z)\,\pi(z)\,\mathrm{d}\omega\,\mathrm{d}z - C(a_1,a_2,a_3) \\ \text{s.t.} \; & \int x(\omega)\,\phi(\omega \mid \xi, m)\,\mathrm{d}\omega = \xi \text{ for all } (\xi, m) \\ & \pi(\xi,m) = \pi_\Xi(\xi)\cdot \delta_{M(\xi)}(m) \text{ for all } (\xi,m) \end{aligned} \tag{46}$$

To complete the argument in the main text, we derive the social planner's objective conditional on a fixed signal structure of the form $\omega_i = \xi + a_3\eta_i$ and a fixed message.

First, see that the problem conditional on the signal choice and message is globally concave and characterized by first-order conditions. Let $\hat{\lambda}(\xi)$ denote the Lagrange multiplier for the first (continuum of) constraints and $\lambda(\xi) = \frac{\hat{\lambda}(\xi)}{\pi(\xi)}$ denote the Lagrange multiplier divided by the prior. The first-order condition for the choice of $x(\omega)$, for each ω, is

$$1 - x(\omega) = \frac{\int\int \lambda(\xi)\,\phi(\omega|z)\,\pi(z)\,\mathrm{d}z}{\int \phi(\omega|z)\,\pi(z)\,\mathrm{d}z} \tag{47}$$

We can rewrite the right-hand-side more illustratively as $\mathbb{E}[\lambda(\xi) \mid \omega]$, where the expectation is taken over the conditional density associated with the joint density $\phi(\omega \mid z)\pi(z)$. We substitute this expression into the resource constraint to obtain the condition

$$1 - \mathbb{E}[\mathbb{E}[\lambda(\xi) \mid \omega] \mid \xi] = \xi \tag{48}$$

using the simpler conditional expectation notation for the integral in the constraint. The Gaussianity of the right-hand-side requires that the left-hand side is also Gaussian. This pins down that the normalized co-state $\lambda(\xi)$ must be Gaussian, which also restricts it to be linear in ξ.

Let us then represent $\lambda(\xi) = \lambda_0 + \lambda_1\xi$ and solve for the undetermined coefficients. Under the assumed signal structure, we can write out the conditional expectation as

$$1 - \lambda_0 - \lambda_1\rho\xi = \xi \tag{49}$$

Matching coefficients such that the previous holds for all realizations of ξ, we recover $\lambda_0 = 1$ and $\lambda_1 = 1/\rho$.

We now solve for optimal consumption, which conditional on any signal choice, maximizes the Lagrangian

$$\mathscr{L} = \int\int \left(x(\omega) - \frac{x(\omega)^2}{2} \right) \phi(\omega|z)\,\pi(z)\,\mathrm{d}\omega\,\mathrm{d}z - \int (\lambda_0 + \lambda_1\xi) \left(\xi - \int x(\omega)\,\phi(\omega \mid \xi, m)\,\mathrm{d}\omega \right) \pi(z)\,\mathrm{d}z \tag{50}$$

which can be reexpressed using the expectation notation as the following:

$$\mathbb{E}\left[x(\omega)-\frac{x(\omega)^2}{2}+\lambda(\xi-x(\omega))\right] \tag{51}$$

Observe that, comparing with the consumer's program, $\lambda=\lambda(\xi)=P(\xi)$ and $x(\omega)=1-\mathbb{E}[\lambda(\xi)\mid\omega]=1-\mathbb{E}[P(\xi)\mid\omega]$. This is identical to the consumer's program, so the same calculations apply to derive the benefits function modulo the replacement of ρ^e with ρ. From this point, the analysis follows from the analysis in the main text.

C.2 Proof of Proposition 2

By the same arguments in the proof of Proposition 1, the social planner's problem can be reduced to choosing a symmetric attention level (ρ) and the message rule's slope in fundamentals (ψ_1) to maximize

$$W(\rho,\psi_1)=b(\rho,\rho)-c_{PD}(\rho,\rho) \tag{52}$$

where b is defined in equation 4 (and derived in the proof of Proposition 1) and c_{PD} is defined in equation 7. Substituting in both functions, we write

$$W(\rho,\psi_1)=b(\rho,\rho)-c(\rho;\psi_1)=-\frac{1}{2\rho}-\frac{\rho}{1-\rho}\psi_1^{-2} \tag{53}$$

By a similar logic, any equilibrium attention level ρ^e solves the fixed-point equation

$$\rho^e\in\arg\max_{\rho}\left\{\frac{\rho-2\rho^e}{2(\rho^e)^2}-\frac{\rho}{1-\rho}(\rho^e)^2\right\} \tag{54}$$

We first show that equilibrium is inefficient. Let $\rho^e\in(0,1)$ be attention obtained in any equilibrium.[39] A sufficient condition for equilibrium efficiency is that the social planner would not want to send a message with a slope different than $1/\rho^e$ (the slope of prices in fundamentals), or

$$\frac{\partial}{\partial\psi_1}W(\rho,\psi_1)|_{\rho=\rho^e,\psi_1=1/\rho^e}=0 \tag{55}$$

By direct calculation,

$$\frac{\partial}{\partial\psi_1}W(\rho,\psi_1)|_{\rho=\rho^e,\psi_1=1/\rho^e}=-\frac{\partial}{\partial\psi_1}c(\rho;\psi_1)|_{\rho=\rho^e,\psi_1=1/\rho^e}=2\frac{\rho}{1-\rho}\psi_1^{-3}|_{\rho=\rho^e,\psi_1=1/\rho^e}=2\frac{(\rho^e)^4}{1-\rho^e}>0 \tag{56}$$

The fact that this is nonzero proves that equilibrium is inefficient; the fact that it is positive proves that the planner strictly prefers to send a message with a higher slope in the fundamental and, hence, a higher variance compared to the equilibrium, among the class of linear messages.

[39] Observe that $\rho^e=0$ and $\rho^e=1$ would be inconsistent with optimization, due respectively to infinitely negative benefit and infinite cost.

C.3 Proof of Proposition 3

This argument closely follows the structure of the proof of Proposition 2. The social planner's problem is to choose (ρ, ψ_1) to maximize

$$W(\rho, \psi_1) = b(\rho, \rho) - c_{VP}(\rho; \psi_1) = -\frac{1}{2\rho} + \left(1 + \sqrt{\frac{1-\rho}{\rho}}\right)\psi_1^{-2} \tag{57}$$

The planner's first-order incentive to increase ψ_1, evaluated at any equilibrium value ρ^e for attention (and corresponding message slope $1/\rho^e$), is

$$\frac{\partial}{\partial \psi_1} W(\rho, \psi_1)|_{\rho=\rho^e, \psi_1=1/\rho^e} = -2\left(1 + \sqrt{\frac{1-\rho^e}{\rho^e}}\right)(\rho^e)^3 < 0 \tag{58}$$

The last inequality follows given $\rho^e > 0$, a condition that it is straightforward to verify as necessary for individual optimization. The fact that this partial derivative is nonzero proves that equilibrium is inefficient; the fact that it is negative proves that the planner strictly prefers to send a message with a lower slope in the fundamental and, hence, a lower variance compared to the equilibrium, among the class of linear messages.

C.4 Proof of Proposition 4

Case 1: Mutual Information with (ξ, ν, p). We first prove the first part of the statement, which concerns an efficient case of the model in which cognitive costs are given by the mutual information of the agent's signal ω with the entire vector (ξ, ν, p).

We start by characterizing equilibrium. Conjecture that the price has the form $p = P(\xi, \nu) = 1 + \psi_1\xi + \psi_2\nu$ and define $V_\omega = (a_1 + a_2\psi_1)^2 + (a_4 + a_2\psi_2)^2 + a_3^2$ as the variance of the signal. The agent's posterior about the price, conditional on receiving their signal, can be calculated via Bayes' rule:

$$\mathbb{E}[p \mid \omega] = \mathbb{E}[p] + \beta(\omega - \mathbb{E}[\omega]) = 1 + \beta(\omega - a_2) \tag{59}$$

where the signal-to-noise coefficient is

$$\beta \equiv \frac{(a_1 + a_2\psi_1)\psi_1 + (a_4 + a_2\psi_2)\psi_2}{V_\omega} \tag{60}$$

The agents' optimal coconut demand is therefore $x(\omega) = 1 - \mathbb{E}[p \mid \omega] = -\beta(\omega - a_2)$.

We next calculate the consumer's flow utility, excluding information costs, conditional on a given information structure and the conjectured price functional. We first observe that $\mathbb{E}[x(\omega)] = 0$, $\mathbb{E}[x(\omega)^2] = \beta^2 V_\omega$, and $\mathbb{E}[p\xi] = \psi_1$. We next calculate

$$\begin{aligned}
\mathbb{E}[P(\xi,\nu)x(\omega)] &= \mathbb{E}[-\beta(\omega - a_2)(1+\psi_1\xi+\psi_2\nu)] \\
&= -\beta\left(a_1\psi_1\mathbb{E}[\xi^2] + a_4\psi_2\mathbb{E}[\nu^2] + a_2\mathbb{E}[(1+\psi_1\xi+\psi_2\nu)^2] - a_2\right) \\
&= -\beta(a_1\psi_1 + a_4\psi_2 + a_2(\psi_1^2+\psi_2^2)) \\
&= -\beta^2 V_\omega
\end{aligned} \tag{61}$$

Using these calculations, we write

$$\max_x \mathbb{E}\left[x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega))\right] = \frac{\beta^2 V_\omega}{2} + \psi_1 \tag{62}$$

We now restrict attention to signals that set $a_2 = 0$. This is without loss since the price, in equilibrium, is affine in the other fundamentals. We specialize the calculations above to write:

$$\beta = \frac{a_1\psi_1 + a_4\psi_2}{a_1^2+a_4^2+a_3^2} \qquad \beta^2 V_\omega = \frac{(a_1\psi_1+a_4\psi_2)^2}{a_1^2+a_4^2+a_3^3} \tag{63}$$

Therefore,

$$\max_x \mathbb{E}\left[x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega))\right] = \frac{1}{2}\frac{(a_1\psi_1+a_4\psi_2)^2}{a_1^2+a_4^2+a_3^3} + \psi_1 \tag{64}$$

We now consider attention costs. The mutual information of ω with (ξ,ν,p), maintaining $a_2 = 0$, is proportional to $\log\left(\frac{a_3^2+a_1^2+a_4^2}{a_3^2}\right)$. Hence the cost can be written as

$$C(a_1,a_2,a_3,a_4) = K\left(\log\left(\frac{a_3^2+a_1^2+a_4^2}{a_3^2}\right)\right) \tag{65}$$

for an increasing and convex K satisfying $K'(0) = 0$.

Next, we use market clearing to derive restrictions on the attention strategies and price functional. Market clearing requires that

$$X(\xi,\nu) = \mathbb{E}[-\beta(\omega - a_2) \mid \xi,\nu] = -\beta(a_1\xi + a_4\nu) = \xi \quad \forall \xi,\nu \tag{66}$$

where the first equality uses the law of large numbers. First, there cannot exist an equilibrium in which $\beta = 0$, as the market would not clear for some realizations of ξ. Second, given this, there cannot exist an equilibrium in which (almost) all agents choose $a_4 \neq 0$, as the market would not clear for some realizations of (ξ,ν). Under these restrictions, and the conjecture that all agents play an attention strategy with non-zero coefficients (a_1^e, a_3^e), the restriction $-\beta a_1 = 1$ yields

$$\psi_1 = -\frac{(a_1^e)^2 + (a_3^e)^2}{(a_1^e)^2} \tag{67}$$

Having established an expression for the price functional, we can rewrite the benefits expression equation 64 as

$$\max_x \mathbb{E}\left[x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega))\right] = \frac{1}{2}\frac{a_1^2}{a_1^2 + a_3^3}\left(\frac{(a_1^e)^2 + (a_3^e)^3}{(a_1^e)^2}\right)^2 - \left(\frac{(a_1^e)^2 + (a_3^e)^2}{(a_1^e)^2}\right) \tag{68}$$

Substituting in $\rho = \frac{a_1^2}{a_1^2 + a_3^3}$, we write

$$\max_x \mathbb{E}\left[x(\omega) - \frac{x(\omega)^2}{2} + p(\xi - x(\omega))\right] = \frac{1}{2}\frac{\rho}{(\rho^e)^2} - \frac{1}{\rho^e} = \frac{\rho - 2\rho^e}{(\rho^e)^2} \tag{69}$$

Similarly, we observe that the costs are $K(-\log(1-\rho))$. Thus the fixed-point problem describing equilibrium can be expressed as

$$\rho^e \in \arg\max_\rho \left(\frac{\rho - 2\rho^e}{(\rho^e)^2} - K(-\log(1-\rho))\right) \tag{70}$$

which is exactly the fixed-point describing equilibrium in equation 5. Thus, any equilibrium in this economy coincides with an equilibrium to the model studied in Proposition 2.

Let us now consider the social planner's problem. Observe again the irrelevance of the message, since (ξ, ν) is necessarily a sufficient statistic for any $m = M(\xi, \nu)$. Hence it is without loss to consider the same mutual information cost described above by (65). If the co-state variable is $\lambda(\xi) = \lambda_0 + \lambda_1 \xi$, we replicate the last subsection's argument that the planner's optimal consumption plan conditional on the information structure is $x(\omega) = 1 - \mathbb{E}[\lambda \mid \omega]$; and plugging into the constraint establishes that $a_4 = 0$ and $\lambda_1 = -1/\rho = -\frac{a_1^2 + a_3^2}{a_1^2}$. The planner's problem is therefore

$$(a_1^*, a_3^*) \in \arg\max_{a_1, a_3}\left(-\frac{1}{2\rho} - K\left(-\log(1-\rho)\right)\right) \tag{71}$$

which is again the planner's problem from equation 6. Thus, applying Proposition 1, the competitive equilibrium is efficient.

Case 2: Mutual Information with p. We now prove the second part of the Proposition. This concerns a model case in which cognitive costs are given by a transformation of the mutual information of the signal ω with *only* the price p.

We start with equilibrium. We continue to assume agents get a signal of the form $\omega_i = a_1\xi + a_2 p + a_3 \eta_i + a_4 \nu$. We make the normalizations $a_2 = 1$ and $\mathbb{E}[p(\omega_i - p)] = 0$. Both of these are without loss, under the maintained conjecture that p is affine in ξ and ν.[40] We moreover maintain, and later verify, the conjecture that $p = P(\xi, \nu) = 1 + \psi_1 \xi + \psi_2 \nu$

The restriction implied by the restriction $\mathbb{E}[p(\omega_i - p)] = 0$ is that

$$\psi_1 a_1 + \psi_2 a_4 = 0 \tag{72}$$

[40] To see the second, consider a signal $\omega_i = p + a_1\xi + a_4\nu + a_3\eta_i$ such that $\mathbb{E}[p(\omega_i - p)] \neq 0$. In this case, we have $\psi_1 a_1 + \psi_2 a_4 = K \neq 0$. Now consider a new signal $\tilde{\omega}_i = p + \frac{a_1 - \frac{K}{\psi_1^2 + \psi_2^2}\psi_1}{1 + \frac{K}{\psi_1^2 + \psi_2^2}}\xi + \frac{a_2 - \frac{K}{\psi_1^2 + \psi_2^2}\psi_2}{1 + \frac{K}{\psi_1^2 + \psi_2^2}}\nu + \frac{a_3}{1 + \frac{K}{\psi_1^2 + \psi_2^2}}\eta_i$. Observe that $\tilde{\omega}_i = \frac{\omega_i}{1 + \frac{K}{\psi_1^2 + \psi_2^2}}$, so it induces the same posterior beliefs as ω_i, and moreover that $\mathbb{E}[p(\tilde{\omega}_i - p)] = 0$.

Thus, the agent's signal can load a non-zero amount on ξ and ν, conditional on p, and still remain orthogonal to p, given that the covariances with ξ and ν balance out in this way.

We now describe the agent's posterior beliefs, information cost, and coconut demand. Define the signal-to-noise ratio

$$\rho_p = \frac{\psi_1^2 + \psi_2^2}{\psi_1^2 + \psi_2^2 + (a_1^2 + a_3^2 + a_4^2)} \tag{73}$$

The consumer's beliefs are $\mathbb{E}_i[p] = \rho_p \omega_i + (1 - \rho_p)$, their the cognitive cost is $K(-\log(1-\rho_p))$, and their coconut demand is $x_i(\omega) = 1 - \mathbb{E}_i[p] = \rho_p(1-\omega_i)$, by direct analogues of the arguments used in Proposition 1.

The market clearing condition is

$$X(\xi, \nu) = \mathbb{E}[\rho_p(1-\omega_i) \mid \xi, \nu] = -\rho_p((a_1 + \psi_1)\xi + (a_4 + \psi_2)\nu) = \xi, \quad \forall \xi, \nu \tag{74}$$

The requirement that the market clear for all realizations of (ξ, ν) yields the following restrictions:

$$\begin{aligned} a_4 + \psi_2 &= 0 \\ -\rho_p(a_1 + \psi_1) &= 1 \end{aligned} \tag{75}$$

or $a_4 = -\psi_2$ and $a_1 = -\psi_1 - \frac{1}{\rho_p}$.

We now solve for information structures that are consistent with market clearing. We first return to equation 72 and, substituting in the first expression above, write $\psi_1 a_1 - \psi_2^2 = 0$ or $\psi_2^2 = \psi_1 a_1$. Combining this with the expression for a_1 compatible with market clearing, $a_1 = -\psi_1 - \frac{1}{\rho_p}$, we derive the condition

$$\psi_2^2 = -\psi_1(\rho_p^{-1} + \psi_1) \tag{76}$$

Next, we rewrite the signal-to-noise ratio as

$$\rho_p = \frac{\psi_1^2 - \psi_1(\rho_p^{-1} + \psi_1)}{\psi_1^2 - 2\psi_1(\rho_p^{-1} + \psi_1) + (\rho_p^{-1} + \psi_1)^2 + a_3^2} \tag{77}$$

Rearranging this expression gives the condition

$$a_3^2 \rho_p = \psi_1^2 + \psi_2^2 - \frac{1}{\rho_p} \tag{78}$$

Given a value for ρ_p, equations 76 and 78 are a system of two (non-colinear) equations with three unknowns, (ψ_1, ψ_2, a_3^2). These admit a continuum of solutions. In particular, these solutions are indexed by $\psi_1 \in [-\rho_p^{-1}, -1]$ and have

$$\begin{aligned} \psi_2 &= \sqrt{-\psi_1(\rho_p^{-1} + \psi_1)} \\ a_3 &= \rho_p^{-1}\sqrt{-1 - \psi_1} \end{aligned} \tag{79}$$

We next derive the fixed-point equation that describes equilibrium. Plugging into the benefits calculation from equation 62, and using equation 79, we derive that

$$b(\rho_p;\psi_1,\rho_p^e) = -\rho_p \frac{\psi_1}{2\rho_p^e} + \psi_1 \tag{80}$$

in terms of the demand-curve slope and the attention paid by others. Thus any pair (ρ,ψ_1) that solves

$$\rho \in \arg\max_{\rho_p} \left[-\rho_p \frac{\psi_1}{2\rho} + \psi_1 - K(-\log(1-\rho_p)) \right] \qquad \text{and} \qquad \psi_1 \in [-\rho^{-1}, -1] \tag{81}$$

is an equilibrium. Since the maximization is strictly concave for any value of ψ_1 and the cost function satisfies an Inada condition, it suffices to look at the first-order-condition associated with the fixed-point:

$$-\frac{\psi_1}{2\rho} = c'(\rho) \tag{82}$$

where $c'(x) = K'(-\log(1-x))/(1-x)$. Let $\rho(\psi_1)$ denote the unique solution to the fixed-point above for a given ψ_1 and observe that is a decreasing function.

We next characterize the set of equilibria. Let $\rho^e \in (0,1)$ denote the unique equilibrium level of attention from Proposition 1; see that $(\rho^e, -1/\rho^e)$ is a solution to equation 81, as the fixed-point equation is the same one studied in the proof of Proposition 1. Next, consider any $\psi_1 < -1/\rho^e$. We have that $\rho(\psi_1) > \rho(-1/\rho^e)$, and hence $\psi_1 < -1/\rho^e < 1/\rho$. Thus there cannot be an equilibrium in which $\psi_1 < -1/\rho^e$. We next consider any $\psi_1 \in (-1/\rho^e, -1]$. We have that $\rho(\psi_1) < \rho(-1/\rho^e)$, and hence $\psi_1 > -1/\rho^e > -1/\rho(\psi_1)$. Thus each $(\rho(\psi_1),\psi_1)$, for $\psi_1 \in (-1/\rho^e, -1]$, is an equilibrium. In all of these equilibria, equation 79 implies that $\psi_2 \neq 0$, and hence prices are not independent from ν.

Additional Remarks. Having proven both parts of the Proposition, we conclude some additional remarks about the structure of equilibrium multiplicity in Part 2.

Note that the equilibrium of $\psi_1 = -1$ implies that $a_3 = 0$ and $\psi_2 = \sqrt{\rho_p^{-1} - 1}$. In this case, $\omega \propto \xi$ and the agent can obtain the first-best. This works with respect to cognitive constraints because the price is sufficiently contaminated with noise that "precise" observation of ξ corresponds with "imprecise" observation of p. This can be understood also as a cognitive externality whereby changing the dependence of prices on ν affects the cost of obtaining a fixed posterior about the fundamental ξ.

Moreover, the equilibria are Pareto-ranked: welfare (the consumer's *ex ante* utility) is strictly higher in the equilibria with "more public noise," or lower $|\psi_1|$, for a fixed ρ_p. We can thus interpret equilibria with less noise in prices as "cognitive traps," where agents fail to correlate their information/inattention in a welfare-improving manner. This reminds the cognitive traps articulated in Tirole (2015). But whereas that particular form of cognitive traps depends on pecuniary or payoff externalities (equivalently, on some inefficiency in the underlying, attentive economy), ours does not: it originates exclusively in the specification of attention costs and, in particular, on the joint violation of invariance and monotonicity (as established by Theorem 1 and Proposition 9).

Finally, see that any equilibrium in which $\psi_1 \neq 0$ (i.e., p varies even slightly with ξ) is dominated by the social planner's allocation in which $m = M(\xi, \nu) = \nu$, $\omega_i = \xi$, and $x_i = \omega_i$. This yields the first-best allocation without any costs of learning, as the mutual information of ω_i with m is 0. This embodies the most extreme possible exploitation of the endogeneity of the price or message and the associated cognitive externality.

C.5 Proof of Proposition 8

In the statement of the result, we refer to "convex and continuous preferences" and "convex and closed production sets" in the following sense:

Definition 14. Reduced preferences are *convex* if, for every π, for every pair $\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}}' \in \mathscr{X}^{|\Theta|}$ such that $\overline{u}^j(\overline{\boldsymbol{x}}, \pi) < \overline{u}^j(\overline{\boldsymbol{x}}', \pi)$ and every $\alpha \in (0,1)$, we have $\overline{u}^j(\alpha\overline{\boldsymbol{x}} + (1-\alpha)\overline{\boldsymbol{x}}, \pi) > \alpha\overline{u}^j(\overline{\boldsymbol{x}}, \pi) + (1-\alpha)\overline{u}^j(\overline{\boldsymbol{x}}', \pi)$. They are *continuous* if, for every π, $\overline{u}^j(\overline{\boldsymbol{x}}, \pi)$ is continuous for every π.

Definition 15. Reduced production sets are *convex* if, for every π, every pair $\overline{\boldsymbol{y}}, \overline{\boldsymbol{y}}' \in \overline{\boldsymbol{F}}(\pi)$, and every $\alpha \in (0,1)$, $\alpha\overline{\boldsymbol{y}} + (1-\alpha)\overline{\boldsymbol{y}}' \in \overline{\boldsymbol{F}}(\pi)$. They are *closed* if, for every π and every convergent sequence $\{\overline{\boldsymbol{y}}_k\}_{k=1}^{\infty}$ with $\overline{\boldsymbol{y}}_k \in \overline{\boldsymbol{F}}(\pi)$ for all k, $\lim_{k\to\infty} \overline{\boldsymbol{y}}_k \in \overline{\boldsymbol{F}}(\pi)$.

Equilibrium Existence

Since an equilibrium in the twin economy trivially maps to an equilibrium in the original economy, it is sufficient to prove to prove existence in the twin economy.

We first prove existence of equilibrium in an economy in which all preferences are production sets are conditioned on the "message" $M(\theta) = \overline{p}$ for some arbitrary $\overline{p} \in \mathbb{R}^N_+$. Let $\pi_{\overline{p}}$ denote the associated prior. We now map our analysis to the setting of Arrow and Debreu (1954) and verify conditions I-IV in that article when preferences for each type j are represented by $\overline{u}^j(\cdot, \pi_{\overline{p}})$ and the production set is $\overline{\boldsymbol{F}}(\pi_{\overline{p}})$. Condition Ia, which requires closed and convex production sets, is assumed in our setting. Ib and Ic, which restrict pathological outcomes like purely positive production plans, are implied by the assumed monotonicity of H and the normalization for a shut-down production of 0 is possible in any state of the world, at any cognitive cost. Condition II, that the consumption set $\mathscr{X}$ is bounded from above and below, is implied by taking $\mathscr{X}$ is a closed rectangle in the first quadrant. Condition III requires first that the utility function is continuous (IIIa) and convex in the sense of Definition 14.2 (IIIc). As stated it requires also the lack of a satiation point, but it is immediate (and remarked upon by the authors) that this can be replaced by there not existing a satiation point that is consistent with feasibility; and this latter point is implied by our assumption in footnote 9. Assumption IV requires that consumers hold positive claims on the firms, summing to one, and endowments bounded above by some element of $\mathscr{X}$. These are part of our environment.

Theorem I in Arrow and Debreu (1954) guarantees the existence of a competitive equilibrium under the stated assumptions. In particular, this equilibrium is supported by some $P : \Theta \to \mathbb{R}^N_+$.[41]

Observe finally, from Lemma 3.5, that under invariance within G^p, preferences and production sets are unchanged conditional on any π of the form $\pi(z) = \pi_\Theta(\theta) \cdot 1_{f(\theta)}(p)$ for any $f : \Theta \to \mathbb{R}^N_+$. More precisely, we

[41] As written, Theorem I in Arrow and Debreu (1954) guarantees the existence of a vector of state-contingent prices $p \in \mathbb{R}^{N|\Theta|}_+$, but this is readily transformed to the price functional under the maintained assumption that $\pi_\Theta(\theta) > 0$ for all states.

can apply Lemma 1 to show $\overline{u}^j(\cdot,\pi_{\overline{p}}) = \overline{u}^j(\cdot,\pi_P)$, where π_P is the prior induced by the price functional $P(\cdot)$; and $\overline{\boldsymbol{F}}(\pi_{\overline{p}}) = \overline{\boldsymbol{F}}(\pi_P)$. It is trivial then to show that we have obtained an equilibrium of the economy in which preferences for each type j are represented by $\overline{u}^j(\cdot,\pi_P)$ and the production set is $\overline{\boldsymbol{F}}(\pi_P)$.

Second Welfare Theorem

The implementation of an equilibrium with transfers, and the notion of a Pareto optimum, are trivially equivalent between the inattentive economy and its attentive twin. It is therefore sufficient to prove the result in the twin economy.

Consider, then, a Pareto optimum in the twin economy that can be written as $(\overline{x}^j(\theta))_{\theta\in\Theta}$, for each j, and implemented with message $m = M(\theta)$. Let π_M be the induced prior over $\theta \times \mathbb{R}^N_+$. We now use Theorem 2 in Debreu (1954) to verify the existence of a price vector that supports such an equilibrium, when preferences and production sets are conditioned on π_M. To do so we verify conditions (I) to (V) in that article. Condition I is that $\mathscr{X}$ is convex, which is assumed. Condition II is convexity of preferences as stated in the main text. Condition III is a form of continuity. In particular, as written, it asks for every $x, x', x'' \in \mathscr{X}^\Theta$ and agent j, that $\{\alpha : \overline{u}^j((1-\alpha)x' + \alpha x'', \pi) \geq \overline{u}^j(x,\pi)\}$ and $\{\alpha : \overline{u}^j((1-\alpha)x' + \alpha x'', \pi)] \leq \overline{u}^j(x,\pi)\}$ are closed subsets of $[0,1]$.[42] See that this is a trivial consequence of the assumed continuity in the utility function in $\mathscr{X}^\Theta$. Condition IV is the convexity of the production set, which is trivial in the endowment economy. And Condition V is that $\mathscr{X}^\Theta$ is finite dimensional, guaranteed by the finite state space.

Theorem 2 in Debreu (1954) guarantees the existence of a price function $P : \Theta \to \mathbb{R}^N_+$ such that[43]

$$\overline{u}^j((\overline{x}'(\theta))_{\theta\in\Theta}, \pi_M) \geq \overline{u}^j((\overline{x}^j(\theta))_{\theta\in\Theta}, \pi_M) \implies \sum_\theta \pi_\Theta(\theta)\, P(\theta) \cdot \overline{x}'(\theta) \geq \sum_\theta \pi_\Theta(\theta)\, P(\theta) \cdot \overline{x}^j(\theta) \tag{83}$$

This implies that $(\overline{x}^j(\theta))_{\theta\in\Theta}$ solves program 18 when the price functional is given by $P(\cdot)$, the utility function by $\overline{u}^j((\overline{x}'(\theta))_{\theta\in\Theta}, \pi_M)$, and income by $\sum_\theta \pi_\Theta(\theta)\, P(\theta) \cdot \overline{x}^j(\theta)$. See that this equilibrium can be implemented with transfers $T^j(\theta) = P(\theta) \cdot (\overline{x}^j(\theta) - e^j(\theta))$ for each θ.

We finally argue that the given allocation is a price equilibrium with transfers. To do this, we apply invariance (Lemma 1) to show $\overline{u}^j(\cdot,\pi_M) = \hat{u}^j(\cdot,\pi_\Theta) = \overline{u}^j(\cdot,\pi_P)$, where π_P is the prior induced by the price functional $P(\cdot)$; and $\overline{\boldsymbol{F}}(\pi_M) = \hat{\boldsymbol{F}}(\pi_\Theta) = \overline{\boldsymbol{F}}(\pi_P)$. Thus we have shown how to implement the Pareto optimum as a price equilibrium with transfers as desired.

C.6 Proof of Proposition 9

An intermediate result, which we state and prove before the main proof of Proposition 9, shows the consumer's strict preference for, and the firm's ability to produce, bundles that average over states of nature that are irrelevant to preferences, endowments and technologies. Define for each consumer type j the set

[42] This already employs two simplifications relative to the article, using the compactness of $\mathscr{X}$ to use the whole interval $[0,1]$ inclusive of endpoints and the existence of a utility representation of preferences.

[43] As written, the theorem guarantees the existence of a vector of state-contingent prices $\overline{p} \in \mathbb{R}^{N|\Theta|}_+$, but this is readily transformed to the price functional under the maintained assumption that $\pi_\Theta(\theta) > 0$ for all states.

of functions $W^j \subseteq \{w : \Theta \to \Theta\}$ that do not separate any two states corresponding to the same payoffs and do not alter prices. That is, $w(\theta) = w(\theta') \implies u^j(x,\theta) = u^j(x,\theta'),\ \forall x \in \mathscr{X}$. For firms we similarly define the transformations that keep intact the state-dependent component of the feasibility constraint: W^F includes all functions such that $w(\theta) = w(\theta') \implies H(y,c,\theta) = H(y,c,\theta'),\ \forall y \in \mathscr{Y}$. The intermediate result is the following:

Lemma 4. *Let attention costs be invariant and monotone with respect to $\mathscr{G}$. The following properties hold:*

1. *Fix a $(\overline{x}(\theta))_\theta$ and a $w \in W^j$, and construct $(\overline{x}'(\theta))_\theta$ such that*

$$\overline{x}'(\theta) = \frac{\sum_{\theta'} \overline{x}(\theta')\pi_\Theta(\theta')1_\theta(w(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta')1_\theta(w(\theta'))} \tag{84}$$

   for each $\theta \in \Theta$. Then $\overline{u}^j((\overline{x}'(\theta))_\theta, \pi) \geq \overline{u}^j((\overline{x}(\theta))_\theta, \pi)$ with equality if and only if $\overline{x}(\theta) = \overline{x}'(\theta)$ for all θ.

2. *Fix a $(\overline{y}(\theta))_\theta \in \overline{\boldsymbol{F}}(\pi)$ and a $w \in W^F$, and construct $(\overline{y}'(\theta))_\theta$ such that*

$$\overline{y}'(\theta) = \frac{\sum_{\theta'} \overline{y}(\theta')\pi_\Theta(\theta')1_\theta(w(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta')1_\theta(w(\theta'))} \tag{85}$$

   for each $\theta \in \Theta$. Then $(\overline{y}'(\theta))_\theta \in \overline{\boldsymbol{F}}(\pi)$.

Proof. We start with part 1. Let us denote by (x,ϕ) and (x',ϕ') any selection of the solutions to the program (17) corresponding respectively to the parameters $(\overline{x}(\theta))_\theta$ and $((\overline{x}'(\theta))_\theta$. See that $\overline{x}(\theta) \neq \overline{x}'(\theta)$ for at least one $\theta \in \Theta$ implies that $\sum_p \phi(\omega \mid \theta, p) \neq \sum_p \phi(\omega \mid \theta', p)$ for at least some pair some (θ,θ') such that $g(\theta) = g(\theta')$.

Let us now construct a lower bound for $\overline{u}^j((\overline{x}'(\theta))_\theta, \pi)$. To do this, we define a transformation as in Definition 4, for some $g = (\theta, p) \mapsto (w(\theta), \overline{p})$ for some $\overline{p} \in \mathbb{R}^N_+$, which defines ϕ'' and π''. We then define the signal structure function (ϕ''', π) such that, for all ω, z,

$$\phi'''(\omega \mid z) = \phi''(\omega \mid g(z)) \tag{86}$$

We then propose the allocation (x, ϕ''') in program (17) corresponding to the parameters $((\overline{x}'(\theta))_\theta$. See that expected utility for the agent is the same under the proposed allocation and under (x,ϕ):

$$\begin{aligned}
\sum_{\omega,\theta} u^j(x(\omega),\theta)\, \phi(\omega \mid f_\pi(\theta))\pi_\Theta(\theta) &= \sum_{\omega,\theta} \sum_{\theta'} u^j(x(\omega),\theta')\, \phi(\omega \mid f_\pi(\theta'))\, \pi_\Theta(\theta')\, 1_\theta(w(\theta')) \\
&= \sum_{\omega,\theta} u^j(x(\omega),\theta) \sum_{\theta'} \phi(\omega \mid f_\pi(\theta'))\, \pi_\Theta(\theta')\, 1_\theta(w(\theta')) \\
&= \sum_{\omega,\theta} u^j(x(\omega),\theta)\, \phi'''(\omega \mid f_\pi(\theta))\, \pi_\Theta(\theta)
\end{aligned}$$

where the first line rewrites the sum; the second uses the definition of W^j and in particular that $u(\cdot,\theta) = u(\cdot,\theta')$ whenever $w(\theta') = w(\theta)$; and the third substitutes the definition of ϕ'''.

We will next show the cognitive cost is strictly lower under the proposed signal technology than under the technology described by ϕ. See first that $C[\phi'',\pi''] < C[\phi,\pi]$, by monotonicity of the cost function, since

$g \subset \mathcal{G}$; and, second, that $C[\phi''', \pi] = C[\phi'', \pi'']$ by informational invariance, as (86) immediately implies the sufficiency relationship. Therefore, $C[\phi''', \pi] < C[\phi, \pi]$. Putting this together with the previous observation,

$$\sum_{\omega,\theta} u^j(x(\omega),\theta)\, \phi'''(\omega \mid f_{\pi'''}(\theta))\, \pi_\Theta(\theta) - C[\phi''', \pi] > \overline{u}^j((\overline{x}(\theta))_\theta, \pi) \tag{87}$$

Observe now that the allocation (x, ϕ''') is feasible in program (17) with parameter $(\overline{x}'(\theta))_\theta$ by the following direct calculation for each state $\theta \in \Theta$:

$$\begin{aligned}
\sum_\omega x(\omega)\, \phi'''(\omega \mid f_\pi(\theta)) &= \sum_\omega x(\omega)\, \frac{\sum_{\theta'} \phi(\omega \mid f_\pi(\theta')) \cdot \pi_\Theta(\theta') \cdot 1_\theta(w(\theta'))}{\sum_{\theta''} \pi_\Theta(\theta'') \cdot 1_{w(\theta')}(w(\theta''))} \\
&= \sum_{\theta'} \frac{\pi_\Theta(\theta)}{\sum_{\theta''} \pi_\Theta(\theta'') 1_\theta(w(\theta''))} \sum_\omega x(\omega)\, \phi(\omega \mid f_\pi(\theta'))\, \pi_\Theta(\theta') \\
&\leq \sum_{\theta'} \overline{x}(\theta') \frac{\pi_\Theta(\theta') 1_\theta(w(\theta'))}{\sum_{\theta''} \pi_\Theta(\theta'') 1_\theta(w(\theta''))} = \overline{x}'(\theta)
\end{aligned} \tag{88}$$

where the first line uses the definition of ϕ'''; the second rearranges terms; and the third uses the feasibility of the original strategy (x, ϕ) in each state of the world.

Since (x, ϕ'''), is feasible, and $\overline{u}^j((\overline{x}'(\theta))_\theta, \pi)$ is the maximized value of the program, we must have

$$\overline{u}^j((\overline{x}'(\theta))_\theta, \pi) \geq \sum_{\omega,\theta} u^j(x(\omega),\theta)\, \phi'''(\omega \mid f_\pi(\theta))\, \pi_\Theta(\theta) - C[\phi''', \pi] \tag{89}$$

This combined with (87) gives $\overline{u}^j((\overline{x}'(\theta))_\theta, \pi) > \overline{u}^j((\overline{x}(\theta))_\theta, \pi)$ as desired.

We now establish the second part of the result, for firms. Let (y, ϕ) denote the original production plan that exists, satisfies the constraints in (19), and satisfies

$$\sum_\omega y(\omega)\, \phi(\omega \mid f_\pi(\theta)) = \overline{y}(\theta), \ \forall \theta \in \Theta \tag{90}$$

construct ϕ'', ϕ''' and π'' just as above, via the change of variables associated with $g = (\theta, p) \mapsto (w(\theta), \overline{p})$. By an analogue of the same argument used for consumers, $C[\phi''', \pi] < C[\phi, \pi]$. We now consider the production constraints. See that a necessary condition for $\phi'''(\omega \mid f_\pi(\theta)) > 0$ is that $\phi(\omega \mid f_\pi(\theta')) > 0$ for some θ' such that $w(\theta') = w(\theta)$. Feasibility of the original allocation plan implies that

$$H\left(y(\omega), C^F[\phi, \pi], \theta'\right) \leq 0 \tag{91}$$

The definition of W^F implies that $H(y(\omega), c, \theta') = H(y(\omega), c, \theta)$. Moreover, monotonicity of the production possibilities function in its second argument means that $H(y(\omega), c', \theta') < H(y(\omega), c, \theta')$ for any $c' < c$. Putting this together with the strict inequality for costs gives

$$H(y(\omega), C^F[\phi''', \pi], \theta) < H(y(\omega), C^F[\phi, \pi], \theta) = H(y(\omega), C^F[\phi, \pi], \theta') \leq 0 \tag{92}$$

so production satisfies the technological constraint. Finally a calculation identical to (88) shows that

$$\sum_{\omega} y(\omega)\,\phi'''(\omega \mid f_\pi(\theta)) = \frac{\sum_{\theta'} \overline{y}(\theta')\pi_\Theta(\theta')1_\theta(w(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta')1_\theta(w(\theta'))} = \overline{y}'(\theta) \tag{93}$$

for all $\theta \in \Theta$. Therefore $(\overline{y}'(\theta))_{\theta\in\Theta} \in \overline{\boldsymbol{F}}[\pi]$ as desired. □

Proof. We now prove Proposition 9. Throughout the proof, we use the notation $\theta^* = Q(\theta)$, $\theta^j = Q^j(\theta)$, and $\theta^F = Q^F(\theta)$ to describe the constructions of Definition 9 via the respective transformations $Q : \Theta \to \Theta$ and $\{Q^x : \Theta \to \Theta, \forall x \in \{1,\ldots,J\} \cup \{F\}\}$.

The claim that the equilibrium is efficient follows from Theorem 1.

We now prove claim 2, that the equilibrium is price-tracking, by contradiction. Assume otherwise. Imagine first that there is at least one consumer type j such that $Q^j(\theta)$ is not a sufficient statistic for θ. Apply the construction in the proof of Lemma 4 with respect to $w(\theta) = Q^j(\theta)$ to generate a new information structure $(\phi^{j\prime}, \pi')$ and consumer bundle $(x^j, \phi^{j\prime})$.[44] See that, because we have assumed that $Q^j(\theta)$ is not a sufficient statistic for θ in the conditional density, this transformation produces a strict improvement in payoffs. Additionally, under the construction, we have $\phi^{j\prime}(\omega \mid \theta, p) = \phi^{j\prime}(\omega \mid Q(\theta), p)$ which implies that $(Q(\theta), p)$ is sufficient for (θ, p).[45]

See next that the new bundle is feasible. By the following elementary calculation, which uses the law of large numbers, the new bundle has the same cost as the original bundle

$$\begin{aligned}
\sum_{\omega,z}(x^j(\omega)\cdot p)\,\phi^j(\omega \mid \theta, p)\,\pi(\theta,p) &= \sum_p p\cdot\left(\sum_\omega x^j(\omega)\sum_\theta \phi^j(\omega \mid \theta,p)\,\pi(\theta,p)\right)\\
&= \sum_p p\cdot\left(\sum_\omega x^j(\omega)\sum_\theta \phi^{j\prime}(\omega \mid \theta,p)\,\pi(\theta,p)\right)\\
&= \sum_{\omega,z}(x^j(\omega)\cdot p)\,\phi^{j\prime}(\omega \mid \theta,p)\,\pi(\theta,p)
\end{aligned} \tag{94}$$

Therefore the feasibility of the original bundle implies the feasibility of the one. The existence of a feasible bundle with strictly higher payoffs contradicts consumer optimality, as we have found something that is higher payoff and feasible. So the proposed equilibrium cannot exist.

Imagine next that, for the firms, $Q^F(\theta)$ is not a sufficient statistic for θ. Apply the construction in the proof of Lemma 4 with respect to $w(\theta) = Q^F(\theta)$ to generate a new production bundle (y, ϕ'). Observe that, according to (92), there is strict slack in the production constraint or

$$H(y(\omega), C^F[\phi', \pi], \theta) < 0 \tag{95}$$

in all ω, θ such that $\phi'(\omega \mid f_\pi(\theta)) > 0$. Because H is continuous and strictly increasing in its first N arguments,

[44] Specifically, $\phi^{j\prime}(\omega \mid z) = \phi'''(\omega \mid z)$ where the right-hand-side is in the Lemma's terminology.

[45] To translate to Definition 5, we would state this as a property of π' defined by (15) with $g(\theta, p) = (Q(\theta), p)$, with respect to π.

there exists an $\epsilon > 0$ such that

$$H(y(\omega) + \epsilon \boldsymbol{e}, C^F[\phi', \pi], \theta) < 0 \tag{96}$$

where $\boldsymbol{e}$ is an $N \times 1$ vector of ones. Hence (in some abuse of notation) the production plan $(y + \epsilon \boldsymbol{e}, \phi')$ is also feasible.[46] Producing this plan increases profits by $\epsilon(p \cdot \boldsymbol{e}) > 0$. Hence the existence of this deviation contradicts profit maximization. So the proposed equilibrium cannot exist.

We now prove claim 3 of the result, that equilibrium is fundamental. Assume, by contradiction, that there exists a non-fundamental equilibrium. Since invariance and monotonicity with respect to $\mathcal{G}$ implies invariance with respect to G^p, the equilibrium is not Pareto dominated by any allocation supported by an arbitrary message (Theorem 1).

We now show a contradiction to Pareto optimality: the social planner could remove the non-fundamental contingency in the allocation to achieve a Pareto improvement. We start by showing that at least one consumer conditions their aggregate demand or production on the non-fundamental state in the equilibrium. Assume not. The market clearing condition is

$$\sum_{j=1}^{J} \mu^j \overline{x}^j(\theta) = \sum_{j=1}^{J} \mu^j e^j(\theta) + \overline{y}(\theta) \tag{97}$$

for all $\theta \in \Theta$. Take any two states θ, θ' such that $Q(\theta) = Q(\theta')$. Under the conjecture, $\sum_{j=1}^{J} \mu^j \overline{x}^j(\theta) - \sum \mu^j e^j(\theta) = \sum_{j=1}^{J} \mu^j \overline{x}^j(\theta') - \sum \mu^j e^j(\theta')$. But this violates the implication of market clearing that $\overline{y}(\theta) \neq \overline{y}(\theta')$. Therefore at least one consumer type conditions on non-fundamental volatility.

We now propose the following allocation. For each consumer, set the allocation constructed in the proof of Lemma 4 using $w(\theta) = Q(\theta)$, which is by construction in W^j for each type j. For each producer, also set the allocation constructed in the proof of Lemma 4 using $w(\theta) = Q(\theta)$, which is by construction in W^F. Lemma 4 guarantees this allocation is strictly preferred by at least one agent type, whose allocation has changed, and is weakly preferred by all others.

We finally show that it is resource feasible. The feasibility constraint for state θ is

$$\sum_{j=1}^{J} \mu^j \overline{x}^{j\prime}(\theta) \leq \sum_{j=1}^{J} \mu^j e^j(\theta) + \overline{y}'(\theta) \tag{98}$$

which is, using the law of large numbers,

$$\sum_{j=1}^{J} \mu^j \frac{\sum_{\theta'} \overline{x}^j(\theta') \cdot \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))} \leq \sum_{j=1}^{J} \mu^j E^j(\theta) + \frac{\sum_{\theta'} \overline{y}(\theta') \cdot \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))} \tag{99}$$

Next see that, because $e^j(\theta) = e^j(\theta')$ for all $\theta, \theta' : Q(\theta) = Q(\theta')$, we can write

[46] For simplicity, we ignore the issue of $y(\omega) + \epsilon \boldsymbol{e} \notin \mathcal{Y}$ for some ω. To "fix" this, we would apply the exact same reasoning as in the proof of Lemma 2: we can restrict attention away from allocations at a "corner" of $\mathcal{Y}$ by making that set sufficiently large, and otherwise construct our deviation to increase $y(\omega)$ only on the (necessarily positive measure subset) of ω in which it is possible.

$$e^j(\theta) = \frac{\sum_{\theta'} e^j(\theta') \cdot \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))}{\sum_{\theta'} \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))} \tag{100}$$

But we know from the feasibility of the original allocation that, for all θ',

$$\sum_{j=1}^{J} \mu^j \overline{x}^j(\theta') \leq \sum_{j=1}^{J} \mu^j e^j(\theta') + \overline{y}(\theta') \tag{101}$$

Multiplying both sides by $1_{Q(\theta)}(Q(\theta')) \cdot \pi_\Theta(\theta')$, then dividing by $\sum_{\theta'} \pi_\Theta(\theta') \cdot 1_{Q(\theta)}(Q(\theta'))$, gives (99).

Therefore we have shown the existence of a feasible allocation that Pareto dominates the original allocation. This is a contradiction. Therefore there cannot exist a non-fundamental equilibrium. □

C.7 Proof of Corollary 5

We proceed in the following two steps. Consider any possible Pareto-dominating allocation. If the proposed message $M(\theta)$ is invertible in θ, then a twin of the argument from the proof of Theorem 1 shows it must be feasible and payoff-equivalent with the message $(\theta, P(\theta))$ and therefore cannot Pareto dominate the equilibrium. In particular, we can apply a map $M(\theta) \mapsto (\theta, P(\theta))$ and use the full invariance condition to prove an equivalent of Lemma 1.

If the proposed message is not invertible in θ, we claim (and prove at the end of this section) that a payoff-equivalent arrangement is possible with the message $(\theta, P(\theta))$. Therefore, were a Pareto-dominating allocation to exist with message $M(\theta)$, our construction has provided a Pareto-dominating allocation with message $(\theta, P(\theta))$ and contradicted the result of Proposition 5.

We now prove the intermediate step. Denote the consumer choices in the proposed arrangement via x^j, ϕ^j for each type j; and the proposed prior distribution, compatible with the message M, by π_M. Let us now construct the signal structure with likelihood distributions

$$\phi_P^j(\omega \mid \theta, P(\theta)) = \phi^j(\omega \mid f_{\pi_M}(\theta)) \quad \forall \omega \in \Omega \tag{102}$$

and the prior distribution

$$\pi_P(\theta, P(\theta)) = \pi_\Theta(\theta) \tag{103}$$

This new construction in particular "un-coarsens" and relabels the state space, maintaining the assumption that agents do not distinguish between the previously coarsened states of nature . See that (ϕ^j, π_M) is a transformation of (ϕ_P^j, π_P) for a mapping $(\theta, P(\theta)) \mapsto M(\theta)$ which is in $\mathcal{G}$. Moreover, under this mapping, π_M is sufficient with respect to π_P with respect to ϕ_P^j in the sense of Definition 5. Therefore, by invariance, $C[\phi^j, \pi_M] = C[\phi_P^j, \pi_P]$. Next, see that the consumer payoffs from program (11) are trivially the same under both information structures when the consumption strategy is fixed at x^j; and aggregate demands $\overline{x}^j(\theta)$ conditional on any θ are also the same. We then can replicate this argument for each consumer type j.

We next turn to feasibility. We use a similar constructive argument to define a new attention strategy that is valid under π_P and leads to the same costs. The feasibility constraint is similarly unaltered. Thus, with this new arrangement, the planner can generate the same aggregate supply $\overline{y}(\theta)$.

Therefore, we have constructed a new feasible arrangement that results in the same payoffs for each agent in the economy. This ends the proof.

C.8 Proof of Lemma 3

Consider an original information structure (ϕ, π) and its transformation $(\tilde{\phi}, \tilde{\pi})$, by some g. The cost of the original distribution can be written using the definition of a posterior separable cost (33) as $C[\phi,\pi] = \sum_{\omega\in\Omega} \phi_\omega(\omega)\cdot T\left[\phi_{z|\omega};\pi\right] - T\left[\pi;\pi\right]$. Now consider the transformation. We use the invariance of the marginal on ω and the normalization $T[\pi,\pi]=0$ to write $C[\phi,\pi] = \sum_{\omega\in\Omega} \tilde{\phi}_\omega(\omega)\cdot T\left[\phi_{z|\omega};\pi\right]$ and then note that, for each ω, $T\left[\phi_{z|\omega};\pi\right] \le T\left[\tilde{\phi}_{z|\omega};\tilde{\pi}\right]$ with equality if and only if $g(z)$ is a sufficient statistic for z in each posterior distribution. Therefore,

$$C[\phi,\pi] = \sum_{\omega\in\Omega} \tilde{\phi}_\omega(\omega)\cdot T\left[\phi_{z|\omega};\pi\right] \le \sum_{\omega\in\Omega} \tilde{\phi}_\omega(\omega)\cdot T\left[\tilde{\phi}_{z|\omega};\tilde{\pi}\right] = C[\tilde{\phi},\tilde{\pi}] \tag{104}$$

with strict inequality if and only if sufficiency from Definition 5 implies sufficiency from Definition 13. Under this claim, C is monotone and invariant in the sense of Definition 6.

It remains to verify the claim about sufficiency. Using the construction of Definition 6 and Bayes' rule, the posterior distributions are

$$\tilde{\phi}_{z|\omega}(z\mid\omega) = \frac{\sum_{z'\in\mathcal{Z}} \phi_{\omega|z'}(\omega\mid z')\cdot\pi(z')\cdot 1_z(g(z'))}{\tilde{\phi}_\omega(\omega)} = \sum_{z'\in\mathcal{Z}} \frac{\phi_{\omega|z'}(\omega\mid z')\pi(z')}{\tilde{\phi}_\omega(\omega)}\cdot 1_z(g(z')) \tag{105}$$

See that the marginal distribution over ω is unchanged, or $\tilde{\phi}_\omega(\omega) = \phi_\omega(\omega)$. Thus the first term of (105) is $\phi_{z|\omega}(z\mid\omega)$, and so

$$\tilde{\phi}_{z|\omega}(z\mid\omega) = \sum_{z'\in\mathcal{Z}} \phi_{z|\omega}(z\mid\omega)\cdot 1_z(g(z')) \tag{106}$$

This expression and the construction for the prior in Definition 6 therefore both fit the construction of $\tilde{\pi}$ and $\tilde{\pi}'$ in Lemma 3, with $\pi = \phi_{z|\omega}$; $\pi' = \pi$; $\tilde{\pi} = \tilde{\phi}_{z|\omega}$; $\tilde{\pi}' = \tilde{\pi}$. We now show the equivalence of the sufficient statistic expressions. Starting with the expression in Definition 6, $\phi_{\omega|z}(\omega\mid z) = \tilde{\phi}_{\omega|z}(\omega\mid g(z))$ for all ω and z, we use the definition of the construction to write and simplify

$$\phi_{z|\omega}(z\mid\omega) = \frac{\tilde{\phi}_{\omega|z}(\omega\mid g(z))\pi(z)}{\phi_\omega(\omega)} = \frac{\tilde{\phi}_{\omega|z}(\omega\mid g(z))\tilde{\pi}(g(z))}{\tilde{\phi}_\omega(\omega)}\cdot\frac{\pi(z)}{\pi(\tilde{g(z)})} = \tilde{\phi}_{z|\omega}(g(z)\mid\omega)\cdot\frac{\pi(z)}{\tilde{\pi}(g(z))} \tag{107}$$

Set $r(z) = \frac{\pi(z)}{\tilde{\pi}(g(z))}$. See that $\tilde{\pi}(g(z)) \ge \pi(z)$ and therefore $r(z)\in[0,1]$, with the natural extension that $r(z)=0$ if $\tilde{\pi}(g(z))=0$. Moreover it is trivial that $\pi(z) = \tilde{\pi}(g(z))\cdot r(z)$.

D Sufficient Conditions for Convex and Continuous Inattentive Economies

In this appendix, we return to the question of how the assumptions on the reduced preferences and technologies invoked in Proposition 8 translate in terms of primitives. Continuity and closedness follow from the appropriate notions of continuity on the primitives and application of Berge's Theorem. Convexity, on the other hand, is related to the question of whether agents can *costlessly* randomize over consumption and production plans. To use language familiar from general equilibrium theory, the key question is whether the optimal design of a noisy signal can subsume the use of lotteries over bundles of goods. Here, we describe a variation of our model economy in which the such convexity is attained.

Set-up. In contrast to our main analysis with a discrete signal space, in this Appendix we consider a model with signal space $\Omega = [0,1]$. Define $L_{\Omega,\mathscr{X}}(\ell)$ in our context as the set of Lipschitz-continuous functions from Ω to $\mathscr{X}$ with constant ℓ, equipped with the sup norm; and $L_{\Omega,\mathscr{Y}}(\ell)$ the equivalent subset of functions mapping Ω to $\mathscr{Y}$.[47] Next, let $B(\Omega)$ denote the set of all probability measures defined on Ω (equipped with the Borel σ-algebra), and equip $B(\Omega)$ with the standard weak topology.[48] We represent such probability measures by cumulative distribution functions $\Phi \in B(\Omega)$. We redefine the domain of the each agent's cost functional as $(B(\Omega))^{|\Theta|} \times \mathscr{P}$, where $\mathscr{P}$ still carries its definition from Section 3.1. The richness of the state space will assist in proving convexity of preferences, as we will see soon.

We next specialize production by assuming that technology can be written as $H(y,c,\theta) = \tilde{H}(y + c \cdot v, \theta)$, for a vector $v \in \mathbb{R}^N_+$ and a function $\tilde{H} : \mathbb{R}^N \times \Theta \to \mathbb{R}$. This assumption is not trivial but easy to motivate: it nests the special case in which attention is "paid for" in the unit of a specific good (the vector v has a single non-zero element) and therefore enters as an additive penalty in profits denominated in the price of that good. As noted earlier, this is the case invariably assumed in macroeconomic applications. Furthermore, this assumption is the mirror in the production side to a simplifying assumption already made in the consumer side of our environment, namely that attention costs subtract linearly from the expected utility of goods.

We finally assume that attention costs are posterior separable (in the sense of Caplin and Dean, 2015). This assumption is maintained in many macroeconomic applications, including all of those which use mutual information costs.

In this environment, we can show that reduced preferences are convex and continuous and reduced production sets are convex and closed. Thus, Proposition 8 holds as stated regarding the twin economies.

Properties of Twin-Economy Preferences. Let us first consider the (appropriately amended) program that defines reduced preferences:

[47] This restriction is technical and, we conjecture, not particularly restrictive given the richness of the signal space.

[48] Let each measure in $B(\Omega)$ be associated with a cumulative distribution function F. A sequence of measures corresponds to a limit measure if the corresponding sequence of cumulative distribution functions F_n converge at all points of continuity to the corresponding limit measure's cumulative distribution, F.

$$\overline{u}^j((\overline{x}(\theta))_{\theta\in\Theta},\pi) \equiv \max_{x,(\Phi(\cdot|z))_{z\in S[\pi]}} \sum_{\theta}\int_{\Omega} u^j(x(\omega),\theta)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta))\ \pi_\Theta(\theta) - C^j[\phi,\pi]$$
$$\text{s.t. } \sum_{\theta}\int_{\Omega} x(\omega)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta)) \le \overline{x}(\theta),\ \forall\theta\in\Theta \tag{108}$$
$$x \in L_{\Omega,\mathscr{X}}(\ell);\quad \Phi(\cdot \mid f_\pi(\theta)) \in B(\Omega),\ \forall\,\theta\in\Theta$$

We assume now that, for fixed π, that $C^j[\cdot,\pi]$ is continuous in the collection of $(\Phi(\cdot \mid z))_{z\in S[\pi]}$, equipped with the weak topology, where $S[\pi]$ denotes the support of π.[49] Observe that the whole objective function is therefore continuous in the choice variables. Next, see that the constraint is continuous in $\overline{x}(\theta)$, in particular because the integral operation $\int_\Omega x(\omega)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta))$ converges for a combined sequence of bounded functions $x_n(\omega)\to x(\omega)$ and $\Phi_n(\omega \mid z)\to\Phi(\omega\mid z)$ by definition.[50] Finally, note that the respective domains for the goods choice and likelihood choices are compact. We can then apply Berge's theorem to program (108) to assert the existence of a maximum and, moreover, that the value function is continuous in the parameter $(\overline{x}(\theta))_{\theta\in\Theta}$. Observe that a similar argument could be used to establish continuity of preferences in our main, discrete-signal-space case in program (17). As stated in the main text, it is trivial to translate this into the required continuity of preferences.

We now discuss the convexity of the preferences defined in (108). We first demonstrate convexity in the posterior separable case.[51] In particular, say that we can write

$$C[\phi,\pi] = \int_\Omega T\left[\phi_{z|\omega};\pi\right]\mathrm{d}\Phi_\omega(\omega) - T\left[\pi;\pi\right] \tag{109}$$

where $\Phi_\omega(\omega)$ is the marginal CDF over ω and $\phi_{z|\omega}\in\mathscr{P}$ is the posterior distribution over z given ω.

Now imagine we have two bundles $(\overline{x}(\theta))_{\theta\in\Theta}$ and $(\overline{x}'(\theta))_{\theta\in\Theta}$, such that the associated solutions of (108) are (x,ϕ) and (x',ϕ'). Our goal is to show that

$$\overline{u}^j((\alpha\overline{x}(\theta)+(1-\alpha)\overline{x}'(\theta))_{\theta\in\Theta},\pi) \ge \alpha\overline{u}^j((\overline{x}(\theta))_{\theta\in\Theta},\pi)+(1-\alpha)\overline{u}^j((\overline{x}'(\theta))_{\theta\in\Theta},\pi) \tag{110}$$

for any $\alpha\in(0,1)$. We do this by constructing a feasible consumption, in the program with constraints $(\alpha\overline{x}(\theta)+(1-\alpha)\overline{x}'(\theta))_{\theta\in\Theta}$, that achieves the payoff on the left of the previous expression. The plan is an "α lottery" across these two plans. In particular, we set

$$\Phi''(\omega\mid z) = \begin{cases} \alpha\Phi\left(\frac{\omega}{\alpha}\mid z\right) & \text{if } \omega\in[0,\alpha] \\ \alpha+(1-\alpha)\Phi'\left(\frac{\omega-\alpha}{1-\alpha}\mid z\right) & \text{if } \omega\in(\alpha,1] \end{cases} \tag{111}$$

for all z. The continuum state space allows us to accommodate such a lottery. For consumption, similarly

[49] In particular, extend the previous definition to hold for all measures in the (finite) collection.

[50] Fixing the measure Φ_n, this requires use of the dominated convergence theorem; and then across the sequence of measures, this follows from the definition of convergence in the weak topology.

[51] Since our problem is specified in terms of signals and decision rules separately, our argument is much simpler (but also less strong) than the result in Denti (2022) showing that the stochastic choice programs associated with posterior separable cost functionals are convex.

define x'' such that[52]

$$x''(\omega) = \begin{cases} x\left(\frac{\omega}{\alpha}\right) & \text{if } \omega \in [0,\alpha] \\ x'\left(\frac{\omega-\alpha}{1-\alpha}\right) & \text{if } \omega \in (\alpha,1] \end{cases} \tag{112}$$

The combination of ϕ'', x'' replicates the strategy x, ϕ with probability α and the strategy x', ϕ' with probability $1-\alpha$.

First, see that the cost decomposes in the following calculation:

$$\begin{aligned} C[\phi,\pi] &= \int_\Omega T\left[\phi''_{z|\omega};\pi\right] \mathrm{d}\Phi''_\omega(\omega) - T[\pi;\pi] \\ &= \int_0^\alpha T\left[\phi''_{z|\omega};\pi\right] \mathrm{d}\Phi''_\omega(\omega) + \int_\alpha^1 T\left[\phi''_{z|\omega};\pi\right] \mathrm{d}\Phi''_\omega(\omega) - T[\pi;\pi] \\ &= \alpha \int_\Omega T\left[\phi_{z|\omega};\pi\right] \mathrm{d}\Phi_\omega(\omega) + (1-\alpha)\int_\Omega T\left[\phi'_{z|\omega};\pi\right] \mathrm{d}\Phi'_\omega(\omega) - T[\pi;\pi] \\ &= \alpha C[\phi,\pi] + (1-\alpha) C[\phi',\pi] \end{aligned} \tag{113}$$

This leverages the fact that the posterior separable cost is linear in posteriors and hence in linear in the operation of "combining" signals in the current way.[53] Next, see that by almost the same logic the expected utility term is linear:

$$\begin{aligned} \sum_\theta \int_\Omega u^j(x''(\omega),\theta)\ \mathrm{d}\Phi''(\omega \mid f_\pi(\theta))\ \pi(f_\pi(\theta)) = \alpha \sum_\theta \int_\Omega u^j(x(\omega),\theta)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta))\ \pi(f_\pi(\theta)) + \\ (1-\alpha)\sum_\theta \int_\Omega u^j(x'(\omega),\theta)\ \mathrm{d}\Phi'(\omega \mid f_\pi(\theta))\ \pi(f_\pi(\theta)) \end{aligned} \tag{114}$$

and finally that the constraints are linear:

$$\int_\Omega x''(\omega)\ \mathrm{d}\Phi''(\omega \mid f_\pi(\theta)) \le (\alpha\overline{x}(\theta) + (1-\alpha)\overline{x}'(\theta)) \tag{115}$$

This allows us to establish a lower bound for $\overline{u}^j((\alpha\overline{x}(\theta) + (1-\alpha)\overline{x}'(\theta))_{\theta\in\Theta},\pi)$, and therefore show (110) as intended.

Convex and Closed Production Sets. We now describe a similar extension for firms. We will first show that production sets are closed, which like the previous argument of continuous utility functions requires no specific properties of the cost function or state space (and could have been shown in our baseline environment).

To establish convexity, we will require the combination of posterior separability of costs with a particular linearity restriction on H. In particular, for some vector $v \in \mathbb{R}^N_+$, we require the representation

$$H\left(y(\omega), C^F[\phi,\pi],\theta\right) = \tilde{H}\left(y(\omega) + vC^F[\phi,\pi],\theta\right) \tag{116}$$

[52] If this construction results in a Lipschitz discontinuous $x''(\omega)$, it can trivially be amended (alongside the construction of Φ'') to leave a small "gap" between the functions which is interpolated continuously.

[53] An additional feature of posterior separable cost functionals, which we do *not* need to use in this construction to achieve weak convexity, is that costs would be strictly reduced if a positive measure of signals corresponding to the same posteriors (or optimal decisions) were combined.

in which we maintain $\tilde{H}(0,\theta)=0$ for all θ and that H is increasing. We also require the regularity condition that $\tilde{H}(\cdot,\theta)$ is continuous for any value of θ.

See that this formulation can capture the "replacement" of any element y_n of y with $y_n + v_n C^F[\phi,\pi]$ in the relevant part of a production function. This allows attention costs to capture requirements of additional inputs and/or destroyed outputs. In addition, building on the discussion in footnote 9, it allows for attention costs to be specified in terms of a "dummy input" with a normalized price (e.g., of 1), so firms effectively maximize profits net of attention costs.

We first state the firm's problem and establish that production sets are closed. We carry over the regularity assumptions stated above for production plans and feasible distributions. We redefine the feasible production set as

$$\begin{aligned}\overline{\boldsymbol{F}}(\pi) \equiv \Big\{(\overline{y}(\theta))_{\theta\in\Theta} : \exists\ (y(\omega),\phi) \text{ s.t.}& \\ &\int_\Omega y(\omega)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta)) \le \overline{y}(\theta), \forall\theta\in\Theta \\ &\tilde{H}\left(y(\omega)+\vec{v}C^F[\phi,\pi],\theta\right) \le 0, \forall(\omega,\theta)\ \phi \text{ a.e.} \\ &y \in L_{\Omega,\mathscr{Y}}(\ell);\quad \phi(\cdot \mid f_\pi(\theta)) \in B(\Omega),\ \forall\,\theta\in\Theta\Big\}\end{aligned} \tag{117}$$

where "ϕ a.e." denotes that the statement holds for any subset of $\Omega\times\Theta$ that has positive measure under the signal distribution.

We first argue that this set is closed. Assume it is not. Then there exists some point $(\overline{y}(\theta))_{\theta\in\Theta} \notin \overline{\boldsymbol{F}}(\pi)$, and a sequence $(\overline{y}_n(\theta))_{\theta\in\Theta} \to (\overline{y}(\theta))_{\theta\in\Theta}$ (in the Euclidean distance of $\mathbb{R}^{N|\Theta|}$) such that $(\overline{y}_n(\theta))_{\theta\in\Theta} \in \overline{\boldsymbol{F}}(\pi)$ for each n. In particular, for any ϵ, there exists some $(\overline{y}_K(\theta))_{\theta\in\Theta}$ such that $\|(\overline{y}_K(\theta))_{\theta\in\Theta} - (\overline{y}(\theta))_{\theta\in\Theta}\| < \epsilon$. Next, consider the program

$$\begin{aligned}\max_{y,(\Phi(\cdot|z))_{z\in S[\pi]}}\ & -\|(\overline{y}'(\theta))_{\theta\in\Theta} - (\overline{y}(\theta))_{\theta\in\Theta}\| \\ \text{s.t. } &\int_\Omega y(\omega)\ \mathrm{d}\Phi(\omega \mid f_\pi(\theta)) \le \overline{y}'(\theta), \forall\theta\in\Theta \\ &\tilde{H}\left(y(\omega)+\vec{v}C^F[\phi,\pi],\theta\right) \le 0, \forall(\omega,\theta)\ \phi \text{ a.e.} \\ &y \in L_{\Omega,\mathscr{Y}}(\ell);\quad \phi(\cdot \mid f_\pi(\theta)) \in B(\Omega),\ \forall\,\theta\in\Theta\end{aligned} \tag{118}$$

See that all the constraints define a compact set. Boundedness is by assumption, given the spaces in which y and ϕ lie. To see closedness, we first establish in the first constraint that for any sequence $y_n \to y$ and $\Phi_n \to \Phi$, $\overline{y}'_n(\theta) = \int_\Omega y_n(\omega)\ \mathrm{d}\Phi_n(\omega \mid f_\pi(\theta))$ is a convergent sequence (using the definition of the weak topology for probability measures); and if each $\overline{y}'_n(\theta) \le \overline{y}'(\theta)$, then also $\lim_{n\to\infty}\overline{y}'_n(\theta) \le \overline{y}'(\theta)$. Next, for the second constraint, see that for any subset of $\Omega\times\Theta$ which has positive measure under the limit signal distribution represented by Φ, there must exist a subsequence of Φ_n for which this set has a positive measure; index this subsequence by k. Along the subsequence, we argue

$$H\left(y_n(\omega)+\vec{v}C^F[\phi_n,\pi],\theta\right) \to H\left(y(\omega)+\vec{v}C^F[\phi,\pi],\theta\right) \tag{119}$$

using the continuity of H and C^F; and then by a similar argue to the above, using the fact that

$$H\left(y_n(\omega)+\vec{v}C^F[\phi_n,\pi],\theta\right)\leq 0 \tag{120}$$

everywhere along the subsequence, argue that $H\left(y(\omega)+\vec{v}C^F[\phi,\pi],\theta\right)\leq 0$.

Having established points, (118) is a maximization of a continuous function on a compact set and by Weierstrauss' theorem must admit a solution. Denote the maximized value of the program as V. Since we have assumed that $(\overline{y}(\theta))_{\theta\in\Theta}\notin\overline{\boldsymbol{F}}(\pi)$, it must be that the value function V of this program satisfies $V<0$. But then, as argued above using the fact that $(\overline{y}(\theta))_{\theta\in\Theta}\in\mathrm{cl}\overline{\boldsymbol{F}}(\pi)$, there exists some implementable solution of (118) that achieves value $-V/2>V$. This is a contradiction. Therefore the set $\overline{\boldsymbol{F}}(\pi)$ must be closed.

Like with the proof of continuous preferences, see that a much less technical version of the same argument could be used to prove the closedness of $\overline{\boldsymbol{F}}(\pi)$ (and the existence of a solution to our profit maximization problem) when the state space was discrete, as in the main model of Section 3.

We now show that the aggregate production set is convex under the restriction in (110). To do this, we show that for any $(\overline{y}(\theta))_{\theta\in\Theta},(\overline{y}'(\theta))_{\theta\in\Theta}\in\overline{\boldsymbol{F}}(\pi)$, we have also $(\alpha\overline{y}(\theta)+(1-\alpha)\overline{y}'(\theta))_{\theta\in\Theta}\in\overline{\boldsymbol{F}}(\pi)$. We show this with the following construction which mirrors the construction for convex preferences. Let y,ϕ and y',ϕ' be the production plan and attention choice that exist and satisfy the conditions in (117) with respect to constraints $(\overline{y}(\theta))_{\theta\in\Theta},(\overline{y}'(\theta))_{\theta\in\Theta}$.

Let us now construct a variant plan y'',ϕ'' in which attention is given by the α lottery as in (111). For the same argument given in the last subsection, $C^F[\phi'',\pi]=\alpha C^F[\phi,\pi]+(1-\alpha)C^F[\phi',\pi]$ on account of the posterior separability. The production plan is more complex than the analogue with consumer demand and is given by the following:[54]

$$y''(\omega)=\begin{cases} y\left(\frac{\omega}{\alpha}\right)+\vec{v}(1-\alpha)(C^F[\phi,\pi]-C^F[\phi',\pi]) & \text{if } \omega\in[0,\alpha] \\ y'\left(\frac{\omega-\alpha}{1-\alpha}\right)+\vec{v}\alpha(C^F[\phi',\pi]-C^F[\phi,\pi]) & \text{if } \omega\in(\alpha,1] \end{cases} \tag{121}$$

See that, since attention costs are bounded, the domain of $\mathscr{Y}$ can be specified such that $y''(\omega)\in\mathscr{Y}$ without upsetting compactness of $\mathscr{Y}$.

See that this satisfies the capacity constraint as

$$\begin{aligned}\int_\Omega y''(\omega)\,\mathrm{d}\Phi''(\omega\mid f_\pi(\theta)) &= \alpha\int_\Omega (y(\omega)+\vec{v}(1-\alpha)(C^F[\phi,\pi]-C^F[\phi',\pi]))\,\mathrm{d}\Phi(\omega\mid f_\pi(\theta)) \\ &\quad +(1-\alpha)\alpha\int_\Omega (y'(\omega)+\vec{v}\alpha(C^F[\phi',\pi]-C^F[\phi,\pi]))\,\mathrm{d}\Phi'(\omega\mid f_\pi(\theta)) \\ &\leq \pi(\theta)(\alpha\overline{y}(\theta)+(1-\alpha)\overline{y}'(\theta))+\alpha(1-\alpha)\vec{v}(C^F[\phi,\pi]-C^F[\phi',\pi]) \\ &\quad +\alpha(1-\alpha)\vec{v}(C^F[\phi',\pi]-C^F[\phi,\pi]) \\ &= \alpha\overline{y}(\theta)+(1-\alpha)\overline{y}'(\theta)\end{aligned} \tag{122}$$

[54] As stated previously: if this construction results in a Lipschitz discontinuous $y''(\omega)$, it can trivially be amended (alongside the construction of Φ'') to leave a small "gap" between the functions which is interpolated continuously.

Next, we check feasibility. See first that, for $\omega \in [0, \alpha]$,

$$\begin{aligned} y''(\omega) + \vec{v} \cdot C^F[\phi'', \pi] &= y''(\omega) + \vec{v} \cdot (\alpha C[\phi, \pi] + (1-\alpha) C[\phi', \pi]) \\ &= y\left(\frac{\omega}{\alpha}\right) + \vec{v}(1-\alpha)(C^F[\phi, \pi] - C^F[\phi', \pi]) + \vec{v} \cdot (\alpha C[\phi, \pi] + (1-\alpha) C[\phi', \pi]) \\ &= y\left(\frac{\omega}{\alpha}\right) + \vec{v}((\alpha + 1 - \alpha) C^F[\phi, \pi] + (1 - \alpha - (1-\alpha)) C[\phi', \pi]) \\ &= y\left(\frac{\omega}{\alpha}\right) + \vec{v}(C^F[\phi, \pi]) \end{aligned} \tag{123}$$

An essentially identical argument shows $y''(\omega) + \vec{v} \cdot C^F[\phi'', \pi] = y\left(2\omega - \frac{1}{2}\right) + \vec{v}(C^F[\phi', \pi])$ when $\omega \in (\alpha, 1]$. Since we know

$$\begin{aligned} H\left(y(\omega) + \vec{v} C^F[\phi, \pi], \theta\right) \leq 0, \forall (\omega, \theta)\ \phi \text{ a.e.} \\ H\left(y'(\omega) + \vec{v} C^F[\phi', \pi], \theta\right) \leq 0, \forall (\omega, \theta)\ \ \phi' \text{ a.e.} \end{aligned} \tag{124}$$

we therefore know the corresponding feasibility condition for y'', ϕ''. Thus all conditions in (117) are satisfied and $(\alpha \overline{y}(\theta) + (1-\alpha) \overline{y}'(\theta))_{\theta \in \Theta} \in \overline{\boldsymbol{F}}(\pi)$.

E A Dynamic Variant With Vanishing Insurance Over Mistakes

In this Appendix, we consider a variant setting in which transfers contingent on signals ω can be vanishingly small. In particular, we study an economy that can be thought as a T-period replica of that in the main analysis. In this economy, the typical consumer faces again and again the same type of consumption and inattention problem as in the main analysis. To ease exposition, we take the cognition state as $z = \theta$; this effectively restricts attention to economies in which our invariance condition holds and isolates the role of complete markets. Provided that mistakes are uncorrelated over time, their cumulative effect on the consumer's budget vanishes as $T \to \infty$. It follows that the equilibria of our original, static economy can be recast as the equilibria of a twin, dynamic economy in which the insurance over signals is vanishingly small—that is, transfers cease to depend on signals with probability one.

E.1 Set-up

We carry over from the baseline model the definition of the state of nature $\theta \in \Theta$, the N-dimensional goods space $\mathscr{X}$, the (finite) signal space Ω, the agent types $j \in \{1, \ldots, J\}$, the type-dependent endowments represented by $e^j : \Theta \to \mathscr{X}$, the prior distribution $\pi_\Theta \in \Delta(\Theta)$, and the (agents' conjectured) price functional $p : \Theta \to \mathbb{R}^N_+$. We define $\Phi \equiv \Delta(\Omega)^{|\Theta|}$ as the space of likelihood functions and define a cost functional $C : \Phi \to \mathbb{R}_+$, suppressing dependence on the prior. We define $u : \mathscr{X} \times \Theta \to \mathbb{R}$ as the consumer's Bernoulli utility function over goods.

There are T time periods. In each period t, the consumer makes two choices: an attention choice, namely a joint distribution $\phi_t \in \Phi$ between the period-t signal ω_t and the underlying state of nature θ; and a consumption choice, namely a mapping x_t from the realized signal to a consumption bundle. The payoff of agent type j is given by

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{\omega_t,\theta} u^j\left(x_t(\omega_t),\theta\right)\phi_t(\omega_t\mid\theta)\,\pi(\theta) - C^j\left(\phi_t\right)\right) \tag{125}$$

That is, their payoffs are additively separable across time, and their per-period payoffs are the same as in our main model. Also, for simplicity, we assume convexity of preferences in the sense of Definition 14. Their budget constraint is given by

$$\sum_{\theta}\sum_{t=1}^{T}\sum_{\omega_t} P_t(\theta)x_t(\omega_t)\phi_t(\omega_t\mid\theta)\pi(\theta) \le \sum_{\theta}\sum_{t=1}^{T}\sum_{\omega_t} P_t(\theta)e^j(\theta)\phi_t(\omega_t\mid\theta)\pi(\theta) \tag{126}$$

where $(P_t(\theta))_{t=1}^T$ is a sequence of prices and $e^j(\theta)$ is a per-period endowment for each state $\theta\in\Theta$. Note that, similarly to the main analysis, the budget constraint embeds complete markets not only vis-a-vis the state of nature θ but also vis-a-vis the realizations of the signals ω. That is, the consumer enjoys insurance against the random mistakes caused by inattention. The question of interest is how "large" this insurance is.

E.2 Result: Vanishing Signal-Contingent Transfers as $T\to\infty$

To address this question, consider first the $T=1$ economy. Clearly, this is the same economy as that studied in the main analysis. Let $(p^*, (x^{j*},\phi^{j*})_{j=1}^J)$ denote any equilibrium of that economy.

Next, for any $T>1$, define the collection $\{P_t^*, (\phi_t^{j*}, x_t^{j*})_{j=1}^J\}_{t=1}^T$ by letting $P_t^*(\theta) = P^*(\theta)$, $\phi_t^{j*}(\omega,\theta) = \phi^{j*}(\omega,\theta)$, and $x_t^{j*}(\omega) = x^{j*}(\omega)$ for all t, θ, ω, and j. Note that this collection constitutes an equilibrium of the T-period economy. This is due to the separability of *both* consumption and attention choices across time; the time-independence of the corresponding per-period preferences, attention costs, and endowments; and the convexity of preferences.[55]

Finally, let

$$b(\omega^T,\theta) \equiv \sum_{t=1}^{T} P_t(\theta)\left[x_t(\omega_t) - e_t(\theta)\right]$$

measure the realized excess lifetime spending of the consumer, conditional on the state of nature is θ and her cumulative cognitive history $\omega^T = (\omega_t)_{t=1}^T$. The consumer's budget constraint can then be rewritten as

$$\sum_{\theta}\sum_{\omega^T} b(\omega^T,\theta)\phi(\omega^T\mid\theta)\pi(\theta) \le 0.$$

where $\phi(\omega^T\mid\theta) = \prod_{\omega_t\in\omega^T}\phi(\omega_t\mid\theta)$. This underscores the presence of insurance vis-a-vis inattention: for any given θ, the consumer is allowed to overspend in some cognitive histories, namely $b(\omega^T,\theta)>0$, provided that she under-spends in other histories. But now note that, along the aforementioned equilibrium, we have

[55] Under convex preferences, were two different consumption and attention strategies optimal in different periods, there would be a contradiction to optimality—a convex combination of these strategies (both consumption and attention) would lead to strictly higher payoffs.

that

$$b(\omega^T, \theta) = \sum_{t=1}^{T} p^*(\theta) \left[x^*(\omega_t) - e(\theta) \right]$$

and that ω_t is i.i.d. across time. By the Strong Law of Large Numbers, we then have that

$$\lim_{T \to \infty} b(\omega^T, \theta) = \bar{b}(\theta), \text{ almost surely, } \forall \theta$$

where $\bar{b}(\theta) \equiv P^*(\theta)\,[\bar{x}^*(\theta) - e(\theta)]$, $\bar{x}^*(\theta) \equiv \mathbb{E}[x^*(\omega)|\theta] = \sum_\omega x^*(\omega)\phi^*(\omega|\theta)$, and "almost surely" is with respect to the distribution of the ω_t process conditional on θ. That is, as the consumer's lifespan gets longer and longer, the transfers $b(\omega^T, \theta)$ cease to depend on the cognitive history ω^T and instead depend only on the state of nature θ. In this sense, insurance takes place only across θ, and the model implies vanishingly small cross-agent, ω-contingent transfers.

This proves the claim made in the beginning. To further illustrate the claim, we specialize to the case in which there is a single type of agent (and drop the superscript j). By market clearing, for all θ, $\bar{x}^*(\theta) = e(\theta)$ and $\bar{b}(\theta) = 0$. We conclude that

$$\lim_{T \to \infty} b(\omega^T, \theta) = 0, \text{ almost surely, } \forall \theta$$

That is, once we remove the scope for insurance across fundamental states θ, we have that transfers are effectively zero across all realizations of uncertainty.

One can then visualize this as follows. In each period, the consumer "does her best" to equate her expenditure to her income, because this is the optimal thing to in the absence of inattention. Inattention may cause the consumer to overspend some times and under-spend other times. Such "mistakes" are absorbed by adjusting the consumer's balance in a checking account, or a credit card. For finite T, this requires that the consumer die with a non-zero balance in her account, and complete markets makes this feasible. But for $T \to \infty$, the consumer's account zeroes out by itself, and the transfers needed to replicate our complete-markets outcome vanish.

To be clear, however, this does not mean that the incomplete markets outcomes converge to the complete markets outcomes as $T \to \infty$. That is, the absence of even small state-contingent transfers could have discontinuous effects on behavior. For example, agents' precautionary motivates to avoid certain low-payoff outcomes (e.g., zero consumption of a specific good) could induce strong precautionary behaviors.

F An Economy with No Insurance Over Mistakes

In this Appendix, we sketch a variant of our model which disallows complete markets over the noise in the agent's signal ω. In the previous Appendix (E), we showed that insurance over signals could be vanishingly small; but, nevertheless, the existence of such insurance was essential to consumers' ability to satisfy their budget, despite the fact that none of their consumption choices could be perfectly measurable in both the signal and fundamental. Now, we want to consider the case in which no insurance is allowed. This raises a challenge—if we were to maintain the assumption that none of their consumption choices could be perfectly measurable in both the signal and fundamental, then it would be impossible to meet budget constraints at equality for all possible states. Following this route would have therefore required the introduction of default, which would take the analysis well outside the Arrow-Debreu framework. Instead, we bypass this problem and ensure that budget constraints are satisfied despite incomplete markets by allowing one of the consumption choices to be fully contingent on both signals and realized states.

If preferences are linear in this residual good, the absence of insurance is inconsequential and the only source of inefficiency remains the potential for cognitive externalities. If instead preferences are not linear, then the absence of insurance leads to different marginal values of wealth conditional on different realizations of the signals and states. Constrained inefficiency, defined as in Geanakoplos and Polemarchakis (1986), could arise for essentially the same reasons that these authors highlight: a social planner can manipulate prices to partially simulate the missing insurance.

F.1 Set-up and Equilibrium

For the example, we restrict the environment in a number of ways to simplify analysis. First, we study a single-consumer-type ($J = 1$) endowment economy. Second, we assume throughout that a "first-order-condition" approach is necessary and sufficient to characterize the optimum for consumers and the social planner. This presumes differentiability and concavity of the utility function. Finally, we allow the good N to be an "adjustment good" which has no (binding) domain constraint and automatically adjusts to meet the budget constraint of the agent given each realization of the signal ω and the physical state θ. Finally, to ease exposition, we take the cognition state as $z = \theta$, as in Online Appendix E. This effectively restricts attention to economies in which our invariance condition holds and isolates the role of (in)complete markets.

We depart from our conventional notation in the following way. We write the consumption vector as $x = (x_{-N}, x_N)$, where x_{-N} is an $N-1$ length vector of the other goods' consumption and x_N is a scalar representing adjustment-good consumption. We similarly write the utility function as $u(x_{-N}, x_N, \theta)$. We write the price vector as $P(\theta) = (P_{-N}(\theta), P_N(\theta))$, where $P_{-N}(\theta)$ is an $N-1$ length vector of prices for "standard" goods and $P_N(\theta)$ is a scalar corresponding to the price of the adjustment good. Finally we let the endowment be written as $e(\theta) = (e_{-N}(\theta), e_N(\theta))$ and the consumption domain as $\mathscr{X} = (\mathscr{X}_{-N}, \mathbb{R})$ with a similar interpretation. Note that the adjustment good has an unrestricted domain and can be consumed in negative amounts.

We make the additional assumption, as in Section 7.1 ("Allowing for Incomplete Markets") and equation 25, that agents have access to complete markets only over the fundamental state θ. Agents have no access to transfers conditional on the signal realizations ω. We capture this by giving agents access to transfers represented by $b : \Theta \to \mathbb{R}$, which must net out in expectation over states θ.

We begin by characterizing necessary conditions for equilibrium. The consumer consumption, θ-contingent transfers, and a signal structure. Their optimal choices solve the following problem:

$$
\begin{aligned}
\max_{x_{-N},x_N,b,\phi} & \sum_\theta \sum_\omega u\left(x_{-N}(\omega), x_N(\omega,\theta), \theta\right) \phi(\omega|\theta)\pi(\theta) - C[\phi,\pi] \\
\text{s.t. } & P_{-N}(\theta)\cdot(e_{-N}(\theta) - x_{-N}(\omega)) + P_N(\theta)(e_N(\theta) - x_N(\omega,\theta)) \le b(\theta) \qquad \forall \omega,\theta \\
& \sum_\theta b(\theta)\pi(\theta) \le 0
\end{aligned}
\tag{127}
$$

with $x_{-N}: \Omega \to \mathscr{X}_{-N}$, $x_N: \Omega\times\Theta \to \mathscr{X}_N$, $b: \Theta \to \mathbb{R}$, and $\phi \in \Phi$. The combination of these constraints capture how the consumer can freely transfer money across θ but not ω.

Equilibrium is characterized by choices that solve (127) taking as given a price functional $P: \Theta \to \mathbb{R}^N_+$ and a prior consistent with that price functional, as well as the market clearing condition

$$
\sum_\omega x_{-N}(\omega)\phi(\omega|\theta) = e_{-N}(\theta) \qquad \forall \theta \tag{128}
$$

The market clearing condition for money is redundant from Walras' law.

We now provide a set of necessary conditions for equilibrium from consumer optimization. Let the Lagrange multipliers for the first and the second constraint of (127) be, respectively, $\chi(\omega,\theta)\phi(\omega|\theta)\pi(\theta)$ and η. The first-order-conditions characterizing optimal the optimal choice of money consumption $x_N(\omega,\theta)$ is

$$
\frac{\partial}{\partial x_N} u\left(x_{-N}(\omega), x_N(\omega,\theta), \theta\right) = P_N(\theta)\chi(\omega,\theta) \qquad \forall\omega,\theta \tag{129}
$$

In words, the marginal value of wealth $\chi(\omega,\theta)$ must equal the marginal utility from consuming money deflated by the price of the adjustment good, in each state $\chi(\omega,\theta)$. Note that, with no restrictions over the domain of x_N and monotonicity in preferences, another implication is that consumption of good N will be a "residual" that makes the budget constraint hold at equality:

$$
x_N(\omega,\theta) = \frac{1}{p_N(\theta)}\Big(b(\theta) - P_{-N}(\theta)\cdot(e_{-N}(\theta) - x_{-N}(\omega))\Big) \qquad \forall\omega,\theta \tag{130}
$$

The first-order conditions for choosing consumption of other goods is

$$
\sum_\theta \left[\frac{\partial}{\partial x_{-N}} u\left(x_{-N}(\omega), x_N(\omega,\theta), \theta\right) - P_{-N}(\theta)\chi(\omega,\theta)\right]\phi(\theta|\omega) = 0 \qquad \forall\omega \tag{131}
$$

where $\phi(\theta \mid \omega)$ is constructed in the standard way via Bayes' rule.

Finally, the first-order condition for choosing money balances is

$$
\sum_\omega \chi(\omega,\theta)\phi(\omega|\theta) = \eta \qquad \forall\theta \tag{132}
$$

In words, insurance over fundamental states allows agents to have the same *average* marginal value of wealth in different realizations θ and θ'. But this does not provide full insurance over signals ω—marginal utility may vary over ω conditional on θ, and how it does so will reflect the exact nature of (optimal) inattention.

Substituting in the marginal value of wealth, equal to the marginal consumption value of money, one derives the following two necessary conditions for consumer optimality and hence necessary conditions for equilibrium:

$$\begin{aligned}\sum_\theta \left[\frac{P_{-N}(\theta)}{P_N(\theta)}\frac{\partial}{\partial x_N}u\left(x_{-N}(\omega), x_N(\omega,\theta),\theta\right)\right]\phi(\theta|\omega) &= \sum_\theta \left[\frac{\partial}{\partial x_{-N}}u\left(x_{-N}(\omega), x_N(\omega,\theta),\theta\right)\right]\phi(\theta|\omega), \quad \forall\omega \\ \eta &= \frac{1}{P_N(\theta)}\sum_\omega \frac{\partial}{\partial x_N}u(x_{-N}(\omega), x_N(\omega,\theta),\theta)\phi(\omega|\theta) \quad \forall\theta\end{aligned} \tag{133}$$

F.2 Efficient Allocations

We now define the problem of a social planner who faces the same "asset-spanning" restriction. Loosely speaking, the planner solves the consumer's program (127) but with control over prices. More specifically, they choose the tuple $(x_{-N}, x_N, b, P, \phi)$ to maximize the following expected utility objective

$$\sum_\theta\sum_\omega u\left(x_{-N}(\omega), x_N(\omega,\theta)\right)\phi(\omega|\theta)\pi(\theta) - C[\phi,\pi] \tag{134}$$

subject to the following feasibility constraints. First, consumption of all goods, including money, is physically feasible in each state θ:

$$\begin{aligned}\sum_\omega x_{-N}(\omega)\phi(\omega \mid \theta) &\le e_{-N}(\theta) \\ \sum_\omega x_N(\omega,\theta)\phi(\omega \mid \theta) &\le e_N(\theta)\end{aligned} \tag{135}$$

Second, each agent has a budget constraint:

$$P_N(\theta)(e_N(\theta) - x_N(\omega,\theta)) + P_{-N}(\theta)\cdot(e_{-N}(\theta) - x_{-N}(\omega)) \le b(\theta) \qquad \forall\omega,\theta$$

Third, transfers must net out in expectation: $\sum_\theta b(\theta)\pi(\theta) \le 0$. Finally, the prior π is consistent with the price functional, or $f_\pi(\theta) = (\theta, P(\theta))$ for each θ. We define an efficient allocation as one that solves the above problem, which is without loss in our economy with one type and symmetric choices.

Observe that this planner's problem is not perfectly parallel with the one in our main analysis, as the prices or messages have an instrumental role in transferring resources across states. We will show that even this "weaker" planner, with a tighter implementability constraint, can often improve upon equilibrium allocations.

We now use a first-order approach to derive necessary conditions for an efficient allocation, again presuming such an allocation exists and it is characterized by first-order conditions. Let the Lagrange multipliers on the three constraints above respectively be $\lambda(\theta)\pi(\theta)$ (x_{-N} feasibility), $\alpha(\theta)\pi(\theta)$ (x_N feasibility), $\tau(\omega,\theta)\phi(\omega|\theta)\pi(\theta)$ (budgets), and γ (portfolio choice). The first-order conditions for x_N, x_{-N}, and b are, respectively,

$$
\begin{aligned}
0 &= \mu(\theta) + P_N(\theta)\tau(\omega,\theta) - \left(\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta)\right) \quad \forall \omega,\theta \\
0 &= \sum_\theta \left[\frac{\partial}{\partial x_{-N}} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) - \lambda(\theta) - P_{-N}(\theta)\tau(\omega,\theta)\right]\phi(\theta|\omega) \quad \forall \omega \\
\gamma &= \sum_\omega \tau(\omega,\theta)\phi(\omega|\theta) \quad \forall \theta
\end{aligned}
\tag{136}
$$

See that these conditions resemble the ones in the consumer problem up to the introduction of shadow values $\mu(\theta)$ and $\lambda(\theta)$ for money and non-money consumption goods, respectively. Next, the first-order condition for the goods prices $P_{-N}(\theta)$, which show up only in the residual spending constraint, is[56]

$$
\sum_\omega \tau(\omega,\theta)\left[e_{-N}(\theta) - x_{-N}(\omega)\right]\phi(\omega|\theta) = 0 \qquad \forall \theta \tag{137}
$$

This condition bears special comment, as it is drives the key wedge between efficiency and equilibrium. By adjusting the price vector $P_{-N}(\theta)$ in any state θ, the planner affects marginal money consumption conditional on each (ω,θ) in proportion to the net endowment $e_{-N}(\theta) - x_{-N}(\omega)$. This has marginal value $\tau(\omega,\theta)$ to the consumer. Such adjustments are exactly what the invisible hand will not do in our environment.

We now develop the above argument mathematically. Like in the consumer's problem, we solve out for $\tau(\omega,\theta)$ and rewrite the first-order conditions of (136) in the following way:

$$
\begin{aligned}
\sum_\theta \left[\frac{\partial}{\partial x_{-N}} u(x_{-N}(\omega), x_N(\omega,\theta),\theta)\right]\phi(\theta|\omega) &= \sum_\theta \left[\frac{P_{-N}(\theta)}{P_N(\theta)}\left(\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) - \mu(\theta)\right) + \lambda(\theta)\right]\phi(\theta|\omega) \\
\gamma + \frac{\mu(\theta)}{P_N(\theta)} &= \frac{1}{P_N(\theta)} \sum_\omega \frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta)\phi(\omega|\theta)
\end{aligned}
\tag{138}
$$

which can be directly compared with the consumer's equilibrium first-order-conditions (133). See that these correspond exactly to the first-order conditions of the consumer's problem if $\mu(\theta) = P_N(\theta)\eta$ and $\lambda(\theta) = P_{-N}(\theta)\eta$, or marginal social values of goods equal their prices (up to scale), and $\gamma = 0$. Note that the last normalization is harmless since, with both resource and budget constraints, the asset constraint was redundant.

The condition for choosing prices $P_{-N}(\theta)$ is

$$
\sum_\omega \frac{1}{P_N(\theta)}\left[\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) - \mu(\theta)\right]\left[e_{-N}(\theta) - x_{-N}(\omega)\right]\phi(\omega|\theta) = 0 \qquad \forall \theta \tag{139}
$$

where, in continuation of the discussion above of using prices for redistribution, $\frac{1}{P_N(\theta)}\left(\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) - \mu(\theta)\right)$ gives the marginal consumption value of the adjustment good (net of the social cost), and the planner changes prices until the average effect on payoffs via the adjustment-good consumption is zero. See that this condition is generically incompatible with the previous conjecture for μ when evaluated at the equilib-

[56] The *dollar* prices $P_N(\theta)$ show up only in the prior and the cognitive cost, and therefore function like the purely informational messages of the main analysis.

rium consumption levels, as there is no analogue for this condition in the equilibrium conditions. This is the sense in which one should not "expect" that an efficient allocation, which necessarily satisfies (139), should also correspond to an equilibrium allocation.[57] Note that the argument makes no reference to the structure of C apart from assuming there is *some* randomness over signal realizations. Thus the logic above applies in a number of simplified settings including exogenously-incomplete-information economies, which underscores our general point that the pathway of inefficiency via incomplete insurance is not a direct consequence of learning or rational inattention.

An important exception is a quasilinear case, in which $\frac{\partial}{\partial x_N} u$ is constant as a function of x_{-N} and x_N (but might depend on θ). In particular, if we can write $\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) = f(\theta)$ for some function f, then the planner's second first-order condition in (138) reduces to $\mu(\theta) = f(\theta)$ and the condition for choosing prices simplifies via

$$
\begin{aligned}
0 &= \sum_{\omega} \frac{1}{P_N(\theta)} \left[\frac{\partial}{\partial x_N} u(x_{-N}(\omega), x_N(\omega,\theta),\theta) - \mu(\theta) \right] [e_{-N}(\theta) - x_{-N}(\omega)]\, \phi(\omega|\theta) \\
&= \sum_{\omega} \frac{1}{P_N(\theta)} \left[f(\theta) - \mu(\theta) \right] [e_{-N}(\theta) - x_{-N}(\omega)]\, \phi(\omega|\theta) \\
&= \sum_{\omega} [0]\, [e_{-N}(\theta) - x_{-N}(\omega)]\, \phi(\omega|\theta) = 0
\end{aligned}
$$

This represents mathematically the logic that, with quasilinear utility (accommodating even θ-dependence in the "slope"), there is no first-order benefit to distributing goods across realizations of ω. And it verifies that the precise role of the quasilinear utility assumption in the example of Section 2 was to substitute for our main model's assumption of insurance over ω.

G Economies with a Broader Cognition State

In this Appendix, we describe how to model economies in which the cognition state includes variables other than the state of nature and prices. We then sketch how an appropriately extended notion of invariance delivers a straightforward extension of Theorem 1 and Corollary 5.

G.1 Environment

We first define an expanded notion of an inattentive market economy, in which the cognitive process (as captured by the definition of z) takes as inputs not only on the exogenous state of nature and the price vector but also the following additional objects: *transfers*, *goods taxes*, *aggregate trades*, and *exogenous signals* (or, more informally, "media").

Equilibrium is defined as in Definition 1 modulo the following changes. As in Section 6.1, we allow each consumer's wealth to include a state-dependent transfer, given by $t^j = T^j(\theta)$, for some exogenously specified

[57] A more formal argument of "generic non-efficiency," as pursued in Geanakoplos and Polemarchakis (1986), would show that even if necessary condition (139) were satisfied evaluated at the equilibrium, then in a small perturbation of the economic environment (appropriately defined) it could be made not to hold.

and type-specific transfer rule $T^j : \Theta \to \mathbb{R}$. We next have consumers and firms take as given after-tax prices $p+\tau$, where the goods taxes are state-dependent, too, or $\tau = T(\theta)$ for some exogenously specified tax rule $T : \Theta \to \mathbb{R}$ The collection $(T, (T^j)_{j=1}^F)$ is restricted to be such that the government budget balances in each state of nature, or $\sum_{j=1}^J (T^j(\theta) + T(\theta) \cdot \overline{x}^j(\theta)) + T(\theta) \cdot \overline{y}(\theta) = 0$ for all $\theta \in \Theta$. Next, we incorporate as part of the equilibrium a set of exogenous signals $s \in \mathbb{R}^Q$, for some $Q > 0$, defined by the exogenous mapping $s = S(\theta)$ for some $S : \Theta \to \mathbb{R}$. And finally, we allow the cognition state in equilibrium to be

$$z = \left(\theta, p, \tau, (t^j)_{j=1}^J, (\overline{x}^j)_{j=1}^J, s\right)$$

We write the domain of this object as $z \in \Theta \times \mathscr{B}_0$ where

$$\mathscr{B}_0 = \mathbb{R}_+^N \times \mathbb{R}^N \times \mathbb{R}^J \times \mathscr{X}^J \times \mathbb{R}^Q \tag{140}$$

is the composition of all the domains for the objects after the state of nature. We define a $\mathscr{B} \supseteq \mathscr{B}_0$ as an even larger possible domain for this state, which the social planner may take advantage of. We correspondingly redefine the admissible set of priors over z corresponding to any particular $B \subseteq \mathscr{B}$, including $B = \mathscr{B}_0$, as

$$\mathscr{P}_B \equiv \left\{\pi : \Theta \times \mathbb{R}_+^N \to [0,1] \text{ s.t. } \pi(\theta, z) = \pi_\Theta(\theta)\, 1_{f(\theta)}(z), \text{ for some } f : \Theta \to B\right\}. \tag{141}$$

Along the same lines, we also redefine, for every $\pi \in \mathscr{P}$ and domain B, the set $Z_{\pi,B} \equiv \{(\theta, b) : \pi(\theta, b) > 0\} \subset \Theta \times B$ and the function $f_{\pi,B} : \theta \to \Theta \times b$ with $f_{\pi,B}(\theta) = \{(\theta, b) : \pi(\theta, b) > 0\} \in \Theta \times B$ for any $\theta \in \Theta$.

Toward our efficiency concept, we start with the feasibility concept in 2 along with the following expanded notion of messages. The definition of an arrangement now includes an image set $B \subseteq \mathscr{B}$, the message is sent via a rule $M : \Theta \to B$.

G.2 Extending Invariance and Monotonicity

For the arguments above to be well-defined, we require an expanded notion of the cost functional. We assume, in particular, that for each $B \subseteq \mathscr{B}$, each agent (and the firm) has a well-defined cost indexed by B:

$$C_B^j : (\Delta(\Omega))^{|\Theta|} \times \mathscr{P}_B \tag{142}$$

maintaining, as in the main text, the assumption that the support of the fundamentals distribution can be ordered.

In this context, we redefine our notions of transformations and invariance. Let $\mathscr{H} \equiv \{h : (\Theta \times B) \to (\Theta \times B'); B, B' \subseteq \mathscr{B}\}$ be the dictionary of transformations from any $\Theta \times B$ to any $\Theta \times B'$. We next define transformations of information structures based on this larger set of functions:

Definition 16 (Transformations of Information Structures, Revisited)**.** Consider two information structures (π, ϕ) and $(\tilde{\pi}, \tilde{\phi})$ and a function $h \in \mathscr{H}$ mapping $\Theta \times B$ to $\Theta \times B'$. We say that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of

(ϕ, π) under h if

$$\tilde{\pi}(z) = \sum_{z'} \pi(z') 1_z(h(z')) \qquad \forall z \in Z_{\pi,B} \tag{143}$$

$$\tilde{\phi}(\omega|z) = \frac{\sum_{z' \in Z_{\pi,B}} \phi(\omega|z')\pi(z') 1_z(h(z'))}{\tilde{\pi}(z)} \qquad \forall \omega \in \Omega, z \in Z_{\tilde{\pi},B'} \tag{144}$$

Sufficiency is similarly extended:

Definition 17. Consider two information structures (ϕ, π) and $(\tilde{\phi}, \tilde{\pi})$ such that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under $h \in \mathscr{H}$ mapping $\Theta \times B$ to $\Theta \times B'$. We say that $\tilde{\pi}$ is sufficient for π with respect to ϕ if $\phi(\omega \mid z) = \tilde{\phi}(\omega \mid h(z))$ for all ω and all z such that $\pi(z) > 0$.

Finally, invariance and monotonicity are written as follows:

Definition 18. Fix a set $H \subseteq \mathscr{H}$. Consider any function $h \in H$ and any two information structures (ϕ, π) and $(\tilde{\phi}, \tilde{\pi})$ such that $(\tilde{\phi}, \tilde{\pi})$ is the transformation of (ϕ, π) under g. A cost functional C is

1. invariant with respect to H if $C[\phi, \pi] = C[\tilde{\phi}, \tilde{\pi}]$ whenever $\tilde{\pi}$ is sufficient for π with respect to ϕ.

2. monotone with respect to H if $C[\phi, \pi] > C[\tilde{\phi}, \tilde{\pi}]$ whenever $\tilde{\pi}$ is not sufficient for π with respect to ϕ

While the wording of these definitions is rather similar, the definition has changed considerably. In particular, invariance and monotonicity can hold along the natural "extension" of a cost functional to different domains or state spaces. What does this extension mean? For posterior separable cost functionals, as defined in Appendix B, this relies on an appropriate extension of the divergence T. And if those divergences are f-divergences as defined in equation 34, reprinted here:

$$T[\pi; \pi'] = \sum_z \pi(z) \cdot f\left(\frac{\pi'(z)}{\pi(z)}\right), \tag{145}$$

this is immediate as the support of the probability distribution always remains finite with a length of at most $|\Theta|$. To use an example, it is reasonable to say that "mutual information costs are invariant and monotone as per Definition 18" *provided* that one presumes the natural extension of "always taking the mutual information of the signal and the state," whatever the (finite-support) state space is.

G.3 Possible Equilibrium Results

Given the above adaptation, it seems safe to conjecture, although we do not formally state and prove, that all of our equilibrium results including efficiency (Theorem 1), existence and implementability of Pareto optima (Proposition 8), and fundamentalness of equilibrium (Proposition 9) readily extend to our environment under invariance and monotonicity for all of $\mathscr{H}$ (e.g., as in our natural extension of mutual information). The content of these statements is that, under fully invariant and monotone costs like mutual information,

- Tax instruments are never optimal for a purely "informational" role, for instance to nudge agents toward or away from learning about certain contingencies.

- Media, expert opinions, blogs, and tweets, as captured in the exogenous signals $s = S(\theta)$, have no instrumental effect on equilibrium, as agents can always construct equivalent signals via unrestricted information acquisition.

- Noise in publicly available signals, which otherwise has no instrumental role, is optimally ignored in equilibrium.

- "Market data" in the form of trades or aggregate consumption statistics cannot be designed to have better informational content.

These lessons follow (we conjecture) from essentially the same premises articulated in our main analysis, but would be made more concrete in this extended environment.