Competitive Markets with Imperfectly Discerning Consumers*

Yair Antler and Ran Spiegler†

August 26, 2025

Abstract

In an adversely selective market model, products generate state-dependent potential hidden charges. Firms have differential abilities to realize this exploitative potential. Unlike firms, consumers do not observe the state. They try to infer hidden charges from headline prices, using idiosyncratic subjective models. Interior competitive equilibrium is uniquely characterized by a "Bellman equation". Relative to rational expectations, equilibrium hidden charges are lower, whereas total price and social welfare are higher. A regulatory intervention that lowers potential exploitation in one state can have adverse effects on consumers in other states. Market responses to shocks display patterns that are impossible under rational expectations.

*Financial support from Sapir Center, The Henry Crown Institute of Business Research, and UKRI Frontier Research Grant no. EP/Y033361/1 is gratefully acknowledged. We thank Meg Meyer, Marco Scarsini, Benny Moldovanu, Eylon Solan, Jakub Steiner, as well as seminar and conference audiences at Bar-Ilan University, Ben Gurion University, Cerge-EI, Nottingham, UPenn, PSE, University of Queensland, UNSW, and the GEA Meetings (Tbilisi).

†Antler: Tel Aviv University. Spiegler: Tel Aviv University and University College London

1 Introduction

One of the deepest ideas in the history of economic thought is that competitive markets aggregate private information through the price mechanism (Hayek (1945), Radner (1979)). According to this idea, competitive-equilibrium prices signal unobserved payoff-relevant features. Under mild assumptions, rational market participants can perfectly invert the equilibrium price signal and effectively make informed choices, as if all payoff-relevant information were public.

This logic relies heavily on the assumption that market participants can make flawless inferences from equilibrium prices. However, in reality, some participants possess a limited grasp of the systematic relation between prices and latent variables. For example, in retail markets, consumers may have a broad sense that price and quality are correlated, or that a low headline price implies hidden costs (as captured by sayings like "this deal is too good to be true," or "if you're not paying for the product, you are the product"), yet lack a precise understanding of such relations.

In this paper, we develop a model of a competitive consumer market in which consumers draw imperfect inferences from equilibrium prices. In our model, consumers who purchase a product pay a headline price as well as a hidden charge, which is an endogenous response by firms to a state of nature (observed by firms alone). We have in mind warranties and service agreements, where consumers may assume comprehensive coverage only to discover that certain repairs are excluded in the fine print. Paying for them thus constitutes a hidden charge.[1] Another example is financial products like credit cards, which frequently advertise attractive terms while hiding penalties in complex contracts.

In our model, there are no "free lunches" in equilibrium: A low headline price tends to signal a high hidden charge. Consumers are broadly aware of latent charges, yet differ in their ability to infer them from the equilibrium headline price. Fully rational consumers make perfect inferences, while others perform imperfect inferences based on idiosyncratic, imprecise subjective models. In other

[1] E.g., see this story about HP's "instant ink" cartridge replacement service, where buyers were surprised to discover that their printer was disabled if they stopped paying for the service: https://www.theatlantic.com/technology/archive/2023/02/home-printer-digital-rights-management-hp-instant-ink-subscription/672913/

words, consumers are *diversely discerning*. Our objective is to explore the market implications of this aspect of consumer bounded rationality.

The supply side in our model consists of a continuum of firms that observe an aggregate state defined by a collection of exogenous variables, before deciding whether to offer (at a cost) one unit of a basic product. Each state defines a distinct potential hidden charge. Firms differ in their ability to realize this potential. Specifically, when a consumer buys from a type-π firm in some state, he pays the firm a headline price as well as a hidden charge — also referred to as a "*ripoff*" — which is a fraction π of the potential in that state. We assume that π is uniformly distributed, which conveniently generates a linear supply function (a substantial part of our analysis is robust to relaxing linearity). Given the product's headline price, the firm types that enter the market in a given state are the ones with high π — i.e., those that are better at exploiting consumers. This feature makes our market adversely selective, in the spirit of Akerlof (1970).

Heterogeneity in π can reflect differences in firms' ingenuity in devising hidden charges (or moral scruples in resisting them).[2] Other sources of heterogeneity among firms are their exposure to regulatory restrictions, or the availability of ex-post opportunities for consumers to substitute away from their latent charges. There are also several interpretations for the state variables that determine the potential ripoff. There may be multiple exploitation channels (e.g., bank fees for various financial services); a state variable can indicate the feasibility of a particular channel. State variables can also represent regulations that constrain hidden charges, or indicate whether buying the basic product enables the firm to extract private information from the consumer and later use it at the consumer's expense. Finally, a state variable can represent a market segment, differentiated by product type, geographic location, or season.

The demand side in our model consists of a continuum of consumers. Each consumer knows his bare willingness to pay for the product. However, he is uninformed about the state or the type of firm he will buy from, and therefore aims to infer the expected ripoff from the headline price. Consumers are classified into "cognitive types"; there is a large measure of each type. A consumer of type

[2] This view of exploitative hidden charges is in the spirit of Heidhues et al. (2016), who regard them as fruits of firms' initiative.

M considers only a subset M of the state variables, ignoring the rest, because he is unaware of them or deems them irrelevant. The consumer infers the state variables in M from the market price and forms an estimate of the expected ripoff based on this inference. When M omits state variables, this is a model of "coarse beliefs" in the spirit of Eyster and Rabin (2005), Jehiel (2005), or Eyster and Piccione (2013). In particular, the exact formula for how consumers infer latent charges from headline prices is reminiscent of Mailath and Samuelson's (2020) "model-based inference".

Competitive equilibrium is an assignment of prices to states, such that each consumer optimizes with respect to his subjective belief given the headline price; each firm offers the product if and only if this is profitable given the headline price, the state, and the firm's type π; and supply equals demand in every state. An equilibrium is *interior* if there are both active and inactive firms in each state. As long as there is variation in consumers' bare willingness to pay (i.e., demand is downward sloping), interior equilibria are *fully revealing*, such that rational consumers can perfectly infer the expected ripoff from the headline price. By comparison, boundedly rational consumers can only partially decipher the signal that equilibrium prices provide, and therefore form wrong ripoff estimates.

Our analysis of interior equilibrium in Section 3 focuses on the limit case in which the variation in consumers' bare willingness to pay is negligible. In this limit, since there are many consumers of each type, the equilibrium price in each state is driven by the cognitive type with the lowest ripoff estimate (which itself is inferred from the headline price). An interior equilibrium exists for a range of values of our primitives. We show that it is *uniquely* characterized by a functional equation that is formally a *Bellman equation* — as if "the market" tries to minimize a discounted average of ripoffs across states. The Bellman equation expresses interdependence across states arising from the presence of imperfectly discerning consumers: Changes in the ripoff in one state can end up having a "contagious" effect on equilibrium outcomes in another state, when the equilibrium price in that state is driven by consumer types whose coarse perception cannot distinguish between the two states.

We use the "Bellman" characterization to probe the structure of interior equilibrium. In particular, we analyze the effects of expanding the set of cognitive types — e.g., when "coarse" consumers are introduced into a population of rational consumers. The expected equilibrium ripoff weakly *decreases* in every state, while the total expected price (i.e., the sum of the headline price and the expected ripoff) weakly *rises*. Thus, making the consumer population cognitively more diverse shifts equilibrium payments from latent to salient components. We use this finding to show that compared with rational-expectations equilibrium, the expected ripoff is weakly lower, yet the total price is weakly higher. In other words, *exploitation is larger and more "naked"*. Trade volume, and therefore social welfare, are larger (trade is socially beneficial in our model). However, this comes at the expense of boundedly rational consumers who overpay for the product because they underestimate the ripoff, and consequently suffer a welfare loss.

We demonstrate the value of the Bellman equation with additional equilibrium characterization results. We also analyze a specific environment in which potential ripoffs are monotone w.r.t state variables. We fully characterize equilibrium prices and hidden charges. This characterization enables us to examine the equilibrium effects of certain regulatory interventions. In particular, lowering the exploitative potential in one state hurts imperfectly discerning consumers in other states.

The equilibrating mechanism behind the Bellman equation can be illustrated by the scenario of expanding the set of cognitive types. This change raises the maximal net willingness to pay in some states. The increase in demand leads to higher headline prices, which in turn impels lower-π firms (which are less adept at devising exploitative hidden charges) to be active in these states. As a result, the average ripoff in these states decreases. Given consumers' belief-formation model, lower ripoffs across states raise their net willingness to pay, thus *reinforcing* the rise in demand.

Our framework also allows us to consider *non-exploitative* add-on product features — e.g., follow-up contracts that benefit both consumers and firms, or hotel add-on spa packages. Specifically, in Section 4 we modify the basic model by assuming that each state generates a distinct latent surplus that *both* consumer and firm enjoy. The unique interior equilibrium is characterized by a quasi-Bellman equation like the one we derive for the basic model, except that the "discount factor" is *negative*. The sign difference reflects *positive market selection* and carries distinct equilibrium implications. E.g., expanding the set of cognitive types need *not* have a uniform effect on latent payoffs across states.

Finally, in Section 5 we generalize our belief-formation model, so that cognitive types are *causal models* represented by so-called *perfect* directed acyclic graphs. This formalism, based on Spiegler (2016), subsumes coarse beliefs as a special case: Our focus on the latter in the basic model is purely to simplify exposition. It captures a wider variety of belief errors — e.g., perceiving that demand for add-ons drives hidden charges, while failing to realize it also influences headline prices. All our results extend to this more general model, which also generates novel effects. First, supply and demand responses to shocks can be virtually independent even though shocks' direct payoff implications are perfectly correlated across market agents. Second, when consumers receive private signals, equilibrium prices can reflect them on top of the payoff-relevant state, although prices fully reveal the latter. Thus, in the presence of imperfectly discerning consumers, equilibrium market outcomes can respond to factors beyond economic fundamentals.

2 The Model

Consider a market for a product with salient and latent components. Let p denote the (headline) price at which the product is traded. We now describe supply and demand in this market, and then define competitive equilibrium.

Supply

There is a measure one of firms. Each firm offers at most one unit of a product. Let $\Theta = \Theta_1 \times \cdots \times \Theta_n$ be a finite set of exogenous *states*. Let $\mu \in \Delta(\Theta)$ be a distribution over states. Unless stated otherwise, μ will have full support. For every $M \subseteq \{1, ..., n\}$, denote $\theta_M = (\theta_i)_{i \in M}$. Let $S : \Theta \to \mathbb{R}_{++}$ be a one-to-one function. Denote $S^{\max} = \max_\theta S(\theta)$, $S^{\min} = \min_\theta S(\theta)$, and $\bar{S} = \sum_\theta \mu(\theta) S(\theta)$. The quantity $S(\theta)$ represents the maximal potential hidden charge in state θ. A firm's *type* is $\pi \sim U[0, 1]$, representing the firm's ability to realize the exploitative potential. When a consumer purchases a product from a firm of type π in state θ, the firm incurs a production cost c, and a subsequent transfer of $\pi S(\theta)$ from the consumer to the firm (in addition to the price p) is realized. The transfer is *latent*, in the sense that consumers do not observe it when purchasing the product. We refer to the transfer as a *hidden charge* or a *ripoff*.

A firm of type π enters the market in state θ given the price p and sells one unit of its product, if and only if it earns a non-negative profit, i.e.,

$$p - c + \pi S(\theta) \geq 0 \tag{1}$$

Let $\pi^\star(\theta, p)$ be the value of π that satisfies (1) bindingly. Total supply under (θ, p) is the measure of active firms, which is equal to $1 - \pi^\star(\theta, p)$ (as long as $\pi^\star(\theta, p) \in [0, 1]$). Thanks to the assumption that π is uniformly distributed, we obtain a linear supply function in each state. We comment on the role of linearity in the concluding section. The hidden charge among active firms given (θ, p) is thus a random variable, denoted ϕ and distributed as follows:

$$\phi \mid (\theta, p) \sim U\left[\pi^\star(\theta, p)S(\theta), S(\theta)\right] \tag{2}$$

The expected ripoff given (θ, p) is

$$\overline{\phi}(\theta) = \frac{1 + \pi^\star(\theta, p)}{2} S(\theta) \tag{3}$$

Note that we can derive market supply independently of demand, since firms do not care with which consumers they trade. The exploitative potential given by S is independent of the type of consumer that buys the product.

Demand

There is a large population of consumers who buy at most one unit of the product. Let v be the consumer's bare valuation of the product, and assume it is distributed continuously over $[v^* - \varepsilon, v^* + \varepsilon]$. When the consumer buys the product in state θ at a headline price p, he is randomly matched with one of the active firms in the market — whose type π is thus drawn from $U[\pi^\star(\theta, p), 1]$ — and his net payoff is thus $v - p - \pi S(\theta)$. Each consumer is informed of his v, yet he is uninformed of θ and the type π of the firm he will interact with when buying the product. He tries to infer the expected ripoff from the market price. We will usually assume that ε is small, such that consumer preferences are nearly homogenous. In the $\varepsilon \to 0$ limit, a transaction generates a social surplus of $\Delta = v^* - c$. We assume that $\Delta > 0$.

Let $\mathcal{M}$ be a finite set of "cognitive types". The measure of consumers of each type is greater than one (recall that the firm population has measure one). Every $M \in \mathcal{M}$ is a distinct subset of the set $\{1, ..., n\}$ of exogenous variables. A type-M consumer is unaware of variables outside M, or finds them irrelevant.

Extend the distribution μ to a measure over triples (θ, p, ϕ). Thus, from now on, μ represents a joint probability measure over both exogenous and endogenous variables. Given μ, a type-M consumer forms the following subjective belief over the ripoff ϕ conditional on the observed price p (as long as p is realized with positive probability under μ):

$$\mu_M(\phi \mid p) = \sum_{\theta_M} \mu(\theta_M \mid p)\mu(\phi \mid \theta_M) \tag{4}$$

A consumer of cognitive type M is active given the price p if $v \geq p + E_M(\phi \mid p)$, where $E_M(\phi \mid p)$ is the expected ripoff conditional on p according to (4). The demand contributed by type-M consumers is the measure of such consumers who satisfy this inequality given μ.

A key property of (4) is that it is unbiased *on average* — i.e.,

$$\sum_p \mu(p)\mu_M(\phi \mid p) \equiv \mu(\phi)$$

Thus, while the consumer may fail to draw correct inferences about latent transfers from headline prices, the forecasts are not systematically biased. This distinguishes our model from a strand in the literature that includes Gabaix and Laibson (2006) and Heidhues et al. (2016,2017), where consumers neglect hidden charges altogether and therefore form systematically biased price evaluations.

Equilibrium

Consider a function h from states θ to prices p. This function, the objective distribution μ over states, and the distribution of active firms given by (2), induce the joint probability measure μ over θ, p, ϕ. In particular, $\mu(p = h(\theta) \mid \theta) = 1$ for every θ. This is the objective distribution that type-M consumers distort into $\mu_M(\phi \mid p)$.

We say that h is a *competitive equilibrium* if for every pair $(\theta, h(\theta))$, total supply is equal to the total demand induced by the distribution μ (which in turn is shaped by h). We say that a competitive equilibrium is *interior* if $\pi^\star(\theta, h(\theta)) \in (0,1)$ for every θ — that is, there are positive measures of both active and inactive firms in each state.

Discussion: Interpretation of (4)
This formula represents a two-stage thought process in the spirit of models of coarse reasoning, as in Jehiel (2005) and Eyster and Rabin (2005). It bears specific resemblance to Mailath and Samuelson's (2020) "model-based inference". In the first stage, the consumer infers the exogenous variables in his subjective model from observed prices, based on correct long-run statistical data. In the second state, he uses this intermediate inference to predict the ripoff, again based on correct long-run data. His error is that he omits exogenous variables that confound the relation between headline price and hidden charge. He also errs by assuming that p and ϕ are independent conditional on θ_M. This assumption has an intuitive graphical representation: $p \leftarrow \theta_M \rightarrow \phi$. Indeed, in Section 5, we embed (4) in a more general formalism in which consumers perceive market regularities through the prism of a subjective causal model, represented by a directed acyclic graph (as in Spiegler (2016)).

The graphical representation visualizes the inference procedure that (4) captures. For example, suppose the basic product is an electrical appliance. The consumer observes its price and tries to infer how much he will need to pay for follow-up service (repairs, ink cartridge replacement, etc.). When the product's headline price is low, the consumer goes through two steps of inference. First, he infers aspects of the product's type: Is it prone to malfunctioning? Does the warranty contain ambiguities that enable firms to charge the customer for certain services under the claim that they lie outside the contract's scope?[3] To what extent does consumer-protection regulation constrain firms in this regard? An alternative story is that the basic product is a hotel room. Here, the aspects that the consumer infers from a low hotel rate can be the hotel's geographic location (are there amenities nearby?) or the season (is this a low season for locals, such that most hotel guests will be foreigners?). In the second step, the consumer

[3]Recall the "instant ink" example of footnote 1. Another example is a claim that mobile-phone malfunctioning is due to corrosion and therefore falls outside the warranty's scope.

predicts the hidden charge based on the inferred product features. In each of the two steps, he relies on accessible statistical data, which enables him to form a quantitative belief based on this two-step inference.

Why does the consumer go through the two-stage process, instead of directly using the long-run conditional probability $\mu(\phi \mid p)$? One possible reason is that the consumer dogmatically believes the causal model given by $p \leftarrow \theta_M \rightarrow \phi$. The Bayesian posterior that is induced by this prior subjective model and the long-run data would be precisely (4). Another reason is that data on the joint distribution of p and ϕ may be unavailable, whereas data on how each of these variables covaries with θ_M is readily available. See Spiegler (2016,2020a) for an elaboration of these arguments and others.

2.1 Full Information Revelation

A basic question in models of competitive markets with imperfectly informed agents is whether equilibrium prices reveal the aggregate state θ. It turns out that interior equilibria in our model are fully revealing.

Proposition 1 *In every interior equilibrium h, $\theta \neq \theta'$ implies $h(\theta) \neq h(\theta')$.*

Proof. Consider an interior equilibrium h. Assume, contrary to the claim, that $h(\theta) = h(\theta') = p$ for some pair of states θ, θ'. This means that consumers cannot distinguish between the two states. As a result, the ripoff forecast $E_M(\phi \mid p)$ is the same in both states for every consumer type M. Consequently, aggregate demand is the same in both states. Turning to the supply side, by assumption $S(\theta) \neq S(\theta')$. Therefore, the L.H.S of (1) is different in the two states, such that $\pi^\star(\theta, p) \neq \pi^\star(\theta', p)$. It follows that supply is different in the two states while the price is the same. This can only be consistent with market clearing if demand is flat around p in θ and θ'. But since demand is downward-sloping around interior-equilibrium prices, we obtain a contradiction. ■

This result means that in any interior equilibrium, a consumer with rational expectations would perfectly deduce the state from the equilibrium price, and therefore have a correct assessment of the expected ripoff according to (3).

The full revelation result establishes a clear benchmark for our analysis: Prices are fully revealing, such that any failure by consumers to infer $\overline{\phi}(\theta)$ from $h(\theta)$ is due to their bounded rationality. The restriction to interior equilibria is instrumental in this regard. Without it, we could construct competitive equilibria in which the headline price is the same in two states in which all firms are active. In this case, both supply and demand are identical in the two states, which is consistent with assigning the same headline price to the two states.

We say that a distribution μ over (θ, p, ϕ) is *fully revealing* if both conditional distributions $(\mu(p \mid \theta))$ and $(\mu(\theta \mid p))$ are degenerate. Proposition 1 means that an interior equilibrium induces a fully revealing μ. In particular, we use $\theta^{\mu}(p)$ to denote the unique value of θ for which $\mu(\theta \mid p) = 1$. This enables us to simplify (4) into

$$\mu_M(\phi \mid p) = \sum_{\theta'} \mu(\theta' \mid \theta'_M = \theta^{\mu}_M(p))\mu(\phi \mid \theta') \tag{5}$$

where

$$\mu(\theta' \mid \theta'_M = \theta_M) = \frac{\mu(\theta')}{\sum_{\theta'' \mid \theta''_M = \theta_M} \mu(\theta'')}$$

Thus, the consumer forms his net willingness to pay for the product as if he learned the realization of the state variables in his model. At the same time, he fails to draw any inference from the event in which he trades with firms, which means that his net willingness to pay is higher than the net willingness to pay of consumers who do not trade. That is, the consumer essentially commits a "winner's curse" fallacy. in this sense, $\mu_M(\phi \mid p)$ is *not* consistent with a rational consumer who is partially informed of θ.

2.2 Rational Expectations Benchmark

Our model includes Rational Expectations Equilibrium (REE) as a special case, when the consumer's type is $M = \{1, ...n\}$ — i.e., he does not ignore any exogenous variable. In this case, $\mu_M(\phi \mid p) \equiv \mu(\phi \mid \theta^{\mu}(p))$. The reason is that by Proposition 1, p is a deterministic, one-to-one function of θ in interior equilibrium.

Full revelation also means that we can analyze equilibria separately for each state when consumers are rational. Let us derive the equilibrium for the *homogenous-preference* limit $\varepsilon \to 0$, where demand is flat because all consumers have a net willingness to pay of

$$v^* - \frac{1+\pi^\star(\theta, h(\theta))}{2} S(\theta) \tag{6}$$

Recall that the measure of the consumer population exceeds the measure of firms. Therefore, the equilibrium price in the $\varepsilon \to 0$ limit is determined by consumers' net willingness to pay. It follows that (6) is the expression for the equilibrium price $h(\theta)$ in the $\varepsilon \to 0$ limit, in terms of the threshold $\pi^\star(\theta, h(\theta))$. By definition, this threshold satisfies (1) bindingly in interior equilibrium when the market price is $h(\theta)$. Combining these equations, we obtain

$$\pi^\star(\theta, h(\theta)) = 1 - \frac{2\Delta}{S(\theta)} \tag{7}$$

It follows that an interior equilibrium exists whenever $2\Delta < S^{\min}$. Plugging (7) into (6), the equilibrium price and expected ripoff in state θ are

$$\begin{aligned} h(\theta) &= v^* + \Delta - S(\theta) \\ \overline{\phi}(\theta) &= S(\theta) - \Delta \end{aligned} \tag{8}$$

The total expected payment in state θ is $h(\theta) + \overline{\phi}(\theta) = v^*$, such that consumers end up paying their net willingness to pay for the product.

The interior REE is socially *inefficient*. Since $\Delta > 0$ and the transfer has zero sum, the efficient outcome is to maximize production — i.e., all firms should be active ($\pi^\star = 0$) in every state. Interior equilibria violate this requirement, by a standard *adverse-selection* argument. The state θ is an aggregate statistic that determines the potential for hidden transfers in the market, yet firms differ in their ability to realize this potential. Even when consumers perfectly infer θ from the market price, the equilibrium involves adverse selection because active firms are those with high ability to generate the exploitative hidden transfer. This lowers consumers' willingness to pay, which in turn lowers the equilibrium price and disincentives low-π firms from entering. The REE volume of trade is thus below the efficient level.

3 Analysis

This section is devoted to characterizing interior equilibrium in our model. We take the following for granted throughout the section. First, we make use of the result (Proposition 1) that interior equilibria are fully revealing. Second, we focus on the $\varepsilon \to 0$ limit, where all consumers' bare valuation of the product is v^*. In any equilibrium h of this limit case,

$$v^* - h(\theta) = \min_{M \in \mathcal{M}} E_{\mu_M}(\phi \mid p = h(\theta)) = \min_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta'_M = \theta_M)\overline{\phi}(\theta') \qquad (9)$$

for every state θ (the second equality makes use of (5)). That is, the equilibrium price in each state is equal to the highest net willingness to pay among all cognitive consumer types. The types that trade with firms in θ are the ones with the lowest (most optimistic) estimate of the ripoff.

We will often make use of a simple relation between equilibrium prices and the expected ripoff in each state:

$$h(\theta) = S(\theta) + c - 2\overline{\phi}(\theta) \qquad (10)$$

This equation follows from (1) and (3), when we plug $p = h(\theta)$ and make use of the fact that (1) is binding at $\pi^*(\theta, h(\theta))$ in an interior equilibrium. Equation (10) allows us to go back and forth between statements about ripoffs and statements about prices.

Consider the following restriction on the model's primitives:

$$S^{\max} - S^{\min} < 2\Delta < S^{\min} \qquad (11)$$

The proof of the following result, as well as some of the later ones, appears in the Appendix.

Proposition 2 *Suppose that condition (11) holds. Then, there exists a unique interior equilibrium. The expected equilibrium ripoff in each state is given by the functional equation:*

$$\overline{\phi}(\theta) = \frac{1}{2}\left[S(\theta) - \Delta + \min_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta'_M = \theta_M)\overline{\phi}(\theta')\right] \tag{12}$$

In particular, $S^{\min} - \Delta \leq \overline{\phi}(\theta) \leq S^{\max} - \Delta$ *for every* θ*.*

Equation (12) has the exact form of a *Bellman equation*, where the "discount factor" is $\frac{1}{2}$; each action corresponds to one of the models in $\mathcal{M}$; and the "transition probability" from θ to θ' induced by M is $\mu(\theta' \mid \theta'_M = \theta_M)$. Thus, in interior equilibrium, "the market" acts as if it tries to solve a Markov Decision Problem of minimizing a discounted sum of ripoffs, where the transition probabilities are derived from consumers' coarse beliefs.

The Bellman equation itself is an immediate consequence of putting the supply and demand equations (9) and (10) together. The proof of Proposition 2 is mostly devoted to establishing that the solution of (12) defines an interior equilibrium. The bounds on $\overline{\phi}(\theta)$ are the REE expected ripoff in the states having extremal values of S, as given by (8).

Unlike REE, the equilibrium equations for different states are *not* mutually independent. The reason is that consumers are imperfectly discerning, hence their willingness to pay in one state can reflect the expected ripoffs in other states. This means that shocks that affect demand in one state can have ramifications in other states. In other words, unlike REE, "*what happens in* θ *does not stay in* θ*.*" The interdependence has an "averaging" effect on expected ripoffs, relative to REE. Specifically, the bounds on the equilibrium levels of expected ripoffs mean that their range is *more compressed* than in REE.

To get a sense of the equilibrating dynamics behind (12), suppose that for some reason, the average ripoff in some state θ' is perturbed downward by a small amount $\eta > 0$. A type-M consumer's ripoff estimate in some other state θ decreases by $x = \mu(\theta' \mid \theta'_M = \theta_M)\eta$. If this type trades both before and after the perturbation, this means that demand (and hence the market price) shifts upward by x in θ. This impels lower firm types to enter the market in θ, causing the ripoff by the marginal active firm type in this state to decrease by x. Consequently, the average ripoff in θ drops by $0.5x$, as indicated by the 0.5 "discount factor" in (12).

Note that condition (11) ensures the existence of interior equilibrium for any $\mathcal{M}$ and μ. In applications that assume specific $\mathcal{M}$ and μ, the condition will often be significantly relaxed. Note also that since our definition of equilibrium focuses entirely on the price function h, uniqueness of interior equilibrium does not extend to allocations. In particular, if two consumer cognitive types happen to have the same ripoff forecast, we are agnostic about how trade is distributed between these two types.

3.1 An Example with Two State Variables

This sub-section presents an example that demonstrates the characterization of interior equilibrium given by (12). Let $n = 2$, $\mu = U\{(0,0),(0,1),(1,0)\}$, and $S(0,0) < S(0,1) \approx S(1,0)$. The set of cognitive types $\mathcal{M}$ consists of all subsets of $\{1,2\}$. Thus, type $\{1,2\}$ has rational expectations; type $\emptyset$ has fully coarse beliefs because he cannot perceive any correlation between headline price and ripoff; finally, types $\{1\}$ and $\{2\}$ have partially coarse beliefs because they omit one variable from their subjective models.

Guess an interior equilibrium: Type $\{1,2\}$ buys the product in state $(0,0)$ (in which his belief assigns probability one to this state); type $\{1\}$ buys the product in state $(0,1)$ (in which his belief is uniform over $(0,0)$ and $(0,1)$); and type $\{2\}$ buys the product in state $(1,0)$ (in which his belief is uniform over $(0,0)$ and $(1,0)$). Type $\emptyset$ never buys the product. Under this guess, (12) is reduced to the following system of linear equations:

$$
\begin{aligned}
2\overline{\phi}(0,0) &= S(0,0) - \Delta + \overline{\phi}(0,0) \\
2\overline{\phi}(0,1) &= S(0,1) - \Delta + \frac{1}{2}\overline{\phi}(0,1) + \frac{1}{2}\overline{\phi}(0,0) \\
2\overline{\phi}(1,0) &= S(1,0) - \Delta + \frac{1}{2}\overline{\phi}(1,0) + \frac{1}{2}\overline{\phi}(0,0)
\end{aligned}
$$

The solution is

$$\begin{aligned}\overline{\phi}(0,0) &= -\Delta + S(0,0) \\ \overline{\phi}(0,1) &= -\Delta + \frac{2S(0,1)+S(0,0)}{3} \\ \overline{\phi}(1,0) &= -\Delta + \frac{2S(1,0)+S(0,0)}{3}\end{aligned} \tag{13}$$

For the solution to define an interior equilibrium, we need $\frac{1}{2}S(\theta) < \overline{\phi}(\theta) < S(\theta)$, which holds whenever $2\Delta < S(0,0)$. This is also the condition for interior REE, which is more lenient than (11).

The following table summarizes the subjective ripoff estimates $E_M(\phi \mid \theta)$ for every type M (we use the abbreviated notation $\phi_{\theta_1\theta_2}$ for $\overline{\phi}(\theta_1, \theta_2)$):

$Type \backslash State$	$0,0$	$0,1$	$1,0$
$\{1,2\}$	ϕ_{00}	ϕ_{01}	ϕ_{10}
$\{1\}$	$\frac{1}{2}(\phi_{00}+\phi_{01})$	$\frac{1}{2}(\phi_{00}+\phi_{01})$	ϕ_{10}
$\{2\}$	$\frac{1}{2}(\phi_{00}+\phi_{10})$	ϕ_{01}	$\frac{1}{2}(\phi_{00}+\phi_{10})$
$\emptyset$	$\frac{1}{3}(\phi_{00}+\phi_{01}+\phi_{10})$	$\frac{1}{3}(\phi_{00}+\phi_{01}+\phi_{10})$	$\frac{1}{3}(\phi_{00}+\phi_{01}+\phi_{10})$

Since $\overline{\phi}(0,0) < \overline{\phi}(0,1) \approx \overline{\phi}(1,0)$, this table confirms our guess of the types with the lowest ripoff estimate in each state.

In the interior equilibrium we derived, the partially coarse types $\{1\}$ and $\{2\}$ earn negative payoffs in the states in which they buy the product, as their net willingness to pay exceeds the rational type's in these states. In contrast, the fully coarse type $\emptyset$, who is intuitively less sophisticated than the partially coarse types, enjoys a "*loser's blessing*": She earns zero payoffs because she never trades. Thus, the types who suffer a welfare loss are sophisticated enough to infer from the observed price that one state variable is favorable, but not sophisticated enough to understand that them buying the product implies that the other state variable is unfavorable. As a result, they underestimate the ripoff and overpay for the product. This kind of *non-monotonicity in consumer sophistication* has been observed in previous works (most relatedly, by Ettinger and Jehiel (2010) and Eyster and Piccione (2013)).

3.2 Characterization Results

In this sub-section we put Proposition 2 to work. Throughout the sub-section, we assume that an interior equilibrium exists (and is therefore unique). Our first result examines how the interior equilibrium changes when we expand the set of cognitive types $\mathcal{M}$ — i.e., when consumers become more diverse in terms of their subjective models. This also means increasing demand, because the measure of each type is above one. The "Bellman" characterization of interior equilibrium means that expanding $\mathcal{M}$ is formally equivalent to expanding the set of actions in a Markov Decision Problem (MDP). This equivalence enables us to tap into standard results on solutions of MDPs and apply them to the present context, where they have very different meaning.

Proposition 3 *Adding a new type M to $\mathcal{M}$ has the following effects on the unique interior equilibrium.*

(i) $\overline{\phi}(\theta)$ weakly decreases in every θ.

(ii) $h(\theta) + \overline{\phi}(\theta)$ weakly increases in every θ.

(iii) Social surplus weakly increases in every θ.

(iv) If $\mathcal{M}$ includes both rational and non-rational types, then consumers incur an aggregate ex-ante welfare loss, which is weakly larger under $\mathcal{M} \cup \{M\}$ than under $\mathcal{M}$.

Proof. By Proposition 2, $\overline{\phi}(\theta)$ is formally the solution to a finite-state MDP of minimizing a discounted expected cost function, where $\mathcal{M}$ is the set of feasible actions in this MDP. Expanding the set of feasible actions weakly improves the value function at each state, which implies (i). Property (ii) then immediately follows from (i) and equation (10).

To see why property (iii) follows from (i), note that by (3), $\overline{\phi}(\theta)$ decreases if and only if $\pi^*(\theta)$ decreases. Therefore, the expansion of $\mathcal{M}$ leads to a weak decrease in $\pi^*(\theta)$ in each state θ. This means that there are more active firms — and hence more trade — in each state. As we saw, in this model social welfare is pinned down by the volume of trade.

As to property (*iv*), note that consumers who do not trade in a given state earn zero payoffs. Consumers who do trade in a state θ earn a net payoff of $v^* - h(\theta) - \overline{\phi}(\theta)$. Plugging (10), this expression becomes $\Delta - S(\theta) + \overline{\phi}(\theta)$. Since the expansion of $\mathcal{M}$ leads to a weak decrease in $\overline{\phi}(\theta)$, active consumers' net payoff in θ weakly decreases, too. When $\mathcal{M}$ includes a rational type, the net payoff of any consumer who trades in any state must be weakly negative, because the equilibrium price is equal to this type's willingness to pay and therefore lies weakly above the rational-expectations willingness to pay. As we saw above, the volume of trade — which is equal to the measure of consumers who trade — weakly increases in each state when we expand $\mathcal{M}$. Thus, not only does the net payoff loss of each trading consumer weakly increases when we expand $\mathcal{M}$, but there are also weakly more consumers who trade in each state. This means that consumers' ex-ante welfare loss weakly goes up. ■

Thus, expanding the set of cognitive types shifts payments from hidden charges to salient prices — i.e., ripoffs decrease while headline prices increase. The total price increases. The basic intuition behind this *"shift towards naked exploitation"* is as follows. An expansion of $\mathcal{M}$ leads to an increase in demand, and therefore higher equilibrium prices in each state. In response, the pool of active firms becomes less adversely selective, as lower-π types enter the market thanks to the higher price. This in turn means that latent exploitation shrinks in equilibrium. Since consumers' ripoff assessments are effectively weighted averages of expected ripoffs across states, this raises consumers' willingness to pay and therefore reinforces the increase in demand.

Recall that in REE, $\overline{\phi}(\theta) = S(\theta) - \Delta$ for every θ. Therefore, the ex-ante expected ripoff in REE is $\bar{S} - \Delta$. The following result draws on Proposition 3 to show that the ex-ante expected ripoff in interior equilibrium is weakly below this REE level. The result also shows that the lowest possible expected ripoff given $\bar{S}$ is approximately sustainable in equilibrium (for a suitable specification of primitives).

Proposition 4 *In interior equilibrium, the expected ripoff is in* $[\frac{1}{2}\bar{S}, \bar{S}-\Delta]$. *Moreover, the lower bound can be approximated arbitrarily well by interior equilibrium for a suitable selection of* $\Theta, S, \mu, \mathcal{M}$ *that is compatible with* $\bar{S}$.

The argument behind the proposition's first part is simple. When $\mathcal{M}$ is a singleton, (12) becomes a linear equation in $\overline{\phi}(\theta)$, for every θ. This linearity, coupled with the unbiasedness-on-average property of consumers' beliefs, implies that the ex-ante expected ripoff in interior equilibrium coincides with the REE level. When we add cognitive types, Proposition 3 implies a drop in the expected ripoff.

The lower bound on the expected ripoff is attained in a large-n variant on the example of Section 3.1. In equilibrium, the trading consumer in every state has an optimistic belief in the sense that he believes that the expected ripoff is close to its lowest possible level ($S = S^{\min}$, which happens when $\pi^* = 0$). The equilibrium outcome is nearly efficient, as $\pi^* \approx 0$ in every state, such that there is virtually no adverse selection, and the equilibrium headline price is close to c in every state.

Proposition 4 and part (iv) of Proposition 3 imply that rational consumers impose a negative externality on boundedly rational ones (reminiscent of a similar effect highlighted by Gabaix and Laibson (2006), although its origin here is different). When $\mathcal{M}$ consists of a single non-rational type, consumers earn zero net expected payoffs, because their ex-ante expected ripoff estimate is consistent with rational expectations. Adding rational consumers leads to a negative aggregate ex-ante consumer payoff. But rational consumers always earn zero payoffs (because when they trade, their total payment equals their willingness to pay). This means that the welfare loss due to the rational type's entry is borne by the non-rational consumers.

Thanks to (10), Proposition 4 has an immediate implication for equilibrium headline prices.

Corollary 1 *The ex-ante expected price in interior equilibrium is weakly above its REE level* $v^* + \Delta - \bar{S}$.

Turning from average price components to their range, recall that by Proposition 2, the range of expected ripoffs in interior equilibrium is compressed relative to REE. The following result obtains an analogous result for prices, as long as there are rational consumers in the market.

Proposition 5 *If $\mathcal{M}$ includes the rational type, then $v^* + \Delta - S^{\max} \leq h(\theta) \leq v^* + \Delta - S^{\min}$ for every θ in the interior equilibrium. Moreover, the R.H.S inequality is binding when $S(\theta) = S^{\min}$.*

Recall that the REE price function is $h(\theta) = v^* + \Delta - S(\theta)$. Thus, Proposition 5 implies that adding imperfectly discerning consumers to a market that already contains rational consumers reduces the extent of equilibrium price fluctuations. Note that when the market does not contain the rational type to start with, adding types to $\mathcal{M}$ may result in wider price fluctuations. To see why, suppose $\mathcal{M}$ consists of a single type $M = \emptyset$. This type has fully coarse beliefs, and therefore his willingness to pay is constant across states. It follows that the equilibrium price is absolutely rigid. Now add the rational type to $\mathcal{M}$ and note that the coarse type's state-independent equilibrium willingness to pay is a convex combination of the rational type's equilibrium willingness to pay in all states. Thus, adding the rational type widens the range of equilibrium prices.

3.3 A Monotone Environment

In this section we provide a complete characterization of interior equilibrium for a natural specification, and use it to show the perverse effects of a certain regulatory intervention.

Let $\theta_i \in \{0, 1\}$ for every $i \in \{1, ..., n\}$. Denote $I(\theta) = \{i \mid \theta_i = 1\}$. We will often write $\theta > \theta'$ when $I(\theta') \subset I(\theta)$. Assume that $\mathcal{M}$ consists of *all* subsets of $\{1, ..., n\}$. Suppose that S is strictly decreasing in θ, such that $I(\theta') \subset I(\theta)$ implies $S(\theta') > S(\theta)$. The distribution μ is *conditionally increasing*: For every $i \in \{1, ..., n\}$ and $J \subset \{1, ..., n\}$ such that $i \notin J$, $\mu(\theta_i = 1 \mid \theta_J)$ is increasing in θ_J. This well-known property in the literature on stochastic orders is weaker than affiliation (see Müller and Stoyan (2002, pp. 125-127)).

The structure of S and μ defines a *monotone environment*. For every i, $\theta_i = 1$ is a "good" state realization because it induces a lower potential ripoff, and also because it is positively associated with similarly good realizations of other state variables.

For every θ, define M^θ to be the set of state variables such that $M^\theta = I(\theta)$. The interpretation of M^θ is that it omits all the state variables that take the value

0 in θ. In other words, M^{θ} is the set of state variables whose realization in θ is good.

Proposition 6 *In the unique interior equilibrium,* $\phi(\theta) = v^* - E_{M^\theta}(\phi \mid \theta)$ *for every* θ*. Moreover, the state-dependent expected ripoff can be calculated recursively:*

$$2\bar{\phi}(\theta) = S(\theta) - \Delta + \sum_{\theta' \geq \theta} \mu(\theta' \mid \theta'_{I(\theta)} = \theta_{I(\theta)})\bar{\phi}(\theta') \tag{14}$$

In equilibrium, the trading consumer type in each state θ is M^{θ}. Thus, equilibrium outcomes are always based on a simple optimistic interpretation of headline prices. The trading consumer focuses on the "good news" that the price reflects (namely, the state variables that take a value that is associated with lower ripoffs) and ignores the "bad news". This property leads to a recursive characterization of the expected ripoff in each state.

The monotone environment illustrates the already-noted motto, "what happens in θ need not stay in θ." Suppose that there is a slight increase in $S(\theta)$ for some state θ. From the recursive representation (14), it is possible to see that this change raises the equilibrium expected ripoff in every state $\theta' \leq \theta$, but only in these states.

Unintended consequences of regulatory intervention

We now use Proposition 6 to illustrate how a policy intended to reduce the potential ripoff in one state may backfire. As an illustration, suppose the different states correspond to distinct markets — e.g., different kinds of electrical appliances, which differ in the looseness of their warranties and therefore in their scope for exploitative ripoffs. Under this interpretation, the consumer's predicament is that he does not know in what kind of market he is in.

Consider a policy intervention that enhances the transparency of contract terms in one particular market, thereby reducing the maximal ripoff there. Under rational expectations, this would lower the expected ripoff in that market, while having no effect on equilibrium outcomes in other markets. The intervention would have no effect on consumer welfare in any market (because the market-clearing price is equal to consumers' willingness to pay, which is based on a correct estimate of the hidden ripoff). The presence of imperfectly discerning consumers gives rise to unintended consequences for consumer welfare. Specifically, consumers in other markets may mistakenly infer a sharper overall decline in ripoff levels than has actually occurred. As a result of this misperception, they are willing to pay more than they should, which in turn leads to increased overpayment and a greater welfare loss.

We now present a result that formalizes this argument. We need a few pieces of notation. Let J be a maximal ripoff function that is identical to S except for a specific state θ^* in which $J(\theta^*) = S(\theta^*) - k$, and denote its induced equilibrium average-ripoff function by $\overline{\phi}_J$. For convenience, we assume that k is not too large, such that the environment's monotonicity is maintained under J. Denote $k(\theta) = \overline{\phi}(\theta) - \overline{\phi}_J(\theta)$.

Proposition 7 *In the interior equilibrium induced by J, consumers of type M^θ with $\theta < \theta^*$ ($\theta > \theta^*$) suffer a greater (the same) welfare loss than in the interior equilibrium induced by S.*

Proposition 7 shows that a regulatory intervention that lowers $S(\theta)$ can reduce the welfare loss of at most one type. This result also establishes that such a policy harms all consumers whose perception is coarser than M^θ. These types effectively hold the mistaken belief that the reduction in $S(\theta)$ leads to a sharp decrease in the hidden ripoff in the state where they buy the product. In turn, this effect increases their willingness to pay for the product more than it should, which results in a greater welfare loss.

4 Mutually Beneficial Add-Ons

So far, we have assumed that the latent quantity is a zero-sum transfer from consumers to firms. In many real-life contexts, however, it can be an add-on feature that generates surplus for both parties. E.g., the add-on can be a follow-up service that the consumer does not expect to be included in the basic warranty, and that can only be provided by the basic product's seller. If demand for this service is linearly downward-sloping, the optimal monopoly price for the service will split the surplus equally between the consumer and the firm. In this section, we present a variant of our model that covers such cases.

Assume that when a consumer buys from a type-π firm in state θ, *each of them* obtains a latent payoff of $\pi S(\theta)$. We refer to $\phi = \pi S(\theta)$ as the *quality* that the consumer gets in this case, and to $\overline{\phi}(\theta)$ (as defined by (3)) as the *average quality* in state θ. The restriction to interior equilibria will require us to assume that $\Delta = v^* - c < 0$, such that add-ons are necessary for gains from trade. All other definitions and assumptions are as in the model of Section 2. In particular, supply is identical, and interior equilibria are fully revealing.

In the homogenous-preference limit $\varepsilon \to 0$, interior equilibrium is given by

$$h(\theta) = v^{\star} + max_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta'_M = \theta_M) \overline{\phi}(\theta')$$

for every state θ. Compare this expression with (9). The equilibrium price in state θ is determined by the consumer type with the *highest* quality estimate in that state (whereas in the basic model, the type with the *lowest* estimate determined the price). Combining this equation for $h(\theta)$ with the supply-driven equation (10), we obtain

$$\overline{\phi}(\theta) = \frac{1}{2} \left[S(\theta) - \Delta - max_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta'_M = \theta_M) \overline{\phi}(\theta') \right] \tag{15}$$

This equation is exactly the same as (12), except for the *minus sign* before the max operator. In other words, it is like a Bellman equation with a negative discount factor. The equation defines a contraction mapping, and so it has a unique solution, pinning down $h(\theta)$ and $\pi^{\star}(\theta, h(\theta))$. The following condition ensures that the unique solution defines an interior equilibrium:

$$-\frac{2}{3}\Delta < S^{min} < S^{max} < -\Delta \tag{16}$$

This mutual-benefit model is qualitatively different from the pure-exploitation model of Section 2. In particular, expanding $\mathcal{M}$ need *not* have a uniform effect on equilibrium quality levels across states. The following example illustrates this effect. Let $n = 1$, $\theta \in \{0, 1\}$, and assume μ is uniform. Let $S(0) = k < 1 = S(1)$ and assume (16) holds. Suppose $\mathcal{M}$ consists of a single, "fully coarse" type $M = \emptyset$. This type's quality estimate is $(\overline{\phi}(0) + \overline{\phi}(1))/2$ in both states. The solution to (15) is $\overline{\phi}(0) = (5k - 4\Delta - 1)/12$ and $\overline{\phi}(1) = (5 - 4\Delta - k)/12$. Now add a rational type to $\mathcal{M}$. We can guess and verify that in equilibrium, the rational type buys the product in $\theta = 1$ and the coarse type buys the product in $\theta = 0$. The solution to (15) is $\overline{\phi}(0) = (6k - 5\Delta - 1)/15$ and $\overline{\phi}(1) = (1 - \Delta)/3$. Thus, as a result of the expansion of $\mathcal{M}$, expected equilibrium quality decreases in $\theta = 1$ and increases in $\theta = 0$.

To appreciate the difference from the exploitative-transfer case, let us track the intuitive equilibrating mechanism following the addition of rational consumers. Whether hidden features are exploitative or mutually beneficial, this change leads to an initial demand increase in one state θ (where rational consumers have a higher willingness to pay than coarse consumers), which pushes the headline price in θ upwards. The ensuing market entry by low-π firm types lowers $\overline{\phi}(\theta)$. This is where the two cases diverge. In the pure-exploitation case, lower hidden charges reinforce the increase in demand across states. In contrast, lower quality *curbs* demand in the mutual-benefit case. In state θ, this has a partially offsetting effect on the initial rise in demand. However, since "what happens in θ does not stay in θ", the drop in demand also occurs in the other state, which never witnessed the initial demand increase in the first place. In that state, the headline price goes down.

Some effects of the basic model continue to hold in the mutual-benefit case, possibly with a change in sign, as long as $\mathcal{M}$ includes a rational type. In a previous version of the paper (Antler and Spiegler (2024)), we showed that under this condition, the interior-equilibrium levels of $\overline{\phi}(\theta)$ and and $h(\theta)$ are weakly below and above their REE level, respectively. Under the same condition, expanding $\mathcal{M}$ leads to a *decrease* in equilibrium social surplus.

The reason for this effect is that in the mutual-benefit case, our market is *positively selective* — i.e., the firm types that enter the market are the ones that create more latent surplus for consumers. In REE, consumers earn zero net payoffs on average, which means that trading with the marginal firm type π^* is harmful for consumers. Since this firm type is indifferent to market entry, trading with it is socially harmful. In other words, the REE volume of trade is excessive from the perspective of social welfare. When $\mathcal{M}$ includes a rational type and we expand this set, even lower-quality firms enter the market, which exacerbates this social harm.

5 A Broader Class of Subjective Models

In this section we revert to the exploitative-transfer version of the model, while extending the consumer belief-formation model presented in Section 2. Following Spiegler (2016), a cognitive type is defined by a *subjective causal model* that is represented by a *directed acyclic graph* (DAG) $G = (N, R)$. A node in N represents an exogenous or endogenous variable, and a directed edge in R represents a perceived causal relation between two variables. Let $\mathcal{G}$ be the set of subjective causal models in the consumer population. Every $G \in \mathcal{G}$ must include nodes that represent p and ϕ. Yet, it is not allowed to have links of the form $p \rightarrow \theta_i$ or $\phi \rightarrow \theta_i$ — i.e., consumers realize that state variables are exogenous whereas p and ϕ are endogenous. As before, the measure of consumers of each type $G \in \mathcal{G}$ is greater than 1.

Label the variables in G as $(x_i)_{i \in N}$ (the enumeration is arbitrary). Abusing notation, $R(i)$ is the set of nodes that send a directed link into i. A node i is *ancestral* if $R(i) = \emptyset$. When the objective joint distribution over (θ, p, ϕ) is μ, G induces the following subjective probabilistic belief:

$$\mu_G(x_N) = \prod_{i \in N} \mu(x_i \mid x_{R(i)}) \tag{17}$$

This is the standard Bayesian-network factorization formula (see Pearl (2009) and Spiegler (2016)). When μ has full support, μ_G is a well-defined probability distribution over x_N.

A DAG $G = (N, R)$ is perfect if, for every triple of nodes $i, j, k \in N$, $i, j \in R(k)$ implies $i \in R(j)$ or $j \in R(i)$. In a perfect DAG, the parents of every node form a clique. The basic model of Section 2 is a special case of the perfect-DAG formalism: M is a subset of the nodes that represent θ; all the nodes in M are mutually linked; and $R(p) = R(\phi) = M$. There are two motivations for adopting the perfect-DAG formalism. First, perfect DAGs subsume earlier equilibrium market models with non-rational expectations as special cases, while making room for new ones. Second, perfect DAGs represent the most general class of DAGs that always satisfy the unbiasedness-on-average property observed in Section 2 (see Spiegler (2020b)). Suppose G is a perfect DAG. Then, for every μ that arises from an interior equilibrium h,

$$\sum_{\theta} \mu(\theta)\mu_G(\phi \mid h(\theta)) \equiv \sum_{p} \mu(p)\mu_G(\phi \mid p) \equiv \mu(\phi) \tag{18}$$

The left-hand identity arises from h being a one-to-one function of θ in interior equilibrium. For the right-hand identity, see Spiegler (2020a,b).

We will use the perfect DAG $G_{ch} : p \leftarrow \theta_1 \rightarrow \theta_2 \rightarrow \phi$ as a running example in this section. This DAG captures consumers who correctly perceive the connection between the exogenous variables, but mistakenly think that the product's salient and latent components are caused by different exogenous variables. The DAG G_{ch} induces the subjective belief $\mu_{G_{ch}}(\theta_1, \theta_2, \phi, p) = \mu(\theta_1)\mu(\theta_2 \mid \theta_1)\mu(p \mid \theta_1)\mu(\phi \mid \theta_2)$, which in turn yields the conditional belief

$$\mu_G(\phi \mid p) = \sum_{\theta_1,\theta_2} \mu(\theta_1 \mid p)\mu(\theta_2 \mid \theta_1)\mu(\phi \mid \theta_2)$$

5.1 Generalizing the Bellman Equation

We now present a lemma that characterizes the conditional belief $\mu_G(\phi \mid p = h(\theta))$ when G is a perfect DAG. We refer to a system of conditional probabilities $\beta = (\beta(\theta' \mid \theta))_{\theta,\theta' \in \Theta}$ as a *transition matrix*. Recall that $\theta^{\mu}(p)$ is the state θ that generates the price p in a fully revealing μ.

Lemma 1 *Fix a perfect DAG $G = (N, R)$. Then, there exists a unique transition matrix β satisfying the following: For every fully revealing distribution μ over (θ, p, ϕ), and for every price p in the support of μ,*

$$\mu_G(\phi \mid p) = \sum_{\theta'} \beta(\theta' \mid \theta^{\mu}(p))\mu(\phi \mid \theta') \tag{19}$$

Moreover, the projection of μ on Θ is an invariant distribution of β_G.

Thus, given μ and G, we have a simple representation of the consumer's belief over the ripoff conditional on the market price.[4] Instead of correctly inferring the ripoff distribution (2) in the state revealed by the market price, the consumer effectively calculates a weighted average of the ripoff distributions associated with various "virtual" states; the weights on virtual states may vary with the actual state. This representation is made possible by the property that μ is fully revealing, such that there is a one-to-one mapping between prices and states.

In the basic model of Section 2, $\beta(\theta' \mid \theta) = \mu(\theta' \mid \theta'_M = \theta_M)$. For the DAG $G_{ch} : p \leftarrow \theta_1 \rightarrow \theta_2 \rightarrow \phi$ introduced above, and given a fully revealing equilibrium h,

$$\mu_G(\phi \mid h(\theta)) = \sum_{\theta'_2} \mu(\theta'_2 \mid \theta_1)\mu(\phi \mid \theta'_2) = \sum_{\theta'_2}\sum_{\theta'_1} \mu(\theta'_2 \mid \theta_1)\mu(\theta'_1 \mid \theta'_2)\mu(\phi \mid \theta'_1, \theta'_2)$$

such that

$$\beta(\theta'_1, \theta'_2 \mid \theta_1, \theta_2) \equiv \mu(\theta'_2 \mid \theta_1)\mu(\theta'_1 \mid \theta'_2)$$

To illustrate this formula, let $n = 2$, $\theta_1, \theta_2 \in \{0, 1\}$, $\mu = U\{(0,0), (1,0), (0,1)\}$. Then,

$$\begin{aligned}
\beta(\cdot, 0 \quad | \quad 0, \cdot) &= \mu(\theta'_2 = 0 \mid \theta_1 = 0)\mu(\theta'_1 = \cdot \mid \theta'_2 = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\
\beta(0, 1 \quad | \quad 0, \cdot) &= \mu(\theta'_2 = 1 \mid \theta_1 = 0)\mu(\theta'_1 = 0 \mid \theta'_2 = 1) = \frac{1}{2} \cdot 1 = \frac{1}{2} \\
\beta(\cdot, 0 \quad | \quad 1, 0) &= \mu(\theta'_2 = 0 \mid \theta_1 = 1)\mu(\theta'_1 = \cdot \mid \theta'_2 = 0) = 1 \cdot \frac{1}{2} = \frac{1}{2}
\end{aligned}$$

Observe that the transition matrix assigns positive weight to $\theta'_1 \neq \theta_1$, even though the consumer correctly infers θ_1 from p.

A fully connected DAG that includes all θ variables induces rational expectations, because in this case (17) becomes the standard chain rule for $\mu(\theta, p, \phi)$. However, this is not the only class of perfect DAGs that are guaranteed to induce correct equilibrium beliefs, because in equilibrium, p is a deterministic function of θ. When a DAG G does not exclude any of the θ variables, and every pair of nodes is linked (except possibly (p, ϕ)), then it represents a rational consumer.

[4] This representation is somewhat reminiscent of a model of misperception of correlations by Ellis and Piccione (2017).

Likewise, a perfect DAG in which p and ϕ are directly linked induces rational expectations. The transition matrix that represents such consumers is the unit matrix, $\beta(\theta \mid \theta) = 1$ for all θ.

Proposition 2 extends to the present belief-formation model whenever $\mathcal{G}$ is a collection of perfect DAGs. The Bellman-like equation (12) is modified into

$$\overline{\phi}(\theta) = \frac{1}{2}\left[S(\theta) - \Delta + \min_{G\in\mathcal{G}} \sum_{\theta'} \beta_G(\theta' \mid \theta)\overline{\phi}(\theta')\right] \tag{20}$$

where β_G is the transition matrix that represents the perfect DAG G. Condition (11) continues to ensure existence and uniqueness of interior equilibrium. All the other results in Section 3 extend as well. The mutually beneficial add-on variant of Section 4 is extended in the same manner. The quasi-Bellman equation that characterizes interior equilibrium is the same as (20), except that the last term in the squared brackets is preceded by a minus sign (and the condition for interior equilibrium is (16)).

However, the DAG formalism offers more than a mere generalization of our previous results. It also generates novel effects beyond the basic model's reach, as the next sub-sections demonstrate.

5.2 The Two-State-Variables Example Revisited

To illustrate the use of (20) to characterize interior equilibrium in the DAG-based extension, revisit the example of Section 3.1, where $n = 2$, $\theta_1, \theta_2 \in \{0, 1\}$, $\mu = U\{(0,0),(0,1),(1,0)\}$, and $S(0,0) < S(1,0) \approx S(0,1)$. The set of cognitive types $\mathcal{G}$ consists of a rational type, and the two chain DAGs $G_1 : p \leftarrow \theta_1 \rightarrow \theta_2 \rightarrow \phi$ and $G_2 : p \leftarrow \theta_2 \rightarrow \theta_1 \rightarrow \phi$. We presented the transition matrix that represents G_1 in the previous sub-section. Using similar calculations, the matrix that represents G_2 is: $\beta(0,0 \mid \cdot,0) = \beta(0,1 \mid \cdot,0) = 0.25$, $\beta(1,0 \mid \cdot,0) = 0.5$, and $\beta(0,0 \mid \cdot,1) = \beta(0,1 \mid \cdot,1) = 0.5$.

We now guess an equilibrium, and later verify that our guess is indeed an equilibrium. As before, the guess-and-verify method is valid because there is at most one interior equilibrium. Suppose the rational type buys the product in state $(0,0)$; type G_1 buys the product in state $(1,0)$; and type G_2 buys the product in state $(0,1)$. Under this guess, (20) takes the exact same form as (12), leading to the same solution (13) for $\overline{\phi}(\theta)$. Let us verify that the type who buys in each state indeed has the lowest ripoff estimate. The following table presents expressions for each type's estimate in each state (we use the abbreviated notation $\phi_{\theta_1\theta_2}$ for $\overline{\phi}(\theta)$):

$Type \backslash State$	$0,0$	$0,1$	$1,0$
$rational$	ϕ_{00}	ϕ_{01}	ϕ_{10}
G_1	$\frac{1}{4}(\phi_{00}+\phi_{10})+\frac{1}{2}\phi_{01}$	$\frac{1}{4}(\phi_{00}+\phi_{10})+\frac{1}{2}\phi_{01}$	$\frac{1}{2}(\phi_{00}+\phi_{10})$
G_2	$\frac{1}{4}(\phi_{00}+\phi_{01})+\frac{1}{2}\phi_{10}$	$\frac{1}{2}(\phi_{00}+\phi_{01})$	$\frac{1}{4}(\phi_{00}+\phi_{01})+\frac{1}{2}\phi_{10}$

Recall that (13) implies $\phi_{00} < \phi_{01} \approx \phi_{10}$, hence our guess is confirmed.

While the expected ripoff in each state is the same as in Section 3.1, the inference behind the trading consumer types' estimates is different. For example, when the state is $(1,0)$, type G_1 correctly infers $\theta_1 = 1$ from the equilibrium price. While this realization by itself is associated with a *high* ripoff (because the only state in which $\theta_1 = 1$ is $(1,0)$), the type's DAG leads him to assign probability $\frac{1}{2}$ to the state $(0,0)$, in which the ripoff is at its *lowest*. Thus, unlike the example in Section 3.1, a pessimistic inference about the state variable the consumer regards as the direct cause of prices leads to an *optimistic* ripoff forecast.

5.3 "Anomalous" Market Fluctuations

In competitive markets, fluctuations in prices and allocations reflect supply and demand responses to external shocks. When REE fully reveals all payoff-relevant information, these responses are as if the information is public. In this sub-section, we demonstrate that under the DAG-based extension of our model, equilibrium supply and demand responses to shocks exhibit patterns that are impossible in REE (or under our basic model). First, we show that although supply and demand shocks in our model are perfectly correlated (negatively for most of the paper, positively in the variant of Section 4), the supply and demand responses can be nearly independent. Second, we extend the model by endowing consumers with private information, and show that even though equilibrium prices fully reveal all payoff-relevant aspects of the state, they can also respond to fluctuations in consumers' private information. The common theme in both sub-sections is that markets with imperfectly discerning consumers are more "jittery" relative to REE.

5.3.1 How Supply and Demand Co-Move

The state θ in our model determines a zero-sum transfer from consumers to firms. Therefore, under rational expectations, supply and demand move in opposite directions in response to fluctuations in θ. Now suppose $\mathcal{G}$ consists of a single "fully coarse" consumer, who does not perceive any correlation between visible and latent features. This consumer will exhibit an absolutely rigid demand, such that equilibrium price fluctuations only reflect supply responses to shocks.

Our model can also generate virtually independent supply and demand movements in response to shocks. E.g., let $\theta = (\theta_1, \theta_2, \theta_3)$, $\theta_i \in \{0, 1\}$ for every i. Assume that $S(\theta) = \alpha_1\theta_1 + \alpha_2\theta_2 + \alpha_3\theta_3 + b$, where $b > 0$ is a constant; and the weights α_i are all positive and different from each other. Moreover, let $\alpha_1, \alpha_3 \approx 0$, whereas α_2 is bounded away from zero, such that the maximal feasible ripoff is almost entirely a function of θ_2. Assume that μ satisfies the following properties: θ_1 and θ_2 are statistically independent, and θ_3 is some function of these two state variables. Under this specification, the supply function mainly responds to fluctuations in θ_2, and exhibits virtually no response to the other state variables conditional on θ_2.

Finally, assume $\mathcal{G}$ consists of a DAG $G : p \leftarrow \theta_1 \rightarrow \theta_3 \rightarrow \theta_2 \rightarrow \phi$. Even though θ_1 and θ_2 are objectively independent, they may be correlated according to the subjective belief μ_G, as long as both θ_1 and θ_2 are correlated with θ_3, since

$$\mu_G(\theta_2 \mid \theta_1) = \sum_{\theta_3'} \mu(\theta_3' \mid \theta_1)\mu(\theta_2 \mid \theta_3')$$

Eliaz et al. (2021) showed that this spurious subjective correlation can be quite large.

In our context, what this observation means is that consumer demand will be highly responsive to prices, because the consumer correctly infers θ_1 from the equilibrium price while exaggerating the correlation between θ_1 and ϕ (as a result of the erroneous perception that θ_1 and θ_2 are correlated). Thus, while supply will be almost entirely a function of θ_2, demand will be a function of θ_1. Since these two state variables are objectively independent, supply and demand responses to external shocks will be virtually orthogonal. This pattern of fluctuations is impossible in our basic model (which subsumes REE as a special case).

5.3.2 Partially Informed Consumers

So far, we have assumed that consumers have no information about the state (other than what they can learn from prices). Since equilibrium prices are fully revealing, this lack of information is irrelevant if consumers have rational expectations. We will now see that this irrelevance no longer holds when consumers are imperfectly discerning.

Extend the model as follows. For every consumer type given by a DAG G, there is a distinct variable w_G which represents a noisy private signal of θ that type-G consumers observe. These consumers admit w_G as a variable in their causal model, such that $R(w_G)$ is contained in the set of nodes that represent θ, and w_G itself is not a parent of any other node. Thus, the consumer understands that w_G is merely a signal of the exogenous state variables, and therefore not a (direct or indirect) cause of any other variable. For instance, G can be

$$\begin{array}{ccccccc} & & w_G & & & & \\ & & \uparrow & \nwarrow & & & \\ p & \leftarrow & \theta_1 & \rightarrow & \theta_2 & \rightarrow & \phi \end{array} \tag{21}$$

Extend μ to be a joint distribution over p, ϕ, and the exogenous variables, θ and $w = (w_G)_{G\in\mathcal{G}}$. Thus, when the market price is p, a type-G consumer uses the conditional subjective belief $\mu_G(\phi \mid p, w_G)$ to predict the ripoff. When G is given by (21), we can see that the consumer infers θ_1 from the market price p, and then uses both this inference and his knowledge of w_G to form a conditional belief over θ_2, and hence ϕ.[5]

The basic result that interior equilibrium fully reveals θ continues to hold in this extended model. That is, in an interior equilibrium h, $\theta' \neq \theta$ implies

[5]Note that p and w_G do not form a clique in G. Consequently, $\mu_G(\phi \mid p, w_G)$ need not satisfy the unbiased-on-average property, even if G is perfect (see Spiegler (2020b)). Since we do not use this property in the sequel, we also drop the assumption that $\mathcal{G}$ consists of perfect DAGs.

$h(\theta', w') \neq h(\theta, w)$. The proof is the same as in the case of Proposition 1, and therefore omitted. However, the next result establishes that equilibrium prices can *also* reflect consumers' private information, even when we hold θ fixed.

The result relies on the following notion of path blocking (in the spirit of similar definitions in the literature on graphical probabilistic models — see Pearl (2009)). We say that a set of nodes M *blocks all non-directed paths* between nodes $i, j \notin M$ if in the non-directed version of G (in which we ignore the direction of links), every path between i and j passes through some $k \in M$. For example, in the DAG (21), $\{\theta_2\}$ blocks all non-directed paths between w_G and ϕ, whereas $\{\theta_1\}$ does not.

Proposition 8 *Suppose $\mathcal{G}$ is a set of DAGs that includes non-rational consumer types. Moreover, suppose that for every non-rational $G \in \mathcal{G}$, $R(p)$ does not block all non-directed paths between w_G and ϕ. Then, assuming the interior equilibrium h does not coincide with REE, there must be a state θ and signals w, w', such that $h(\theta, w) \neq h(\theta, w')$.*

Thus, the presence of imperfectly discerning consumers can create excessive price fluctuations, in the sense that equilibrium prices respond to factors beyond economic fundamentals. (This is distinct from the observation, made in Section 3, that the *range* of equilibrium prices is narrower that in REE.) Specifically, they can reflect consumers' private information, whereas this would not happen if consumers had rational expectations. For instance, when G is given by (21), equilibrium prices respond to w_G because consumers do not infer θ_2 from prices.

6 Related Literature

Our paper contributes to a small literature on competitive markets with asymmetric information, in which consumers' beliefs deviate from rational expectations. Eyster and Piccione (2013) study dynamic competitive markets for financial securities without short-selling. Their traders have diversely coarse models of an exogenous state that determines the interest rate and dividend. Our baseline behavioral model of Section 2 generates similar behavior as, since interior equilibrium is fully revealing, agents behave as if they have a coarse perception of the state space. Using a similar model in which states evolve in continuous time while trading periods are discrete, Steiner and Stewart (2015) show that as the duration of trading periods vanishes, equilibrium asset prices become measurable with respect to the meet of the partitions of the state space that traders' subjective models induce.

Apart from the different economic settings — a dynamic financial market vs. a static consumer market — the main difference between these works and the present model is that traders in the Eyster-Piccione and Steiner-Stewart models do not draw any inferences from current prices, whereas the heart of our model is consumers' imperfect attempt to infer latent variables from current prices.

Eyster et al. (2019) study patterns of speculative trade in a financial-market model, in which traders exchange risky and riskless assets after observing public and private signals about the risky asset's value. Building on Eyster and Rabin (2005), they define "cursed expectations" competitive equilibrium, where traders beliefs are a mixture of the rational-expectations benchmark (which involves flawless inference from equilibrium prices) and fully coarse beliefs (which involve no inference from equilibrium prices).[6]

Piccione and Rubinstein (2003) analyze a simple example of a dynamic, complete-information competitive market, in which producers differ in their ability to perceive temporal price patterns, and hence in their ability to predict market prices when making costly production decisions. They demonstrate "the existence of equilibrium fluctuations that are unrelated to fundamentals..." (Piccione and Rubinstein (2003, p. 218)), thus offering a precursor to Section 5.3 in our paper.

Our model fits naturally into the Behavioral Industrial Organization literature (see Spiegler (2011) for a textbook treatment and Heidhues and Kőszegi (2018) for a review). A prominent strand in this literature analyzes market competition when firms use hidden charges as part of their competitive strategy. Most of this

[6] The partial-cursedness approach, by which beliefs are a convex combination of rational and coarse beliefs, is trivial in our model. As long as $\mathcal{M}$ includes the rational and fully coarse types, admitting partially cursed types would have no effect. The reason is that their hidden-charge estimates would be bounded between the rational and fully coarse types' estimates, and so they would never determine the market-clearing price.

literature (going back to Gabaix and Laibson (2006)) has assumed that consumers are unaware of the hidden charges and evaluate market alternatives as if they do not exist. In Spiegler (2006), products have many dimensions, and consumers base their product evaluation on a single, randomly drawn dimension.

A few exceptions have examined market models in which consumers have a coarse understanding of what drives market prices. Spiegler (2011, Ch. 8) synthesizes examples of bilateral-trade models with adverse selection (extracted from Eyster and Rabin (2005), Jehiel and Koessler (2008), and Esponda (2008)), in which the uninformed party has a coarse perception of price formation.[7] At the extreme, this agent's belief is entirely coarse, such that he correctly perceives average prices without having any understanding of how they depend on the state of Nature.

In a similar vein, Murooka and Yamashita (2023) study a bilateral-trade setting in which, with some probability, the buyer believes that product quality is independent of the price in which it is traded. Ispano and Schwardmann (2023) study a model in which consumers fail to understand that only high-quality firms have an incentive to disclose their quality. Schumacher (2023) studies a model in which firms sell a superior product that only charges a base price and an inferior product that also includes an add-on component. Coarse consumers know the average add-on charge across products but incorrectly believe it is independent of the product type. Thus, as in our model, consumers are aware of hidden charges but have limited ability to predict them based on their information. Antler (2023) analyzes a model of multilevel marketing and pyramid schemes, where a principal exploits a network of agents having coarse expectations regarding the network formation process. These models are all game-theoretic, and they lack this paper's key feature, namely consumers' heterogeneous ability to draw inferences from market prices.

[7] In Esponda (2008), as in the present paper, consumers' assessment of firms' types is based on the empirical distribution of *active* firms at the equilibrium price. There is no aggregate uncertainty in Esponda's model and therefore no need to ask how consumers infer an aggregate state from equilibrium prices.

7 Conclusion

The standard theory of competitive markets gives a central role to equilibrium prices' ability to aggregate information. This property, however, relies on market participants' ability to decipher the price signal. This paper developed a new model of a competitive market in which consumers differ in this regard, and explored the theoretical implications of this "cognitive friction" for the way equilibrium outcomes respond to exogenous shocks.

The paper's methodological contributions consist of our novel supply function (arising from firms' differential ability to realize state-dependent latent profit), our model of how consumers infer latent quantities from market-clearing prices, and the tractable "Bellman" characterization of interior equilibrium. The paper's substantive conclusions include the deviation of equilibrium price components from their rational-expectations benchmarks, and the demonstration that market outcomes respond to exogenous variables in ways that are impossible in REE.

A key economic insight of our paper is that a change that induces an increase in demand in some states (e.g., introducing a new cognitive consumer type) leads to higher total prices and lower ripoffs across states. The changes can be strict in states that did not see the original shock. This "contagion" effect arises from a combination of two fundamental features: The adverse-selection element of the market environment, and the fact that imperfectly discerning consumers' estimates of the ripoff/quality in a given state also reflect its value in other states. The effect is robust, in the sense that it would appear in models having these two features, even if they differ from ours in specifics.

We conclude the paper with a discussion of two of our modeling procedures.

The homogenous-preference limit

Our analysis has focused on the $\varepsilon \to 0$ limit. A criticism of this approach is that on our full-revelation result (Proposition 1) relies on preference heterogeneity, yet our equilibrium analysis assumes this heterogeneity is negligible. However, our procedure is analogous to a common practice in the repeated games literature (e.g., Mailath and Samuelson (2006)): Assuming players apply a discount factor δ to future payoffs and then studying equilibria in the $\delta \to 1$ limit. The justification for that procedure is that while discounting captures a key behavioral motive in long-term interactions, assuming this motive is weak enables a clean understanding of the logic of long-run cooperation. Likewise in our context, preference heterogeneity allows equilibrium prices to reflect supply-side responses to external shocks. This ensures that consumers' task of deciphering equilibrium prices is meaningful. At the same time, weak taste heterogeneity enables us to focus on consumers' diverse forecasts of latent product features.

The distribution of firm types
The uniform distribution of π facilitates our analysis by generating linear supply. The Bellman equation (12) arises from the combination of two equations: The indifference condition for the *marginal* firm type $\pi^*(\theta, h(\theta))$, and consumers' maximal willingness to pay for the product in state θ (which is equal to $h(\theta)$ in the $\varepsilon \to 0$ limit). The latter equation involves the *average* active firm type $\bar{\pi}(\theta', h(\theta'))$ in various states θ'. When π is uniformly distributed, $\pi^*(\theta', h(\theta'))$ and $\bar{\pi}(\theta', h(\theta'))$ are linearly related, which enables us to conveniently substitute one for the other. This also ensures that (12) defines a contraction mapping.

If π does not obey a uniform distribution, the tractable linear structure of (12) is lost. However, as long as the firm type distribution is *log-concave* and has mean $\frac{1}{2}$, the equilibrium equations continue to define a contraction mapping, ensuring a unique interior equilibrium under the same condition on S.[8] This in turn implies that Proposition 3 extends to this case, too.

References

[1] Akerlof, G. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, 84, 488-500.

[2] Antler, Y. (2023): "Multilevel marketing: pyramid-shaped schemes or exploitative scams?", *Theoretical Economics*, 18, 633-668.

[3] Antler, Y. and R. Spiegler (2024): "Competitive Markets with Imperfectly Discerning Consumers", arXiv preprint arXiv:2409.14885.

[8] If the mean is not $\frac{1}{2}$, then the bound on $S(\theta)/S(\theta')$ needs to be modified.

[4] Eliaz, K., Spiegler, R., and Weiss, Y. (2021): "Cheating with models", *The American Economic Review: Insights* 3, 4, 417-434.

[5] Ellis, A. and Piccione, M. (2017): "Correlation Misperception in Choice", *The American Economic Review*, 107, 4, 1264-92.

[6] Esponda, I. (2008): "Behavioral Equilibrium in Economies with Adverse Selection", *The American Economic Review*, 98,4, 1269-91.

[7] Ettinger, D. and Jehiel, P. (2010): "A Theory of Deception", *American Economic Journal: Microeconomics*, 2, 1, 1-20.

[8] Eyster, E. and Rabin, M. (2005): "Cursed Equilibrium", *Econometrica*, 73, 1623–1672.

[9] Eyster, E. and Piccione, M. (2013): "An Approach to Asset Pricing Under Incomplete and Diverse Perceptions", *Econometrica*, 81, 4, 1483-1506.

[10] Eyster, E., M. Rabin, and D. Vayanos (2019): "Financial Markets where Traders Neglect the Informational Content of Prices", *Journal of Finance* 74, 1, 371-399.

[11] Gabaix, X. and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets", *Quarterly Journal of Economics* 121, 505-540.

[12] Hayek, F. A. (1945): "The Use of Knowledge in Society", *The American Economic Review*, 35, 4, 519-530.

[13] Heidhues, P. and Kőszegi, B. (2018): "Behavioral Industrial Organization", In *Handbook of Behavioral Economics*, 1, 6, 517–612.

[14] Heidhues, P., B. Kőszegi, and T. Murooka (2016): "Exploitative Innovation", *American Economic Journal: Microeconomics* 8, 1-23.

[15] Heidhues, P., B. Kőszegi, and T. Murooka (2017): "Inferior Products and Profitable Deception," *Review of Economic Studies* 84, 323-356.

[16] Ispano, A. and Schwardmann (2023): "Cursed Consumers and the Effectiveness of Consumer Protection Policies", *Journal of Industrial Economics,* 71, 2, 407-440.

[17] Jehiel, P. (2005): "Analogy-Based Expectation Equilibrium", *Journal of Economic Theory*, 123, 81–104.

[18] Jehiel, P. and Koessler, F. (2008): "Revisiting Games of Incomplete Information with Analogy-Based Expectations", *Games and Economic Behavior*, 62, 533–557.

[19] Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations: Long-Run Relationships*, New York, Oxford university press.

[20] Mailath, G. and L. Samuelson (2020): "Learning under Diverse World Views: Model-Based Inference", *American Economic Review* 110, 1464-1501.

[21] Murooka, T. and Yamashita, T. (2023): "Optimal Trade Mechanisms with Adverse Selection and Inferential Naivety", *Mimeo.*

[22] Müller, A. and Stoyan, D. (2002): *Comparison methods for stochastic models and risks*, John Wiley & Sons, Chichester.

[23] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*, Cambridge, Cambridge University Press.

[24] Piccione, M. and Rubinstein, A. (2003): "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns", *Journal of the European Economic Association,* 1, 212–223.

[25] Radner, R. (1979): "Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices", *Econometrica*, 47, 3, 655-678.

[26] Schumacher, H. (2023): "Competitive Markets, Add-On Prices, and Boundedly Rational Expectations", *Mimeo.*

[27] Steiner, J. and C. Stewart (2015). "Price Distortions under Coarse Reasoning with Frequent Trade", *Journal of Economic Theory* 159, 574-595.

[28] Spiegler, R. (2006): "Competition over Agents with Boundedly Rational Expectations", *Theoretical Economics* 1, 207-231

[29] Spiegler, R. (2016): "Bayesian Networks and Boundedly Rational Expectations", *Quarterly Journal of Economics*, 131, 1243–1290.

[30] Spiegler, R. (2020a): "Behavioral Implications of Causal Misperceptions", *Annual Review of Economics*, 12, 81-106

[31] Spiegler, R. (2020b): "Can Agents with Causal Misperceptions be Systematically Fooled?", *Journal of European Economic Association* 18, 583-617.

Appendix: Omitted Proofs

Proposition 2

Equation (12) is an immediate consequence of (9) and (10). By definition, $\overline{\phi}(\theta) \in [\frac{1}{2}S(\theta), S(\theta)]$ for every θ. Thanks to the $\frac{1}{2}$ coefficient on the R.H.S of (12), it is then clear that the equation defines a contraction mapping over a compact and convex Euclidean space. By the contraction mapping theorem, it has a unique solution. This also uniquely pins down the values of $h(\theta)$ and $\pi^\star(\theta, h(\theta))$ for every θ.

We now obtain the bounds on $\overline{\phi}(\theta)$. Equation (12) implies $2\overline{\phi}(\theta) \leq \max_\theta S(\theta) - \Delta + \max_{\theta'} \overline{\phi}(\theta')$ for every θ. Therefore, $2\max_\theta \overline{\phi}(\theta) \leq \max_\theta S(\theta) - \Delta + \max_\theta \overline{\phi}(\theta)$, such that $\max_\theta \overline{\phi}(\theta) \leq S^{\max} - \Delta$. Likewise, (12) implies $2\overline{\phi}(\theta) \geq \min_\theta S(\theta) - \Delta + \min_{\theta'} \overline{\phi}(\theta')$ for every θ. Therefore, $2\min_\theta \overline{\phi}(\theta) \geq \min_\theta S(\theta) - \Delta + \min_\theta \overline{\phi}(\theta)$, such that $\min_\theta \overline{\phi}(\theta) \geq S^{\min} - \Delta$.

It remains to show that $\pi^\star(\theta, h(\theta)) \in (0, 1)$ for every θ — i.e., the equilibrium is interior. Equivalently, we need to show that for every θ, $\frac{1}{2}S(\theta) < \overline{\phi}(\theta) < S(\theta)$. Assume $\overline{\phi}(\theta) \geq S(\theta)$ for some θ. Then, (12) implies

$$S(\theta) - \Delta + \min_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta_M)\overline{\phi}(\theta') \geq 2S(\theta)$$

Since $\overline{\phi}(\theta') \leq S^{\max} - \Delta$ for every θ', $S^{\max} - S(\theta) \geq 2\Delta$. By definition, this means $S^{\max} - S^{\min} \geq 2\Delta$, contradicting (11). Therefore, $\overline{\phi}(\theta) < S(\theta)$ for every θ. Now assume $\overline{\phi}(\theta) \leq \frac{1}{2}S(\theta)$ for some θ. Then, (12) implies

$$S(\theta) - \Delta + \min_{M \in \mathcal{M}} \sum_{\theta'} \mu(\theta' \mid \theta_M)\overline{\phi}(\theta') \leq S(\theta)$$

Since $\overline{\phi}(\theta') \geq S^{\min} - \Delta$ for every θ', $S^{\min} - 2\Delta \leq 0$, contradicting (11). ■

Proposition 4

Suppose $\mathcal{M} = \{M\}$, where M is arbitrary. Take an expectation of both sides of (12) with respect to μ. Then,

$$2\sum_{\theta} \mu(\theta)\overline{\phi}(\theta) = \sum_{\theta} \mu(\theta)S(\theta) - \Delta + \sum_{\theta'}\sum_{\theta} \mu(\theta)\mu(\theta' \mid \theta'_M = \theta_M)\overline{\phi}(\theta')$$

As we observed above,

$$\sum_{\theta} \mu(\theta)\mu(\theta' \mid \theta'_M = \theta_M) = \mu(\theta')$$

Therefore, the expected Bellman equation becomes

$$2\sum_{\theta} \mu(\theta)\overline{\phi}(\theta) = \sum_{\theta} \mu(\theta)S(\theta) - \Delta + \sum_{\theta} \mu(\theta)\overline{\phi}(\theta)$$

such that $\sum_{\theta} \mu(\theta)\overline{\phi}(\theta) = \bar{S} - \Delta$, which is the REE level. Thus, for any singleton $\mathcal{M}$, the expected ripoff in interior equilibrium coincides with its REE level. By Proposition (3), for any $\mathcal{M}' \supset \{M\}$, the expected ripoff in interior equilibrium is weakly lower in each state than under $\{M\}$. It follows that the ex-ante expected ripoff is weakly below the REE level $\bar{S} - \Delta$.

Recall that by (3), $\overline{\phi}(\theta) \geq \frac{1}{2}S(\theta)$ for every θ. Therefore, the ex-ante expected ripoff cannot fall below $\frac{1}{2}\bar{S}$. We now construct primitives $\Theta, \mu, S, \mathcal{M}$ that satisfy $\sum_{\theta} \mu(\theta)S(\theta) = \bar{S}$ and condition (11), and show that the expected ripoff in the interior equilibrium under this specification is arbitrarily close to $\frac{1}{2}\bar{S}$. Let $\theta_i \in \{0, 1\}$ for every $i = 1, ..., n$, where n is arbitrarily large. Let e_i denote the state θ for which $\theta_i = 0$ and $\theta_j = 1$ for all $j \neq i$. distribution μ as follows: $\mu(0, ..., 0) = \alpha$ and $\mu(e_i) = (1 - \alpha)/n$ for every i. We will pin down α below. Define the function S as follows: $S(0, ..., 0) \gtrapprox 2\Delta$, and $S(e_i) \lessapprox 4\Delta$ for every $i = 1, ..., n$. Fix α such that $\bar{S} \approx \alpha \cdot 2\Delta + (1-\alpha) \cdot 4\Delta$, i.e., $\alpha \approx (4\Delta - \bar{S})/2\Delta$. Finally, let $\mathcal{M}$ consist of the following types: the rational type $\{1, ..., n\}$, and the coarse types $\{i\}$ for every $i = 1, ..., n$.

Guess an equilibrium in which the rational type buys the product in the state $(0, ..., 0)$; and the coarse type $\{i\}$ buys the product in the state e_i, for every $i = 1, ..., n$. The Bellman-like equations are thus reduced to

$$2\overline{\phi}(0, ..., 0) = S(0, ..., 0) - \Delta + \overline{\phi}(0, ..., 0)$$

and

$$2\overline{\phi}(e_i) = S(e_i) - \Delta + \frac{\alpha}{\alpha + \frac{1-\alpha}{n}}\overline{\phi}(0, ..., 0) + \frac{\frac{1-\alpha}{n}}{\alpha + \frac{1-\alpha}{n}}\bar{\phi}(e_i)$$

for every $i = 1, ..., n$. It follows that $\overline{\phi}(0, ..., 0) = S(0, ..., 0) - \Delta \approx \Delta$; and as $n \to \infty$, the solution to the remaining equations is $\overline{\phi}(e_i) \approx 2\Delta$. It is straightforward to confirm that the types that buy the product in each state have the lowest ripoff estimate in that state. The ex-ante equilibrium ripoff is approximately $\alpha \cdot \Delta + (1-\alpha) \cdot 2\Delta \approx \bar{S}/2$. ■

Proposition 5

A rational type's willingness to pay in state θ is $v^* - \overline{\phi}(\theta)$. Therefore, $h(\theta) \geq v^* - \overline{\phi}(\theta)$ for every θ. Plugging the upper bound on $\overline{\phi}(\theta)$ given by Proposition 2, we obtain

$$h(\theta) \geq v^* - (S^{\max} - \Delta) = 2v^* - c - S^{\max}$$

Now consider the state θ for which $S(\theta) = S^{\min}$. The rational type's willingness to pay in this state is $v^* - \overline{\phi}(\theta)$. The willingness to pay of an arbitrary type M is

$$v^* - \sum_{\theta'} \mu(\theta' \mid \theta_M)\overline{\phi}(\theta') \tag{22}$$

Guess a solution to (12) for which $\overline{\phi}(\theta) = S^{\min} - \Delta$. Then, the rational type's willingness to pay in state θ is $v^* - (S^{\min} - \Delta)$. By the lower bound on $\overline{\phi}(\theta)$ given by Proposition 2, this expression is weakly above (22) for any M. Then, guessing that the rational type has the highest willingness to pay in θ is consistent with a solution to (12) in this state, and it gives $h(\theta) = v^* - (S^{\min} - \Delta)$. The remaining equations in (12) for all other states deliver a unique solution, hence the guess is consistent with the entire system of equations. It follows that when $\mathcal{M}$ contains a rational type, the upper bound on equilibrium prices given in part (ii) is binding. ■

Proposition 6

The following lemma draws a simple conclusion from the definition of conditionally increasing distributions.

Lemma 2 *Suppose μ is conditionally increasing. Then, for every disjoint subsets $A, B \subset \{1, ..., n\}$, $\theta'_B > \theta_B$ implies that $\mu(\theta_A \mid \theta'_B)$ first-order stochastically dominates $\mu(\theta_A \mid \theta_B)$.*

Proof. Enumerate the variables in A and B as follows: $A = \{1, ..., m_1\}$, $B = \{\ell, ..., m_2\}$. If μ is conditionally increasing, then: $\mu(\theta_1 = 1 \mid \theta_B)$ is increasing in θ_B; $\mu(\theta_2 = 1 \mid \theta_1, \theta_B)$ is increasing in θ_1, θ_B; and extending the same logic, $\mu(\theta_k = 1 \mid \theta_1, ..., \theta_{k-1}, \theta_B)$ is increasing in $\theta_1, ..., \theta_{k-1}, \theta_B$, for every $k \leq m_1$. This means that an increase in θ_B leads to a first-order stochastic upward shift in the conditional distribution $\mu(\theta_A \mid \theta_B)$. ■

Let us guess (and verify later) that $\bar{\phi}(\theta)$ is decreasing in θ in the unique interior equilibrium. We first show that if $M^\star \in \arg\min_{M'} E_{M'}(\phi \mid \theta)$ for some θ, then $M^\star = I(\theta)$. Assume there is θ for which this is not the case. Suppose there is $i \notin M^\star$ such that $\theta_i = 1$. For every θ', $\mu(\theta' \mid \theta_{M^*})$ can be written as

$$\mu(\theta'_i = 1 \mid \theta_{M^*})\mu(\theta'_{-(M^* \cup \{i\})} \mid \theta_{M^*}, \theta'_i = 1) + \mu(\theta'_i = 0 \mid \theta_{M^*})\mu(\theta'_{-(M^* \cup \{i\})} \mid \theta_{M^*}, \theta'_i = 0)$$

By (14), $\mu(\theta'_{-(M^* \cup \{i\})} \mid \theta_{M^*}, \theta'_i = 1)$ first-order stochastically dominates $\mu(\theta'_{-(M^* \cup \{i\})} \mid \theta_{M^*}, \theta'_i = 0)$. Therefore, $\mu(\theta' \mid \theta_{M^*}, \theta_i = 1)$ first-order stochastically dominates $\mu(\theta' \mid \theta_{M^*})$. It follows that

$$\sum_{\theta'} \mu(\theta' \mid \theta_{M^* \cup I(\theta)})\bar{\phi}(\theta') < \sum_{\theta'} \mu(\theta' \mid \theta_{M^*})\bar{\phi}(\theta')$$

An analogous argument can be applied to shrinking M^* by removing nodes i for which $\theta_i = 0$, implying

$$\sum_{\theta'} \mu(\theta' \mid \theta_{M^* \cap I(\theta)})\bar{\phi}(\theta') < \sum_{\theta'} \mu(\theta' \mid \theta_{M^*})\bar{\phi}(\theta')$$

It follows that $E_{M^\star}(\phi \mid \theta) > E_{I(\theta)}(\phi \mid \theta)$, contradicting $M^\star \in \arg\min_{M'} E_{M'}(\phi \mid \theta)$.

Having found the cognitive type with the highest willingness to pay in every state, we can use equation (12) to pin down the expected ripoff in every state and verify our initial guess. In state $(1, ..., 1)$, the equation is reduced to $\bar{\phi}(1, ..., 1) = S(1, ..., 1) - \Delta$, which pins down $\bar{\phi}(1, ..., 1)$. Next, we move to the equations for all states θ in which $\theta_i = 0$ for *exactly one* component i. Such an equation expresses $\bar{\phi}(\theta)$ in terms of $\bar{\phi}(\theta)$ and $\bar{\phi}(1, ..., 1)$, and so the expected ripoff in all of these states is pinned down. Next, we move to the equations for states θ for which $\theta_i = 0$ for *exactly two* components i, and so forth. We can then use (12) to obtain (14). This functional equation confirms our guess since $S(\theta)$ is decreasing in θ. ■

Proposition 7

By Proposition 6, type M^θ is the one with the lowest average-ripoff estimate in state θ, under both S and J. Under S, the welfare loss of a type-M^θ consumer in state θ is

$$L(\theta) = \overline{\phi}(\theta) - \frac{\sum_{\hat{\theta} \geq \theta} \mu(\hat{\theta} | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})\overline{\phi}(\hat{\theta})}{\sum_{\hat{\theta} \geq \theta} \mu(\hat{\theta} | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})}$$

By (14), $\overline{\phi}(\theta) = \overline{\phi}_J(\theta)$ for every $\theta > \theta^*$, and so consumers of type M^θ, $\theta > \theta^*$, incur the same expected loss under $S(\cdot)$ and $J(\cdot)$.

Consider state $\theta < \theta^*$. When we move from S to J, $L(\theta)$ increases by

$$\frac{\sum_{\hat{\theta} \geq \theta} \mu(\hat{\theta} | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})k(\hat{\theta})}{\sum_{\hat{\theta} \geq \theta} \mu(\hat{\theta} | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})} - k(\theta) \tag{23}$$

Using the definition of $k(\theta)$ and rearranging (14), we obtain

$$\overline{\phi}_J(\theta) = \frac{S(\theta) - \Delta + \sum_{\hat{\theta} > \theta} \mu(\hat{\theta} | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})(\overline{\phi}(\hat{\theta}) - k(\hat{\theta}))}{2 - \mu(\hat{\theta} = \theta | \hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})}$$

It follows that

$$k(\theta) = \overline{\phi}(\theta) - \overline{\phi}_J(\theta) = \frac{\sum_{\hat{\theta}>\theta} \mu(\hat{\theta}|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})k(\hat{\theta})}{2 - \mu(\hat{\theta} = \theta|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})}$$

Plugging into (23) yields

$$\frac{\sum_{\hat{\theta}\geq\theta} \mu(\hat{\theta}|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})k(\hat{\theta})}{\sum_{\hat{\theta}\geq\theta} \mu(\hat{\theta}|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})} - \frac{\sum_{\hat{\theta}>\theta} \mu(\hat{\theta}|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})k(\hat{\theta})}{2 - \mu(\hat{\theta} = \theta|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})} \tag{24}$$

Note that $2 - \mu(\hat{\theta} = \theta|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)}) > 1 > \sum_{\hat{\theta}\geq\theta} \mu(\hat{\theta}|\hat{\theta}_{I(\hat{\theta})} = \theta_{I(\theta)})$. By (14), $\overline{\phi}(\theta) = \overline{\phi}_J(\theta)$ for every $\theta > \theta^*$. Hence, (14) also implies that $k(\theta^*) > 0$. Applying (14) again, we obtain that $k(\tilde{\theta}) > 0$ for every $\tilde{\theta} < \theta'$. We can conclude that (24) is strictly positive, and so $L(\theta)$ increases by a strictly positive amount. Finally, note that $k(\theta) > 0$ implies that the volume of trade in state θ increases. Thus, not only do individual consumers who buy in state $\theta < \theta'$ experience a greater welfare loss, but there is also a larger mass of consumers who incur this loss. ■

Lemma 1

Since G is perfect, there is an equivalent DAG G' (in the sense that $\mu_G \equiv \mu_{G'}$) in which p is an ancestral node (see Spiegler (2020a,b)). Therefore, we can regard p as ancestral, without loss of generality. If there is a direct link $p \to \phi$, then (p, ϕ) form a clique in G, and hence perfection implies $\mu_G(\phi, p) \equiv \mu(\phi, p)$, hence $\mu_G(\phi \mid p) \equiv \mu(\phi \mid p)$ whenever $\mu(p) > 0$. Since μ is fully revealing, $\mu(\phi \mid p) \equiv \mu(\phi \mid \theta_\mu(p))$. In this case, the unique transition matrix β for which (19) holds is $\beta(\theta \mid \theta) \equiv 1$.

Now suppose there is no path from ϕ to p. Then, ϕ is independent of p according to μ_G, such that $\mu_G(\phi \mid p) \equiv \mu(\phi)$. We can thus rewrite

$$\mu_G(\phi \mid p) = \sum_{\theta'} \mu(\theta')\mu(\phi \mid \theta')$$

In this case, the unique transition matrix β for which (19) holds is $\beta(\theta' \mid \theta) \equiv \mu(\theta')$.

Now suppose there is a path from ϕ to p, but the two nodes are not directly related. Note that all nodes along all paths from ϕ to p represent θ variables. Let

C denote the set of nodes to which p sends direct links, and let D denote the set of nodes that send direct links into ϕ. Then,

$$\mu_G(\phi \mid p) = \sum_{\theta_C} \mu(\theta_C \mid p) \sum_{\theta_D} \mu_G(\theta_D \mid \theta_C)\mu(\phi \mid \theta_D)$$

Since μ is fully revealing, $\mu(\theta_M \mid p)$ assigns probability one to the projection of $\theta_\mu(p)$ on the variables represented by C, denoted $\theta_C(p)$.

Therefore, $\mu_G(\phi \mid p)$ is equal to

$$\sum_{\theta_D} \mu_G(\theta_D \mid \theta_C(p))\mu(\phi \mid \theta_D) = \sum_{\theta_D} \mu_G(\theta_D \mid \theta_C(p)) \sum_{\theta'} \mu(\theta' \mid \theta_D)\mu(\phi \mid \theta')$$

Denote

$$\beta(\theta' \mid \theta) = \sum_{\theta''_D} \mu_G(\theta''_D \mid \theta_C) \sum_{\theta'} \mu(\theta' \mid \theta_D)$$

The R.H.S of this equation is pinned down by G and μ. Thus, it is the unique transition matrix for which (19) holds. Moreover, the property that μ is an invariant distribution of β is an immediate consequence of (18). ■

Proposition 8

Assume the contrary — i.e., $\mathcal{G}$ satisfies the premises of the result, and yet the interior equilibrium h is purely a function of θ. By assumption, $h(\theta)$ deviates from the REE price in some θ. In that state,

$$h(\theta) = v^* - \int_\phi \mu_G(\phi \mid h(\theta), w_G)\phi$$

for all realizations of w_G, since by assumption, h is unresponsive to w_G given θ. By assumption, $R(p)$ does not block all paths in G between w_G and ϕ. For generic μ, this means that type G's belief over ϕ conditional on $h(\theta)$ is not invariant to w_G, hence his willingness to pay varies with w_G given θ. As a result, h cannot be constant in w_G given θ, a contradiction. ■