

# Strategic Mistakes

Joel P. Flynn\*

MIT

Karthik A. Sastry†

MIT

August 14, 2021

## Abstract

To study the equilibrium implications of imperfect optimization, we introduce a model of costly control in continuum-player games in which agents interact via an aggregate of the actions of others. We find primitive conditions such that equilibria exist, are unique, are efficient, and feature monotone comparative statics for action distributions, aggregates, and the size of agents' mistakes. We use our results to provide robust equilibrium predictions in a class of generalized beauty contests, which we apply to study the implications of imperfect optimization for financial speculation, price-setting, and the business cycle. We contrast our model with the mutual information model (Sims, 2003), which in the same games can produce non-unique predictions and non-monotone comparative statics.

---

\*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: [jpflynn@mit.edu](mailto:jpflynn@mit.edu).

†MIT Department of Economics, 50 Memorial Drive, Cambridge, MA 02142. Email: [ksastry@mit.edu](mailto:ksastry@mit.edu).

An earlier version of this paper was entitled “Strategic Mistakes in Large Games.” We thank Daron Acemoglu, George-Marios Angeletos, Oğuzhan Çelebi, Roberto Corrao, Drew Fudenberg, Chen Lian, Nicolas Minudri, Stephen Morris, George Nikolakoudis, Chris Sims, Alp Simsek, John Sturm, Iván Werning, Alexander Wolitzky and participants in the MIT Macro Lunch, MIT Theory Lunch, MIT Behavioral Economics Lunch, and MIT Sloan Finance Lunch for helpful comments. First posted version: July 29, 2020.

# 1 Introduction

Economic agents commonly make mistakes that affect others. Consider a firm choosing its price or production level to maximize profits, taking into account projected demand for their product, current input prices, and choices by other firms. The complexity of real firms' decision-making processes makes clear that, even though the problem and its parameters are well-defined and an ideal solution surely exists, determining that solution is rather difficult in practice. Any resulting deviation from the ideal point will in turn affect all other firms' incentives to invest and to rein in their own mistakes when doing so. Observed production choices therefore arise from a process of *strategic mistakes*, or the combination of imperfect optimization and strategic interaction.

Previous studies have applied such a notion of strategic mistakes to a number of settings in macroeconomics and finance. Our hypothetical firm's problem is a core ingredient in models introduced by Woodford (2003), Maćkowiak and Wiederholt (2009), Angeletos and La'O (2010, 2013), Benhabib, Wang, and Wen (2015), and Chahrour and Ulbricht (2019) to study how “noisy” decision-making shapes various properties of business cycles. The notion of strategic mistakes is also an important motivation for the *linear beauty contest* framework which is analyzed by Morris and Shin (2002), Angeletos and Pavan (2007), Bergemann and Morris (2013), and Huo and Pedroni (2020) and commonly applied to study business cycles as well as other topics like financial speculation and the Keynesian cross.<sup>1</sup> Each of these studies has either restricted to specific parametric models of informational or behavioral frictions to allow for analytical comparative statics analysis, or studied less parametric models (e.g., the mutual information model of Sims, 2003) numerically to explore richer behaviors. In the second category, models of unrestricted information acquisition like rational inattention with mutual-information costs have an appealing generality but admit few comparative statics results, particularly in equilibrium.<sup>2</sup>

This paper introduces a model of non-parametric stochastic choice in continuum-player games to study equilibrium properties with imperfect optimization in a way that is tractable, robust to functional form restrictions on payoffs and the stochastic environment, and able to capture rich behavioral patterns. In the model, agents face a problem of *costly control*: conditional on their conjecture for fundamentals and others' actions, each agent picks a stochastic choice pattern that balances high expected payoff with a cost that penalizes playing

---

<sup>1</sup>For the first, see Allen, Morris, and Shin (2006). For the second, in a dynamic setting, see Garcia-Schmidt and Woodford (2019), Gabaix (2019), and Angeletos and Lian (2018).

<sup>2</sup>In this context, we refer specifically to the class of posterior-separable cost functionals defined and studied by Caplin and Dean (2015) and Denti (2019), which includes mutual information. The strongest equilibrium results in these settings are results in Angeletos and Sastry (2019) and Hébert and La'O (2020) showing sufficient conditions for efficiency and the lack of non-fundamental volatility in equilibrium.

too precisely. Our key simplifying assumption is that this control problem is state-separable. This assumption makes our model applicable to scenarios in which agents’ difficulty is purely in identifying the “right thing to do” in a given state of the world, and not confusion regarding the identity of the state. We argue that this perspective is justified in a number of the previously mentioned applications, in which agents could be aware of state variables like interest rates or asset prices but still optimize imperfectly conditional on them.

Our main results provide sufficient conditions for equilibrium existence, uniqueness, and monotonicity of distributions, aggregates and mistakes. We show how these results can be used to study the equilibrium implications of strategic mistakes in a manner that is robust to the extent or character of the stochastic choice friction, applicable to games that cannot be tractably studied using common approaches like parametric signal extraction, and capable of making rich aggregate and cross-sectional predictions. In example applications, we derive mappings from primitive properties of an abstract financial market and monetary economy to claims such as: aggregate investment is monotone in the strength of fundamentals, the cross-firm price distribution is monotone in the money supply, and cross-firm price dispersion is pro-cyclical.

**Framework and Main Results.** In the model, a continuum of agents take a continuous action. Their payoffs depend on their own action, an exogenous state and a one-dimensional aggregate of the cross-sectional distribution of others’ actions. Agents choose a stochastic choice rule describing the non-parametric distribution of actions they play in each state of the world to maximize expected utility net of a cost that penalizes strategies that are more sharply peaked. Our specific class of cost functional can be interpreted as an expected-utility generalization of the Additive Perturbed Utility preferences studied by [Fudenberg, Iijima, and Strzalecki \(2015\)](#) in a decision-theoretic context, or as a *likelihood-separable* specialization of the stochastic choice costs introduced by [Morris and Yang \(2019\)](#) to study binary-action coordination games.<sup>3</sup> The preferences capture a notion of costly control that can be conditioned strategically on each state of the world.

Equilibrium is characterized by a functional fixed-point equation for the continuous distribution of actions. Our strategy for proving equilibrium properties is to find primitive conditions such that the operator describing this fixed-point equation is a contraction map. We first require conventional conditions such that the underlying game has supermodular but sufficiently concave payoffs, and that aggregation is monotone and concave in level shifts of the action distribution. We next identify and require a condition on stochastic choice

---

<sup>3</sup>[Denti, Marinacci, and Rustichini \(2019\)](#) formulate a class of cost functional that they similarly call likelihood-separable. These differ from ours as the likelihoods to which they refer are defined over a fixed signal space in a model with an explicit information-acquisition interpretation.

such that these conventional restrictions on payoffs and aggregation translate into the required properties of each agent’s optimal stochastic choice, which are appropriate notions of monotonicity and discounting.

We first use the contraction-mapping approach to establish when there exists a unique equilibrium (Theorem 1). We next use the same approach to show that, if payoffs are jointly supermodular in actions, aggregates and the state, then action distributions and aggregates are monotone in the state (Theorem 2).<sup>4</sup> Under a further condition that payoffs are quadratic in one’s own action and the loss from misoptimization is monotone in both the endogenous aggregate and exogenous state of the world, we establish that the size of agents’ mistakes is monotone in the state (Theorem 3). Finally, we provide a simple, cost-function-independent necessary condition for efficiency of the resulting unique equilibrium: that the average marginal benefit of increasing the aggregate is zero (Proposition 1).

To illustrate these results, we introduce a class of generalized beauty contest games that model imperfect aggregate coordination. The generalized beauty contest allows for exogenous and endogenous state-dependence in the cost of mistakes, non-linear responsiveness of optimal actions to the same states, and non-linear aggregation. These features are missing from the linear beauty contest (Morris and Shin, 2002) and incompatible with standard approaches for tractably studying imperfect optimization within games (e.g., assuming Gaussian fundamentals and signals). But they are important for describing incentives in macroeconomic and financial applications. In particular, we show how the generalized beauty contest intuitively captures incentives for coordination and precise optimization in the context of speculation in asset markets, price-setting in monetary economies, consumption choice in the Keynesian cross, production in Real Business Cycle models, and Bertrand competition within an industry.

We use the strategic mistakes framework to establish properties of equilibrium in each aforementioned application that are robust to the extent of the cognitive friction, the stochastic properties of the state of nature, or, within specific bounds, the exact structure of payoffs and aggregation. This translates our abstract results into economically interpretable conditions in each context that guarantee uniqueness; monotonicity of aggregates, distributions, and the propensity to make mistakes; and efficiency. These conditions, in particular, provide robust rationalizations for empirical patterns of cyclical volatility and shock responsiveness in the aforementioned domains, as we later detail.

**Strategic Mistakes *vs.* Mutual Information.** Although its foundations are in information theory, the mutual information model of Sims (2003) has become a standard in

---

<sup>4</sup>Monotonicity for the action distribution is in the sense of first-order stochastic dominance.

the literature for modelling imperfectly optimizing agents.<sup>5</sup> We thus explore how mutual-information costs and strategic mistakes compare as models of controlled mis-optimization in equilibrium. To make the comparison as direct as possible, we focus on a “strategic mistakes cousin” of the mutual information model which takes the control cost equal to the expected entropy of the mixed strategy played in each state. Extending a result in [Matějka and McKay \(2015\)](#), we provide necessary and sufficient conditions on payoffs for the existence of a prior such that *some equilibrium* of the mutual information model coincides with the unique equilibrium of the entropy-cost strategic mistakes model (Proposition 2).

Away from equivalence, we explore the difference numerically in a linear beauty contest. We show that the mutual information model features multiple equilibria and non-monotone comparative statics while the strategic mistakes model has a unique equilibrium and the comparative statics expected from the theory. We illustrate how the mutual information model’s properties arise from the specific way in which the model has agents “anchor” to often-played actions, which breaks the contractive properties of the equilibrium map.<sup>6</sup> Our results highlight that these non-uniqueness and non-monotonicity properties stem from cross-state interactions in the cost of information and not from stochastic choice *per se*.

**Additional Related Literature.** The literature has proposed many models to rationalize *state-dependent stochastic choice*, or random decisionmaking whose distribution may depend on an objective state of the world. These generally fall into the following partially overlapping categories: costly learning about fundamentals or “rational inattention” ([Sims, 2003](#); [Caplin and Dean, 2015](#); [Pomatto, Strack, and Tamuz, 2018](#); [Denti, 2019](#); [Hébert and Woodford, 2020](#)), imperfect perception of choice attributes ([Woodford, 2012](#)), imperfect reasoning ([Gabaix, 2014](#); [Ilut and Valchev, 2020](#)), random utility ([McFadden, 1973](#)), and trembling hands or control costs ([Harsanyi, 1973](#); [Selten, 1975](#); [Myerson, 1978](#)). Our approach to modelling misoptimization relates most closely to the last of these. Within this literature, of most relevance is the work of [Van Damme \(1991\)](#) and [Stahl \(1990\)](#) who add control costs to normal form games and study the resulting refinement of Nash equilibrium as the cost of control vanishes.<sup>7</sup> Relative to these papers, we are instead concerned with establishing uniqueness and comparative statics of equilibria when a broad class of control costs are

---

<sup>5</sup>See the review article by [Maćkowiak, Matejka, and Wiederholt \(2020\)](#) for specific applications, and in particular Section 2.4.3 about connections with both information processing and behavioral theories of imperfect optimization.

<sup>6</sup>This anchoring is studied by [Matějka and McKay \(2015\)](#) and shown by [Caplin, Dean, and Leahy \(2019\)](#), [Jung, Kim, Matějka, and Sims \(2019\)](#), [Matějka \(2015\)](#), and [Stevens \(2019\)](#) to result in “endogenous consideration sets” or sparse supports for action profiles. Our model can generate sparse supports for *conditional* action distributions, in a way that does not go through an “anchoring” channel and is therefore compatible with contraction-mapping.

<sup>7</sup>[Mattsson and Weibull \(2002\)](#) axiomatize entropy control costs in a model where agents incur disutility of effort in restraining mistakes.

non-vanishing.

This focus on the case with non-vanishing mistakes relates our analysis to the Quantal Response Equilibrium (QRE) of [McKelvey and Palfrey \(1995\)](#), who add type-I extreme value noise to agents' utility functions to smooth best responses via logit play.<sup>8</sup> Our model with an entropy kernel of stochastic choice generates the same choice and equilibrium patterns as QRE, but in our large game concept. Our model therefore differs in the crucial respect that we consider more general choice patterns by allowing for different kernel functions and consider a model in which agents interact via an aggregator of the cross-sectional action distribution.

Our paper thus also contributes to the literature on aggregative games (see [Jensen, 2018](#), for a review). Within this literature, the most related work is by [Acemoglu and Jensen \(2010, 2013, 2015\)](#), who provide equilibrium results and comparative statics in aggregative games with a continuum of players. Our analysis differs in that we study the equilibrium consequences of strategic misoptimization in such a setting.

Finally, our analysis relates to a large literature on uniqueness in games with strategic complementarity (e.g., [Morris and Shin, 1998, 2002](#); [Weinstein and Yildiz, 2007](#); [Yang, 2015](#)). Our proof strategy is most closely related to [Frankel, Morris, and Pauzner \(2003\)](#) and [Mathévet \(2010\)](#), in that we use contraction mapping techniques. Our results on comparative statics are similar in spirit to those of [Van Zandt and Vives \(2007\)](#), but differ in that we study different games and provide comparative statics for action distributions.

**Outline.** Section 2 introduces the model. Section 3 presents our main results on equilibrium properties. Section 4 discusses applications of our main results. Section 5 compares the strategic mistakes model to the mutual information model. Section 6 concludes.

## 2 Model

### 2.1 Primitives and Definitions

A continuum of identical agents is indexed by  $i \in [0, 1]$ . They take actions  $x_i \in \mathcal{X} = [\underline{x}, \bar{x}] \subset \mathbb{R}$ . Cross-sectional distributions of actions are aggregated by an aggregator  $X : \Delta(\mathcal{X}) \rightarrow \mathbb{R}$ . There is an underlying and payoff-relevant state of the world  $\theta \in \Theta \subset \mathbb{R}$ . The state space

---

<sup>8</sup>See a review by [Palfrey, Holt, and Goeree \(2016\)](#). Contemporaneously, [Allen and Rehbeck \(2021\)](#) generalize QRE with finite players by allowing for more general kernel and utility functions and establish equilibrium existence. By contrast, we study a model with a continuum of players and establish existence while additionally providing results on uniqueness, efficiency, and monotone comparative statics. Also contemporaneously, [Melo \(2020\)](#) studies when QRE is unique in finite games with general additive random utility. This differs from our analysis in both the games under consideration, the random utility approach adopted, and our further results on comparative statics and efficiency.

$\Theta$  is a finite set, over which the agent has prior  $\pi \in \Delta(\Theta)$ .<sup>9</sup> Agents have identical utility functions  $u : \mathcal{X} \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  where  $u(x, X, \theta)$  is an agent’s utility from playing  $x$  when the aggregate is  $X$  and the state is  $\theta$ .<sup>10</sup> We assume throughout that  $u$  and  $X$  are continuous and bounded.<sup>11</sup>

Each agent chooses a stochastic choice rule  $P : \Theta \rightarrow \Delta(\mathcal{X})$  with  $P(x|\theta)$  describing the cumulative distribution of actions  $x$  taken by the agent in state  $\theta$ . When this admits a density function, we denote a stochastic choice rule by  $p(x|\theta)$ . We call the set of measurable stochastic choice rules  $\mathcal{P}$ . We model the cost of “controlling mistakes” via a cost functional  $c : \mathcal{P} \rightarrow \bar{\mathbb{R}}$  which returns how costly any given stochastic choice rule is for the agent.

The agent’s problem, put together, is to maximize expected utility net of the control cost given a conjecture  $\tilde{X} : \Theta \rightarrow \mathbb{R}$  for the aggregate, as summarized in the following program:

$$\max_{P \in \mathcal{P}} \sum \int_{\mathcal{X}} u(x, \tilde{X}(\theta), \theta) dP(x|\theta) \pi(\theta) - c(P) \quad (1)$$

The strategic mistakes model further specializes the cost functional to belong to what we call the likelihood-separable class:<sup>12</sup>

**Definition 1** (Likelihood-Separable Cost Functional). *A cost functional  $c$  has a likelihood-separable representation if there exists a strictly convex function  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that for any stochastic choice rule  $P$  with corresponding density  $p$ :*

$$c(P) = \sum \int_{\mathcal{X}} \phi(p(x|\theta)) dx \pi(\theta) \quad (2)$$

*with the convention that the cost is  $\infty$  if  $P$  does not have a density.*

Likelihood-separable cost functionals capture the idea that it is costly for agents to avoid “mistakes” or misoptimizations in a specific, state-separable way. These mistakes are strategic because agents’ incentives to avoid misoptimization depend on the behavior (and the mistakes) of other agents.

We will refer to the collection of all primitives (payoffs, costs of stochastic choice, aggregator, prior, and domains) as a *game of strategic mistakes*  $\mathcal{G}$ . An equilibrium of a game of

---

<sup>9</sup>All arguments carry over to the case with a continuum of states so long as one ensures that agents choose the optimal stochastic choice rule in the interim, for each state. We omit this to simplify exposition.

<sup>10</sup>This assumes that all agents care about a common state. In Flynn and Sastry (2021), we show that it is possible to allow for idiosyncratic states in a specific special case of the model considered here.

<sup>11</sup>Throughout, our notion of continuity for functionals will be under the sup norm.

<sup>12</sup>Denti, Marinacci, and Rustichini (2019) formulate a class of cost functional that they similarly call likelihood-separable. These differ from ours as the likelihoods to which they refer are defined over a different space.

strategic mistakes is defined as the following:

**Definition 2** (Equilibrium). *An equilibrium  $\Omega$  of game  $\mathcal{G}$  is a tuple comprising a collection of stochastic choice rules  $\{P_i\}_{i \in [0,1]}$  and an equilibrium law of motion for aggregates  $\hat{X} : \Theta \rightarrow \mathbb{R}$  such that:*

1. *Each agent's stochastic choice rule  $P_i$  solves program (1) under the correct conjecture that  $\tilde{X}(\theta) = \hat{X}(\theta)$  for all  $\theta \in \Theta$ .*
2. *The equilibrium law of motion is consistent with the stochastic choice rules played by agents, or*

$$\hat{X} = X \circ \int_{[0,1]} P_i \, di \tag{3}$$

*An equilibrium is symmetric if  $P_i = P$  for all  $i \in [0, 1]$ .*

Our high-level approach in defining the game with stochastic choice is to recast the underlying game with incomplete information in the interim as an *ex ante* game with complete information and a strategy space sufficiently rich to embed all profiles of state-dependent mixed strategies. This approach has also recently been adopted and studied by [Morris and Yang \(2019\)](#) in the binary action context. Moreover, the same approach (with different  $c$ ) is applicable to games with costly learning with general posterior separable cost functionals ([Caplin, Dean, and Leahy, 2017](#); [Denti, 2019](#)), such as mutual-information costs ([Sims, 2003](#); [Matějka and McKay, 2015](#)), log-likelihood ratio costs ([Pomatto, Strack, and Tamuz, 2018](#)), or neighborhood-based costs ([Hébert and Woodford, 2020](#)).

The interpretation and properties of any stochastic choice game then hinge greatly on the specification of the cost functional  $c$ . In particular, it is known that games with different cost functionals have different implications for efficiency ([Angeletos and Sastry, 2019](#); [Hébert and La'O, 2020](#)) and uniqueness ([Morris and Yang, 2019](#)).

## 2.2 Interpreting Strategic Mistakes

We now draw out the specific interpretation of our proposed class of likelihood-separable cost functionals. Succinctly, we argue that they capture a notion of costly control within any state when there is no confusion as to the identity of the state. We split the discussion into two sub-parts: interpreting the costly control problem in a fixed state and then interpreting the separability of these problems across states.



### 2.2.1 Costly Control within a State

Our leading interpretation of the kernel  $\phi$  is that it represents a cost of control in the spirit of [Van Damme \(1991\)](#). We heuristically illustrate this (assuming for exposition that  $\phi$  is twice differentiable) with the following combined first-order condition for the probability for playing any two choices  $x, x^\theta \in \mathcal{X}$  in state  $\theta$ , conditional on those choices being interior:

$$u(x^\theta, \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) = \phi^\theta(p(x^\theta|\theta)) - \phi^\theta(p(x|\theta)) \quad (4)$$

Since  $\phi$  is convex, higher payoffs for action  $x^\theta$  imply a higher probability on that action. But the shape of  $\phi$  controls how much more often the agent plays  $x^\theta$  than  $x$ . When  $\phi^\theta$  is small, the agent is more “decisive” and translates small utility differences into large differences in probability of play; when  $\phi^\theta$  is large, the agent is “trembling” and plays the better action only a little more often. This distinction depends on the value of the utility function  $u$  and the (slope of the) control cost  $\phi$ , but not independently on the value of the action  $x$ . That is, the agent wavers between their best options in utility terms without additional constraints related to the “distance” between those actions.

Importantly, our model can accommodate zero probabilities of playing certain actions or *endogenous consideration sets* in the terminology of the literature on rational inattention with mutual-information costs ([Caplin, Dean, and Leahy, 2019](#)). This is the sense in which our cost functional differs from that axiomatized by [Fudenberg, Iijima, and Strzalecki \(2015\)](#) in decision problems without uncertainty or strategic interaction. In particular, these authors impose that all choice probabilities are strictly positive and consequently require that  $\phi$  satisfies an Inada condition, ruling out such consideration sets.

Our model also accommodates *certain* random utility models and thereby allows for an interpretation as a model in which there are cross-sectional idiosyncratic differences in preferences across agents. For example, our model with an entropy kernel  $\phi(p) = p \log p$  corresponds to a model where agents have logistic choice patterns. That is, our agents behave as if they have utility function:

$$\tilde{u}(x, X, \theta) = u(x, X, \theta) + \varepsilon_x \quad (5)$$

where  $\varepsilon_x$  is distributed type-I extreme value and IID across agents and actions. Thus, with an entropy kernel our model corresponds to a continuum version of Quantal Response Equilibrium ([McKelvey and Palfrey, 1995](#)). However, the model is neither nested by nor nests all random utility models, as shown by [Fudenberg, Iijima, and Strzalecki \(2015\)](#).<sup>13</sup>

---

<sup>13</sup>There is a tighter mapping between Additive Perturbed Utility models and a type of variational prefer-

We finally note that our model does not allow the cost of control to depend directly on what others do, as summarized by the probability distribution of the aggregate  $X$ . This precludes direct externalities of agents' actions on others' costs of cognition, as studied by [Angeletos and Sastry \(2019\)](#) and [Hébert and La'O \(2020\)](#) in the context of market economies and games, respectively.

### 2.2.2 Expected Control Cost Across States

The second ingredient of our model is the linearity of control costs across states. The key economic assumption is a notion of independence of decisions about both payoffs and control costs across states of nature. Concretely, this allows for the agent's choice about costly control in some state  $\theta_0 \in \Theta$  to be independent of their choice about costly control in another state  $\theta_1 \in \Theta$ . This is the assumption that will generally be violated in any alternative model with an interpretation of information acquisition, since actions are measurable in signals that, if not perfectly revealing, will conflate multiple states of nature.

This prior-independence assumption underscores that our model cannot be very precisely invoked when the economic environment necessarily involves learning and/or economically important disagreement about the state of the world. Our model is more relevant when the right notion of “mistakes” involves not knowing the optimal action *conditional* on the state of the world about which the agent is certain in the interim, like if agents plausibly observe goods or asset prices and still make sub-optimal purchase decisions.<sup>14</sup> In this context, independence is still a strong assumption but one with costs and benefits familiar from standard expected utility theory. In our case, it will enable tractable single-agent comparative statics and, in turn, tractable equilibrium analysis. This underscores that our results will capture the interaction between stochastic choice and strategic considerations within each state, but not conflation across states. Nevertheless, our results may still be of interest to those studying information acquisition in games, in which the sole additional feature to the analysis is interaction across states.

## 3 Main Results

We now prove existence, uniqueness, efficiency and equilibrium monotone comparative statics for both the aggregate and the cross-sectional action distribution. Our approach will be to

---

ence, whereby Nature chooses payoff-specific shocks (subject to costs) adversarially and the decision maker *ex ante* maximizes expected utility. See Section 5.1 of [Fudenberg, Iijima, and Strzalecki \(2015\)](#).

<sup>14</sup>Here, there is also a connection to the study of imperfect decision rules by [Ilut and Valchev \(2020\)](#), though their model is better understood as introducing noise in the best response functions.

establish that the correct notion of a “best response function” for the aggregate action  $X$  is a contraction map.

### 3.1 Assumptions: Payoffs and Aggregator

We first identify conditions on payoffs, aggregators and stochastic choice functionals sufficient to guarantee uniqueness. For payoffs, we first require complementarities in the underlying game in the form of supermodularity in payoffs between an agent’s own action and the aggregate. Second, we require that these complementarities are not too strong in the sense that payoffs are sufficiently concave to outweigh them.<sup>15</sup>

**Assumption 1** (Supermodularity and Sufficient Concavity). *Payoffs are supermodular in  $(x, X)$ , or for all  $x^\theta \geq x, X^\theta \geq X, \theta$ , the following holds:*

$$u(x^\theta, X^\theta, \theta) - u(x, X^\theta, \theta) \geq u(x^\theta, X, \theta) - u(x, X, \theta) \quad (6)$$

*Moreover, payoffs feature sufficient concavity to outweigh supermodularity, or for all  $\alpha \in \mathbb{R}_+, x^\theta \geq x, X$ , and  $\theta$ , the following holds:*<sup>16</sup>

$$u(x, X + \alpha, \theta) - u(x - \alpha, X, \theta) \geq u(x^\theta, X + \alpha, \theta) - u(x^\theta - \alpha, X, \theta) \quad (7)$$

More informally, the latter part of the assumption ensures that agents’ stochastic best-response functions giving the optimal  $x$  as a function of conjectured  $X$  and  $\theta$  are not unduly sensitive to the value of  $X$ . To gain intuition for the role of these assumptions, observe that they ensure, in the case where  $u$  is twice continuously differentiable, that the slope of the unconditionally optimal best reply to  $X$  is between zero and one.

Having identified conditions on payoffs, we now turn to finding conditions on the aggregator of the cross-sectional distribution of actions. Naturally, we require that the aggregator is monotone in the sense of first-order stochastic dominance. We further require that the aggregator satisfies discounting, which is to say that it is concave in level shifts of the cross-sectional action distribution.<sup>17</sup>

<sup>15</sup>The condition we identify also appears in the literature on uniqueness in Bayesian games. See [Weinstein and Yildiz \(2007\)](#) for an example.

<sup>16</sup>In stating this assumption, we are implicitly extending the domain of  $u$  so that it is well-defined under such translations.

<sup>17</sup>If we weaken this assumption to allow  $\beta = 1$ , our results will carry so long as the second inequality in Assumption 1 is made strict. This assumption nests the important class of linear aggregators:

$$X(g) = \int_{\mathcal{X}} f(x)g(x) dx \quad (8)$$

**Assumption 2** (Monotone and Discounted Aggregator). *For all  $g, g^\theta \in \Delta(\mathcal{X})$ :*

$$g^\theta \succeq_{FOSD} g \implies X(g^\theta) \geq X(g) \quad (9)$$

*Moreover, there exists  $\beta \in (0, 1)$  such that for any distribution  $g \in \Delta(\mathcal{X})$  and any  $\alpha \in \mathbb{R}_+$ :*

$$X(\{g(x - \alpha)\}_{x \in \mathcal{X}}) \leq X(\{g(x)\}_{x \in \mathcal{X}}) + \beta\alpha \quad (10)$$

### 3.2 Intermediate Result: Properties of Stochastic Choice

Assumption 2 suggests a path toward ensuring that equilibrium is described by a contraction map if, in response to level shifts in the aggregate, the optimal stochastic choice pattern increases in the sense of first-order stochastic dominance (monotonicity) but remains dominated by the level shift itself (discounting). These are intuitive properties given the supermodularity and concavity of payoffs, which encode that level shifts in the (conjectured) aggregate globally increase the attractiveness of playing higher  $x$ , but in a way that is less than one-for-one. We now show an interpretable sufficient condition within the likelihood-separable class which guarantees that monotonicity and discounting translate appropriately to stochastic choice.

We first define a new property of a function that we label quasi-monotone-likelihood-ratio-property (quasi-MLRP). This condition allows us to relate the underlying cost functional to the distribution of actions induced by optimality.

**Definition 3** (Quasi-MLRP). *A function  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  satisfies quasi-MLRP if for any two distributions  $g^\theta, g \in \Delta(\mathcal{X})$ :*

$$\left( f(g^\theta(x^\theta)) - f(g^\theta(x)) \geq f(g(x^\theta)) - f(g(x)) \quad \forall x^\theta \geq x \right) \implies g^\theta \succeq_{FOSD} g \quad (11)$$

It is moreover worthwhile to note that important functions satisfy quasi-MLRP and that the class of  $f$  satisfying the quasi-MLRP property  $\mathcal{F}$  is non-empty. Indeed, as the name suggests, quasi-MLRP is a strict weakening of MLRP. More specifically, it can be verified that both  $f(x) = x$  and  $f(x) = \log x$  satisfy quasi-MLRP.<sup>18</sup>

With this definition in hand, we can now state our final assumption on stochastic choice functionals, which will ensure we can always translate dominance in payoff units to domi-

---

when  $f : X \rightarrow \mathbb{R}$  is a differentiable function with derivative bounded strictly between zero and one.

<sup>18</sup>Lemma 2 in Online Appendix C establishes this formally. We moreover provide an additional Lemma 3, which establishes a general method by which one may be able to identify other members of this class. When we say that quasi-MLRP is a weakening of MLRP, we mean the following: when  $f = \log(\cdot)$ , the implication in Equation 3 holds.

nance in terms of distributions:

**Assumption 3** (Quasi-MLRP Kernel). *Costs have a differentiable kernel  $\phi$  such that  $\phi^\theta$  satisfies quasi-MLRP.*

Example kernels that satisfy this assumption are the entropy kernel  $\phi(x) = \lambda x \log x$  and the quadratic kernel  $\phi(x) = \lambda x^2$  for any scaling  $\lambda > 0$ . The latter does not feature an Inada condition and therefore allows for consideration sets, where some actions lie outside the support of the optimal stochastic choice rule.

We can now state a Lemma using this assumption and our earlier assumptions on payoffs to establish monotonicity and discounting of the solution of the stochastic choice problem:

**Lemma 1** (Monotone and Discounted Stochastic Choice). *Consider the stochastic choice program with payoffs satisfying Assumption 1 and cost functional satisfying Assumption 3. Then,*

1. *The optimal stochastic choice rule  $p$  is weakly increasing in the sense that if  $\hat{X}^\theta \geq \hat{X}$  then  $p(\theta; \hat{X}^\theta) \succeq_{FOSD} p(\theta; \hat{X})$  for all  $\theta \in \Theta$ .*
2. *The optimal choice rule is discounted in the sense that when  $\hat{X}$  and  $\hat{X}^\theta = \hat{X} + \alpha$  for  $\alpha \in \mathbb{R}_+$ , we have that  $p_\alpha(\theta; \hat{X}) \succeq_{FOSD} p(\theta; \hat{X}^\theta)$  for all  $\theta \in \Theta$ , where  $p_\alpha$  denotes the translation of  $p$  to the right by  $\alpha$ .*

*Proof.* See Appendix A.1. □

The key to both parts is that quasi-MLRP allows us to “invert” dominance relationships in payoffs to dominance relationships between distributions. For the first part, we show that the dominance of payoffs for playing higher  $x$  from supermodularity implies dominance of distributions under quasi-MLRP. For the second part, we use the property of payoffs from (7) that concavity of utility exceeds strategic complementarity, to show the optimal stochastic choice rule is dominated by the claimed level shift in the rule.

Lemma 1 is the core to our environment’s tractability. It is in principle the ingredient that might be replaced in an alternative model of stochastic choice like a form of unrestricted information acquisition. But, to our knowledge, such monotonicity and discounting results do not exist for any form of information acquisition in general environments. Moreover, this is not merely a technical glitch. A very relevant mechanism, anchoring toward frequently played actions, fights such monotonicity and discounting in information acquisition models. In a numerical example with the mutual-information cost (Sims, 2003) in Section 5, this leads to violations of monotonicity and discounting in the single-agent problem, and to non-uniqueness and non-monotone comparative statics in the equilibrium of an example game.

### 3.3 Existence and Uniqueness

We can now state our main existence and uniqueness result:

**Theorem 1** (Existence, Uniqueness, and Symmetry). *Under Assumptions 1, 2 and 3, there exists a unique equilibrium. The unique equilibrium is symmetric.*

*Proof.* See Appendix A.2. □

As alluded to above, we show this result by defining an equilibrium operator that maps the law of motion of the aggregate in the state to the resulting optimal stochastic choice rule and then maps this back to a law of motion of the aggregate, and then determining that said operator is a contraction map. More formally, let  $\mathcal{B} = \{\hat{X}|\hat{X} : \Theta \rightarrow \mathbb{R}\}$  be a space of (bounded) functions endowed with the sup norm. We define the operator  $T : \mathcal{B} \rightarrow \mathcal{B}$ :

$$T\hat{X} = X \circ p(\hat{X}) \tag{12}$$

To show uniqueness of the equilibrium law of motion of aggregates, it then suffices to prove that  $T$  is a contraction map. We prove this by showing that, under the given assumptions,  $T$  satisfies both of Blackwell's conditions of monotonicity and discounting. Given the unique equilibrium-consistent law of motion which satisfies  $T\hat{X} = \hat{X}$ , the equilibrium stochastic choice rule is then the unique solution of the stochastic choice problem given that law of motion, or  $p(\hat{X})$ . This extends classic uniqueness results to the realm of stochastic choice.<sup>19</sup>

### 3.4 Monotone Comparative Statics

Once we lie in the realm of unique equilibria, it is well-posed to consider comparative statics in equilibrium. We provide two such results, showing when the action distribution and aggregate action are monotone in the state and when the precision of agents actions is monotone in the state.

#### 3.4.1 Monotonicity of Action Distributions

To show monotonicity of distributions and aggregates, we require a stronger supermodularity assumption that not only are individual actions and aggregate actions complements, but so too is the underlying state itself a complement to both individual actions and aggregates in utility:

---

<sup>19</sup>One could dispense with Assumptions 1, 2, and 3 and prove existence in our setting only by applying the Schauder fixed-point theorem. We omit this result as it is simple, and because our analysis will proceed afterward under Assumptions 1, 2, and 3.

**Assumption 4.** *The payo function  $u$  is supermodular in  $x$  and  $(X, \theta)$ :*

$$u(x^\theta, X^\theta, \theta^\theta) - u(x, X^\theta, \theta^\theta) \geq u(x^\theta, X, \theta) - u(x, X, \theta) \quad \forall \theta^\theta \geq \theta, X^\theta \geq X, x^\theta \geq x \quad (13)$$

**Theorem 2** (Monotone Actions and Aggregates). *Under Assumptions 1, 2, 3, and 4, the unique equilibrium action distribution is monotone increasing in the sense of FOSD and the law of motion of the aggregate is increasing in the underlying state.*

*Proof.* See Appendix A.3. □

The intuition for this result is simple: higher  $\theta$  makes higher actions more desirable so that the distribution of actions in higher states dominates the distribution in lower states. The proof strategy makes use of the contraction mapping property used in the uniqueness proof. In particular, it shows that monotonicity is preserved by the fixed point operator and therefore that the fixed point must itself be monotone.

### 3.4.2 Monotonicity of Action Precision

We now turn to establishing when the precision of, or extent of mistakes in, agents' actions is monotone in the state in equilibrium. To this end, in our context with flexible stochastic choice, we first need a non-parametric notion of precision:

**Definition 4** (Precision). *Fix an  $h : \mathbb{R} \rightarrow \mathbb{R}$ . A symmetric distribution  $g$  is more precise about a point  $x$  than  $g^\theta$  about  $x^\theta$  under  $h$  if  $h \circ g(|x - x|)$  is faster decreasing in  $|x - x|$  than is  $h \circ g^\theta(|x^\theta - x^\theta|)$  in  $|x^\theta - x^\theta|$ .<sup>20</sup>*

Informally, this definition requires that a distribution is more precise than another if its density is more rapidly decreasing away from the point about which precision is being considered. This definition generalizes the property that Gaussian distributions are more precise about their mean when they have a lower standard deviation to cases with non-Gaussian densities by exactly capturing the idea that a distribution is more precise if its tails decay faster from the point about which a distribution is centered.<sup>21</sup>

<sup>20</sup>On an asymmetric support, we call a distribution  $g$  symmetric if  $g(x) = g(-x)$  whenever both  $g(x)$  and  $g(-x)$  are defined. For any symmetric functions  $\xi, \hat{\xi} : A \rightarrow \mathbb{R}$ , we say that  $\xi$  is faster decreasing than  $\hat{\xi}$  in their arguments if  $\xi(0) - \xi(|x|) \geq \hat{\xi}(0) - \hat{\xi}(|x|)$  for all  $x \in A$ .

<sup>21</sup>To see this, recall that a Gaussian random variable with mean  $\mu$  and standard deviation  $\sigma$  has pdf:

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\} \quad (14)$$

Thus, for two Gaussian distributions with means  $\mu, \mu'$  and standard deviations  $\sigma, \sigma'$  such that  $\sigma < \sigma'$ , we

Having defined precision, we now state sufficient assumptions on payoffs for precision to be monotone. To show this result, we specialize to a quadratic-in-own-action environment, which we refer to throughout as *generalized beauty contest* payoffs:

**Assumption 5** (Generalized Beauty Contests). *The utility function is given by:*

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \quad (15)$$

where  $\gamma(X, \theta)$  is monotonically increasing in  $(X, \theta)$  and  $\beta(X, \theta)$  is positive, for every  $(X, \theta)$ .

Such a utility function can be justified via a second-order approximation of any smooth, concave utility function. In such an interpretation,  $\gamma(X, \theta)$  is the optimal action conditional on  $(X, \theta)$  and  $\beta(X, \theta)$  measures the curvature of payoffs, or second-order loss of mis-optimization, around that point. Our results on monotone precision will rely on monotonicity of  $\beta(X, \theta)$  in both of its arguments. It may be useful to interpret such conditions via the following example of consumers' *precautionary savings*. Standard analysis of this issue (e.g., [Kimball, 1990](#)) emphasizes that the second derivative of payoffs  $|u_{xx}|$  is monotonically decreasing in an agent's own action when they have convex marginal utility (and increasing when concave). The former case induces more cautious behavior when consumption is low, while the opposite case induces more risky behavior in the same context. Our model is quadratic in  $x$  but can feature endogenous state-dependence in  $u_{xx}$ , which can capture a "macro" version of the same phenomenon: agents can have incentives to behave more (or less) cautiously when the exogenous or endogenous state in the game is lower.

We now state the result, which encapsulates the idea that precision is higher when the losses from mis-optimization are higher for endogenous or exogenous reasons:

**Theorem 3** (Monotone Precision). *Under Assumptions 1, 2, 3, 4, and 5,  $p(\theta) \in \Delta(\mathcal{X})$  is more precise about  $\gamma(\hat{X}(\theta), \theta)$  than  $p(\theta^0)$  about  $\gamma(\hat{X}(\theta^0), \theta^0)$  under  $\phi^\theta$*

1. *For any  $\theta^0 \leq \theta$  if  $\beta(X, \theta)$  is monotone decreasing in both arguments.*
2. *For any  $\theta^0 \geq \theta$  if  $\beta(X, \theta)$  is monotone increasing in both arguments.*

*Proof.* See Appendix [A.4](#). □

The proof of this result shows that the agents' incentives to transfer probability mass from the ideal point  $\gamma(\hat{X}(\theta), \theta)$  to any other  $x \in \mathcal{X}$  are strictly lower when  $\beta(\hat{X}(\theta), \theta)$  is

---

have that  $h \circ g(|x - \mu|)$  is faster decreasing than  $h \circ g'(|x - \mu'|)$  whenever  $h$  is monotone. Thus, under monotone  $h$ , we have that Gaussian distributions with lower standard deviations are more precise about their mean under our definition of precision.



larger, which translates directly to our notion of precision. This calculation relies on the symmetry of quadratic payoffs around  $\gamma(\hat{X}(\theta), \theta)$ .<sup>22</sup> It then leverages the fact that  $\hat{X}$  is monotone in  $\theta$  in equilibrium, because of Theorem 2, which in turn implies monotonicity of the mapping  $\theta \mapsto \beta(\hat{X}(\theta), \theta)$ , decreasing in case 1 and increasing in case 2. Put differently, “endogenous” and “exogenous” stakes amplify one another in equilibrium, so we can infer that precision is monotone in the state.<sup>23</sup>

### 3.5 Efficiency

A further question of interest is when equilibria of our model are efficient. As our agents are symmetric, *ex-ante* Pareto efficiency and utilitarian efficiency are equivalent. We therefore say that a stochastic choice rule is efficient if it maximizes utilitarian welfare:

**Definition 5.** *A stochastic choice rule  $P^E \in \mathcal{P}$  is efficient if it solves the following program:*

$$P^E \in \arg \max_{P \in \mathcal{P}} \sum \int_X u(x, X(p(\theta)), \theta) dP(x|\theta) \pi(\theta) - c(P) \quad (17)$$

An efficient stochastic choice rule both fully internalizes the effect choices have on aggregates and the costs of stochastic choice. Thus, this notion of efficiency takes seriously that agents do incur the cognitive cost of constraining their mistakes. We now ask, when will equilibrium be efficient? The following result relates the answer to this question to the *absence of aggregate externalities* in physical (non-cognitive) payoffs, which is most clearly illustrated in a specialization with a linear aggregator:<sup>24</sup>

**Proposition 1.** *Suppose that there exists a unique efficient  $P^E$  obtained as an interior solution of the efficient program and  $u$  is differentiable in its second argument  $X$ . Moreover,*

---

<sup>22</sup>Assumption 5 can in fact be relaxed to consider objective functions of the form:

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)\Gamma(|x - \gamma(X, \theta)|) \quad (16)$$

for monotone increasing  $\Gamma$  with the normalization  $\Gamma(0) = 0$ . The proof of Theorem 3 then carries exactly as written under this slightly more general specification. We specialize to the quadratic form for expositional purposes given the justification of quadratic  $\Gamma$  by an approximation argument.

<sup>23</sup>Unsurprisingly, we cannot state a general result when  $\beta(X, \theta)$  is not strictly and symmetrically monotone in its two arguments; but we could of course still use part of the previous argument to compare precision in any two states  $(\theta, \hat{X}(\theta)), (\theta', \hat{X}(\theta'))$  after solving for equilibrium.

<sup>24</sup>It is worth noting that the condition of a linear aggregator is weak as the utility function can transform this aggregator. For example, this formulation accommodates constant elasticity of substitution (or more generally quasi-arithmetic) aggregation. Note, however, this clearly does not mean that imply that only mean aggregation is without loss. Indeed, our assumptions on the payoff function must be understood as holding under this transformation.

suppose that the aggregator is linear:

$$X(g) = \int_X f(x) dG(x) \tag{18}$$

for some non-constant function  $f$ . A necessary condition for efficiency of an equilibrium stochastic choice rule  $P$  to be efficient is that:

$$\int_X u_X(x, X(p(\theta)), \theta) dP(x|\theta) = 0 \tag{19}$$

for all  $\theta \in \Theta$ .

*Proof.* See Appendix A.5. □

This result leverages the fact that externalities appear only through  $u(\cdot)$  and not  $c(\cdot)$ , since the latter depends on aggregates only through choices. Such a restriction to externalities only in physical payoffs is also among the sufficient conditions for efficiency in studies of competitive equilibria and games with unrestricted information acquisition, by [Angeletos and Sastry \(2019\)](#) and [Hébert and La'O \(2020\)](#), respectively. In particular, using those papers' language, our likelihood-separable cost functional is suitably *invariant* to the actions of others: the exact value of others' actions does not affect any individual's "technology" for controlling mistakes. This is not to say, however, that stochastic choice behavior cannot open up novel avenues for inefficiency, a point to which we will return in our analysis of generalized beauty contests.

## 4 Applications

We now show how our abstract results can shed light on equilibrium behavior in a class of *generalized beauty contest* games. These games accommodate three features which are abstracted away from in the canonical linear-beauty-contest framework of [Morris and Shin \(2002\)](#): state-dependent costs of misoptimization, non-linear best responses, and non-linear aggregation. We first show how these features naturally arise in applications to asset-market speculation and price-setting, and how our analysis can shed light on the mapping from primitives to equilibrium properties in each case. We then sketch additional applications to consumption in the Keynesian cross, production in Real Business Cycles models, and price-setting under Bertrand competition in a single industry.

## 4.1 Generalized Beauty Contests and Financial Speculation

### 4.1.1 Set-up

We specialize payoffs to have the following state-dependent quadratic form originally introduced in Assumption 5 and Equation 15:

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \quad (20)$$

The function  $\gamma(X, \theta)$  is the agent’s pay-off maximizing “best reply,” absent informational or choice frictions. We assume that it is increasing in both arguments, to capture respectively coordination motives and payoff monotonicity in the state. The function  $\beta(X, \theta)$  allows the penalty of failing to play  $\gamma(X, \theta)$  to depend on both the endogenous and exogenous state variable. The endogenous state variable  $X$  is determined by an aggregator that satisfies Assumption 2 (monotonicity and discounting).

In the language of financial speculation, the context studied informally by Keynes (1936) and formally by Morris and Shin (2002) as a coordination game, we can think of  $x$  as an individual investor’s position,  $X$  as the aggregate trading in the market, and  $\theta$  as fundamental value. The monotonicity of  $\gamma(X, \theta)$  in both arguments implies that higher overall activity and stronger fundamentals justify more individual investment. But the analyst need not impose that either this best-response function or the aggregate  $X$  is linear in others’ trades. The appropriate monotonicity of  $\beta(X, \theta)$  is ambiguous without more structure. Standard theories of risk-pricing would imply that dollar payoffs are more valuable in lean times, so the natural units for  $u$  are risk-adjusted payoffs and  $\beta$  is a decreasing function of one or both arguments. On the other hand, Keynes (1936) remarks in his exposition of the beauty-contest metaphor that “worldly wisdom” implies “it is better for reputation to fail conventionally than to succeed unconventionally.” By this interpretation, the natural units of  $u$  are reputation-adjusted profits and  $\beta$  would increase in both arguments to make the “conventional failure” of low profits during lean times less damaging.

As throughout, the important friction in our analysis is that agents make mistakes in their choice of asset positions. Our analysis is germane to cases in which working out the right position to take is a challenging activity that requires some amount of costly contemplation, even in the presence of awareness of the fundamental and/or other market statistics. This is formalized through our likelihood separable costs of stochastic choice, which we assume to have a kernel  $\lambda \hat{\phi}$  for some  $\hat{\phi}$  satisfying the Quasi-MLRP condition (Assumption 3) and  $\lambda > 0$ . Resulting from this friction, we define the bias and dispersion of a stochastic choice

rule  $P$  in state  $\theta$  around optimal point  $\gamma(X(P), \theta)$  as

$$\begin{aligned} \text{Bias}[P, \theta] &\equiv \int_{\mathcal{X}} (x - \gamma(X(P), \theta)) \, dP(x|\theta) \\ \text{Disp}[P, \theta] &\equiv \left( \int_{\mathcal{X}} (x - \gamma(X(P), \theta))^2 \, dP(x|\theta) \right)^{\frac{1}{2}} \end{aligned} \quad (21)$$

#### 4.1.2 Results and Interpretation

Applying our general results, we can provide conditions under which a generalized beauty contest has a unique equilibrium with a number of economically relevant properties:

**Corollary 1.** *The following properties hold under any extent of friction  $\lambda > 0$  and the additional stated conditions.*

1. **Uniqueness.** *There exists a unique equilibrium if the following holds for all  $x \in \mathcal{X}$ ,  $X \in \mathcal{X}$  and  $\theta \in \Theta$ :*

$$-(1 - \gamma_X(X, \theta)) < \frac{\beta_X(X, \theta)}{\beta(X, \theta)} (x - \gamma(X, \theta)) < \gamma_X(X, \theta) \quad (22)$$

2. **Monotone investment.** *The cross-sectional distribution of investment and aggregate investment are monotone in the fundamental if, in addition to the condition (22), the following holds for all  $X \in \mathcal{X}$ ,  $x \in \mathcal{X}$ , and  $\theta \in \Theta$ .<sup>25</sup>*

$$\frac{\beta_\theta(X, \theta)}{\beta(X, \theta)} (x - \gamma(X, \theta)) < \gamma_\theta(X, \theta) \quad (23)$$

3. **Monotone precision.** *The precision of investors' investment about the optimal level of investment  $\gamma$  under  $\phi^\theta$  is increasing (decreasing) in the strength of fundamentals if, in addition to (22) and (23),  $\beta$  is monotone decreasing (increasing) in both arguments.*
4. **Efficiency.** *A necessary condition for efficiency of the stochastic choice rule  $P$  when the efficient stochastic choice rule is unique and interior is that*

$$\begin{aligned} 0 = & \alpha_X(X(P(\theta)), \theta) - \beta_X(X(P(\theta)), \theta) (\text{Disp}[P(\theta), \theta])^2 \\ & + 2 \gamma_X(X(P(\theta)), \theta) \beta(X(P(\theta)), \theta) \text{Bias}[P(\theta), \theta] \end{aligned} \quad (24)$$

*Proof.* See Appendix A.6. □

---

<sup>25</sup>Where, for simplicity, we allow  $\beta$  to be defined for all states in a closed interval that contains  $\Theta$ , and assume it is differentiable in its second argument.

Under the given conditions, all of the above results hold *without* any specific reference to the extent of cognitive cost.<sup>26</sup> Our predictions are therefore robust across two dimensions: they allow flexibility for “physical” parameters within the bounds specified in Corollary 1, and arbitrary flexibility for the scaling of cognitive costs.

We now discuss each of the conditions and the intuition behind them in turn. First, by Theorem 1, equilibrium existence and uniqueness (Point 1) can be established by verifying that Assumptions 1 and 2 are satisfied and concavity condition (7) is strict. In condition 22, the right inequality corresponds to complementarity, and is loosened for high  $\gamma_X$ , while the left inequality corresponds to concavity, and is tightened for higher  $\gamma_X$ . Were  $\beta_X \equiv 0$ , this condition would reduce to the more familiar condition  $\gamma_X(X, \theta) \in (0, 1)$ , for all  $(X, \theta)$ , which bounds the slope of the agent’s fully-attentive best-response function. With unrestricted mis-optimization, however, the condition must be adjusted to accommodate strategic interactions via shifting the endogenous cost of mis-optimization. These additional interactions, via our two interpretations of  $\beta(X, \theta)$ , correspond with investors’ shifting aggregate risk-bearing capacity or the reputational standards by which the market evaluates them.

To verify that aggregate trading increases in fundamental value (Point 2), by Theorem 2 we require that payoffs are jointly supermodular in investment, aggregate investment and fundamentals. This is ensured by Equation 23. In parallel to previous discussion, this condition simplifies to a monotone best response  $\gamma_\theta > 0$  when there is no state-dependence in  $\beta$ . Otherwise it needs to be adjusted to account for the shifting costs of mis-optimization. In particular, sufficiently large shifts in action dispersion, driven by shifting incentives to optimize precisely, may undermine monotonicity of the action distribution should the condition fail.

To verify the state-dependence of investing mistakes (Point 3), we inspect the conditions underlying Theorem 3. Under our risk-pricing interpretation above ( $\beta$  is monotone increasing), our result verifies that investors are more precise in lean times; and under our interpretation of Keynes’s (1936) reputational story ( $\beta$  is monotone decreasing), our result verifies that investors are more careless in lean times and careful during booms. Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) provide empirical evidence that actively managed US equity mutual funds’ portfolios are more dispersed from one another during recessions.<sup>27</sup> Interpreting  $x$  as a portfolio weight on a “aggregate-fundamentals-tracking” investment (like

---

<sup>26</sup>The inclusion of vanishingly small costs  $\lambda \downarrow 0$  in the previous result contrasts with known results in the literature regarding “cognitive traps,” or equilibrium multiplicity that arises *only* when a cognitive friction is added to the game (see, e.g., Tirole, 2015; Hellwig and Veldkamp, 2009; Myatt and Wallace, 2012). Our analysis further underscores how such results rely on features of information-acquisition environments in specific games, and do not generalize to our class of coordination games with imperfect optimization.

<sup>27</sup>See Table 3 (Columns 1 and 2) in Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016).

the stock market), our model explains this finding as the outcome of the reputational mechanism of Keynes (1936) dominating the risk pricing mechanism.

Efficiency of this asset market (Point 4) requires the following three economic conditions under the premise that dispersion and strategic interaction  $\gamma_X$  are necessarily non-zero: (i) zero pure externalities, (ii) zero dependence of objective-function curvature on aggregate actions, and (iii) zero aggregate bias in actions. The first channel is standard. The second channel uncovers a possibly important interaction between the “average extent of mistakes,” summarized by the dispersion of actions, and the endogenous state-dependence of the *cost* of mistakes, summarized by the slope of  $\beta$  in  $X$ . The final channel is a first-order effect of moving the the optimal action  $\gamma$  whenever optimal stochastic choice does not equal  $\gamma$  on average. Such asymmetry is natural if, for instance, agents face a non-negativity (no-short-selling) constraint for asset positions.

## 4.2 Price-Setting in a Monetary Economy

We next adopt the generalized beauty contest model to study price-setting in a monetary economy, another macroeconomic scenario commonly modeled as a coordination game. Our studied game, unlike the linear beauty contest frameworks of Woodford (2003) and Maćkowiak and Wiederholt (2009), accommodates non-linear best responses and state-dependent costs of mis-optimization, and correspondingly delivers richer predictions for the entire distribution of prices in the economy.

### 4.2.1 Set-up

Let  $x$  denote the price set by a given firm (to allow consistent notation with our main model);  $q$  denote the quantity sold by a given firm;  $X$  denote the aggregate price level; and  $\theta$  denote an aggregate nominal demand shock (e.g., to the money supply). We assume that each firm faces a demand curve of the form

$$q = q_0(X, \theta) - q_1(X, \theta)x \tag{25}$$

where we assume  $q_0, q_1 > 0$ .<sup>28</sup> This allows the aggregate conditions to affect both the intercept and slope of the demand curve faced by firms. The firm produces at constant marginal cost  $c_0(X, \theta) > 0$ , where dependence of marginal costs on others’ prices and the fundamental represents the dependence of wages on these forces in an integrated labor market. Firms’

---

<sup>28</sup>Observe that this is a linear approximation of the implied demand system given the aggregator.

profits are then given by:

$$\Pi(x, X, \theta) = (x - c_0(X, \theta))(q_0(X, \theta) - q_1(X, \theta)x) \quad (26)$$

By completing the square around the maximum point, we can re-write the objective (26) in the generalized beauty-contest form (20) with

$$\begin{aligned} \alpha(X, \theta) &= \Pi(\gamma(X, \theta), X, \theta) \\ \beta(X, \theta) &= q_1(X, \theta) \\ \gamma(X, \theta) &= \frac{1}{2} \left( \frac{q_0(X, \theta)}{q_1(X, \theta)} + c_0(X, \theta) \right) \end{aligned} \quad (27)$$

Each firm chooses its stochastic choice rule over prices to maximize expected profits net of likelihood-separable control costs, assumed to have a kernel  $\phi$  that satisfies Assumption 3. Concretely this embodies the assumption that firms know the strength of demand but, as motivated in the Introduction, find it costly to work out the optimal price they should set conditional on this knowledge.

The model is closed by specifying the aggregate price level as a function of the distribution of good-specific prices, via any aggregator that satisfies Assumption 2.

#### 4.2.2 Results and Interpretation

Since this economy is a generalized beauty contest, its theoretical properties follow exactly from Corollary 1 with appropriate changes in the language. We now go through the economic intuition for these conditions as applied to the price-setting context.

When is there a unique equilibrium in this monetary economy? It is natural to assume that the optimal price  $\gamma(X, \theta)$  is monotone in both others' price  $X$  and the monetary shock  $\theta$ , which is implied by the monotonicity of  $q_0/q_1$  and  $c_0$ . The former implies that higher prices for other products and higher nominal demand increase the *intercept* of demand relative to its slope. The second implies that input prices go up under the same circumstances. Assuming the condition of Equation 22, which ensures that this monotonicity is not too strong and that the slope of the demand curve does not render the game one of strategic substitutes, then guarantees existence and uniqueness of this price-setting game with strategic mistakes by Theorem 1.

When are the aggregate price level and all increasing functionals of the price distribution (e.g., quantiles) monotone in the monetary shock in this economy? These results are obtained as long as optimal reset prices are increasing in the monetary shock  $\theta$  and the associated increase in the strength of demand, and the curvature of firms' payoff functions is not so

high as to render lower prices more attractive when demand is higher.

Do firms make greater mistakes in price-setting when nominal demand is high or low? Under our specification, the monotonicity of  $\beta(X, \theta)$  is left unclear. In particular, it may be instructive to consider the demand curve parameterization

$$q_0(X, \theta) = q_1(X, \theta) \cdot \hat{q}(X, \theta) \tag{28}$$

such that  $\hat{q}(X, \theta)$  controls the monotonicity of  $\gamma(X, \theta)$  and  $q_1(X, \theta)$  can be manipulated independently. See that, under this parameterization, monotone decreasing  $\beta(X, \theta)$  corresponds to monotone decreasing  $q_1(X, \theta)$ , or higher  $X$  and  $\theta$  “flattening” the mapping from prices to quantities while leaving its intercept with the  $q = 0$  axis constant; monotone increasing  $q_1(X, \theta)$  similarly involves “steepening” the mapping from prices to quantities. Intuitively, the “flatter” mapping from prices to quantities always lowers the effect of a fixed price-setting mistake on revenues. Thus, we predict that if the inverse demand curve is endogenously and exogenously flatter (steeper) when nominal aggregate demand is low, that firms will pay less (more) attention in equilibrium under those conditions.

Empirical evidence from [Munro \(2021\)](#) and [Stroebel and Vavra \(2019\)](#) suggests that the elasticity of demand faced by US firms increases during recessions, or in the previous paragraph’s language, that the inverse demand curve gets steeper in recessions. Through the lens of our model, this implies that price-setters would be more responsive to nominal shocks in recessions, a prediction empirically verified by [Berger and Vavra \(2019\)](#).

### 4.3 Additional Applications

**Consumption and the Keynesian Cross.** Several recent studies have modeled the Keynesian cross with imperfect optimization to study equilibrium properties including the effectiveness of forward guidance about monetary policy.<sup>29</sup> These models, up to approximation, can be described as variants of dynamic linear beauty contests. Here, we describe a simpler static beauty-contest model that retains the original insights of [Keynes \(1936\)](#) and [Hicks \(1937\)](#) and fits into our generalized beauty contest framework. The agents are households who choose their consumption  $x$ , aggregate consumption is  $X$  and  $\theta$  is a primitive shifter of demand capturing expectations of future income or preferences to consume today. Households are aware of their income, but nevertheless find it hard to plan and select the optimal level of consumption and face likelihood-separable costs of stochastic choice. The ideal point  $\gamma(X, \theta)$  corresponds to the optimal level of consumption today, which is increasing in both

---

<sup>29</sup>See, for instance, [Garcia-Schmidt and Woodford \(2019\)](#), [Gabaix \(2020\)](#), [Farhi and Werning \(2019\)](#) and [Angeletos and Lian \(2018\)](#).



aggregate consumption, which raises household incomes, and  $\theta$ , in its role as capturing a desire to consume instead of save. The cost of misoptimization  $\beta(X, \theta)$  arises naturally from over/under-saving and is the curvature of household utility in consumption. Monotone decreasing  $\beta(X, \theta)$  then corresponds to and is well-motivated by a kind of precautionary motive, where households are more averse to both over-saving today (because marginal utility of consumption is high) and under-saving today (because marginal utility of consumption will also be high tomorrow). Our results in this context describe when there is a unique equilibrium in which the distribution of consumption is increasing in the sense of first-order stochastic dominance in the strength of demand, and whether mistakes in consumption are smaller when demand is low. [Macaulay \(2020\)](#) shows that UK households more reliably choose high-interest-rate savings technologies in downturns than in booms, and interprets this as evidence of “counter-cyclical attention to saving.” These results are qualitatively consistent with the precautionary-savings case of our model.

**Production in Real Business Cycle Models.** Our results can be used to understand production over the business cycle as studied by [Angeletos and La’O \(2010\)](#), [Benhabib, Wang, and Wen \(2015\)](#), and others. We sketch a stylized version of the full model of [Flynn and Sastry \(2021\)](#) and defer mathematical details to that paper. Agents are firms producing differentiated goods in amount  $x$ , the aggregate  $X$  is total GDP, and the fundamental  $\theta$  shifts common, aggregate productivity. Firms understand the level of aggregate productivity, but as motivated in the introduction, find it challenging to work out the correct level to produce and therefore make mistakes as if they had a likelihood-separable cost of stochastic choice. The ideal point  $\gamma(X, \theta)$  satisfies  $\gamma_\theta(X, \theta) > 0$  and  $\gamma_X(X, \theta) > 0$ , encoding that higher productivity increases the frictionlessly optimal production level and demand externalities, on net, make firms’ production choices strategic complements.<sup>30</sup> Finally,  $\beta(X, \theta) = \beta(X)$  and is everywhere decreasing. This reflects the fact that the firms are owned by risk-averse households consuming  $X$ . Under our established conditions, there is a unique equilibrium level of aggregate GDP that is monotone increasing in aggregate productivity and firms make smaller mistakes in the cross-section during recessions. [Flynn and Sastry \(2021\)](#) provide direct empirical evidence of smaller mistakes during recessions, and hence a validation of this model mechanism, in US public firms’ input choices.

**Bertrand Competition.** The previous subsection’s price-setting model can also be used to study a continuous analogue of Bertrand competition in a single industry. Such a model is used by [Angeletos and Pavan \(2007\)](#) to stylize the finite-player models of [Vives \(1984\)](#) and [Raith \(1996\)](#). The definition of the fundamental is changed to an industry-specific, instead of

---

<sup>30</sup>Formally, the second requires primitive conditions such that the aggregate demand externality is stronger than the force toward strategic substitutability, which comes from input prices.

macroeconomic, demand shifter, while the rest of the interpretation of both assumptions and results remains the same. Our results deliver conditions for uniqueness and monotonicity of price distributions and mistakes.

## 5 Strategic Mistakes *vs.* Mutual Information

We have introduced a new model of cognitive frictions in games and demonstrated both its tractability and broad applicability. We now compare the strategic mistakes model with the mutual-information cost rational inattention model of Sims (2003) which, despite its micro-foundation in optimal encoding of information, is often used in applications as a more general model of imperfect optimization.<sup>31</sup> We highlight important similarities and differences between the two models both in theory and practice.

### 5.1 Definitions and an Equivalence Result

We first provide abstract conditions under which a version of the strategic mistakes model makes identical equilibrium predictions to the mutual information model, to build intuition about the comparability and differences of the two approaches.

All information acquisition models that have a posterior separable representation, including mutual information, can be recast as a choice over stochastic choice rules in  $\mathcal{P}$  subject to some convex cost functional  $c$  (Denti, 2019). The mutual information cost of a stochastic choice rule  $P \in \mathcal{P}$  can be decomposed into two terms which we label below:<sup>32</sup>

$$c^{MI}(P) = \underbrace{\sum \int_X p(x|\theta) \log p(x|\theta) dx \pi(\theta)}_{\text{Likelihood-Separable Term}} - \underbrace{\int_X p(x) \log p(x) dx}_{\text{Cross-State Interactions}} \quad (29)$$

The first term is in fact identical to the likelihood-separable representation (2) with the (quasi-MLRP) kernel  $\phi(p) = p \log p$ . We label the resulting cost function  $c^{LSM}$ , or *logit strategic mistakes*. In a stochastic choice interpretation, this term encodes the agent’s desire to *increase* the entropy of the conditional action distributions or play randomly. The second term equals the entropy of the unconditional action distribution and encodes the agents’ desire to, on average, anchor toward commonly played actions. This force is absent in the logit strategic mistakes model, and therefore characterizes  $c^{MI}$  model compared to its

<sup>31</sup>See the review article of Maćkowiak, Matejka, and Wiederholt (2020) for a list of applications of the mutual information model, and the review of Woodford (2020) for discussion of how the mutual information model compares to other approaches for introducing imprecision into choice.

<sup>32</sup>In this expression, we use the definition of the marginal distribution  $p(x) = \sum_{\theta} p(x|\theta)\pi(\theta)$ .

“strategic mistakes cousin”  $c^{LSM}$ . Moreover, this decomposition makes clear that there is no conceptual difference in modelling any stochastic choice game with mutual information versus entropic stochastic choice other than that agents have different cost functions, and therefore preferences.

Matějka and McKay (2015) show that the second term (“anchoring”) has marginally zero influence on actions when agents’ actions are *ex ante* exchangeable, or agents play each action  $x$  with equal unconditional probability. From the analyst’s perspective, the key free parameter for engineering such exchangeability is the prior  $\pi(\cdot)$ . We extend this result to show, constructively, that an analyst free to specify the prior can re-construct the equilibrium of a logit strategic mistakes model as an equilibrium of an equivalent game with a mutual information friction provided that a technical condition on payoffs, which ensures that all actions can be made *ex ante* equally attractive, holds:

**Proposition 2** (Equilibrium Equivalence). *Suppose that the action space  $\mathcal{X}$  is finite. Let  $\Omega = (P, \hat{X})$  be a symmetric equilibrium for the game  $\mathcal{G}^{LSM} = (u(\cdot), \lambda c^{LSM}(\cdot), X(\cdot), \pi^\theta(\cdot), \Theta, \mathcal{X})$ . There exists some  $\pi^\theta(\cdot) \in \Delta(\Theta)$  such that  $\Omega$  is an equilibrium of  $\mathcal{G}^{LSM}$  and  $\mathcal{G}^{MI} = (u(\cdot), \lambda c^{MI}(\cdot), X(\cdot), \pi^\theta(\cdot), \Theta, \mathcal{X})$  if and only if the following linear system has a solution for  $\pi^\theta \in \Delta(\Theta)$ :*

$$\tilde{U}\pi^\theta = \frac{1}{|\mathcal{X}|} \quad (30)$$

where  $1$  is a  $|\Theta|$  length vector, and  $\tilde{U}$  is a  $|\mathcal{X}| \times |\Theta|$  matrix with entries:

$$\tilde{u}_{x_i, \theta_j} = \frac{\exp\{u(x_i, \hat{X}(\theta_j), \theta_j)/\lambda\}}{\sum_{x_k \in \mathcal{X}} \exp\{u(x_k, \hat{X}(\theta_j), \theta_j)/\lambda\}} \quad (31)$$

*Proof.* See Appendix A.7. □

The proof establishes from first-order conditions that (30) corresponds with a flat unconditional distribution over actions. Technically, the condition ensures that there exists a prior such that all actions yield ex-ante equal payoffs. More heuristically, it is likely to fail if the state space does not have many realizations or if some actions in  $\mathcal{X}$  are unappealing regardless of the state.

This result has two practical implications. First, an analyst who is unsure of the physical prior distribution can think of the logit strategic mistakes model as a selection criterion for the mutual information model, across *games indexed by different priors* and, within each prior, a *potentially non-singleton set of equilibria*. This is a general-equilibrium analogue of Matějka and McKay’s (2015) insight about the relationship between logit and mutual-information models for individual choice: the former approximates the latter when the analyst does

not take a specific stand on anchoring toward defaults. Second, comparative statics in the strategic mistakes model which perturb payoffs  $u(\cdot)$  or compare across states  $\theta \in \Theta$  may be interpreted, under the conditions of Proposition 2, as comparative statics in a mutual information model *jointly* across the aforementioned features and the physical prior and given a specific equilibrium selection rule.

## 5.2 Numerically Revisiting The Beauty Contest

We now return to the beauty contest model to illustrate the differences between the strategic mistakes and mutual information models in a practical scenario that maps to the applications of Section 4. Because closed-form solutions are not available for equilibrium action profiles under the mutual information cost, we instead make a feasible approximation of the model on a gridded action space.<sup>33</sup> We will show in this context sharp differences between the predictions of the logit strategic mistakes and mutual information models regarding equilibrium multiplicity and comparative statics, and that these stem from the cross-state interactions embedded in the mutual information cost functional.<sup>34</sup>

### 5.2.1 Environment and Solution Method

For the simplest exposition and comparison to existing work, we use a version of our model that reduces to the linear beauty contest. In payoffs (20), we set  $\alpha(X, \theta) \equiv 0$ , eliminating the pure externality;  $\beta(X, \theta) \equiv 1$ , giving constant costs of misoptimization; and  $\gamma(X, \theta) = (1 - r)\theta + rX$  with  $r = 0.85$ .<sup>35</sup> The aggregator is the mean. The state space has two points of support,  $\Theta = \{\theta_0, \theta_1\} = \{1.0, 2.0\}$ . The action space  $\mathcal{X}$  is approximated with a 40-point grid between lower endpoint  $\underline{x} = 0$  and upper endpoint  $\bar{x} = 3$ . We use a flat prior with  $\pi(\theta_0) = \pi(\theta_1) = \frac{1}{2}$ . And we scale both logit and mutual information costs by  $\lambda = 0.25$ .

Let  $p(\hat{X}) \in \Delta(\mathcal{X})^2$  return each agent’s (unique) optimal stochastic choice rule, expressed as pair of probability mass functions, when they conjecture the equilibrium law of motion

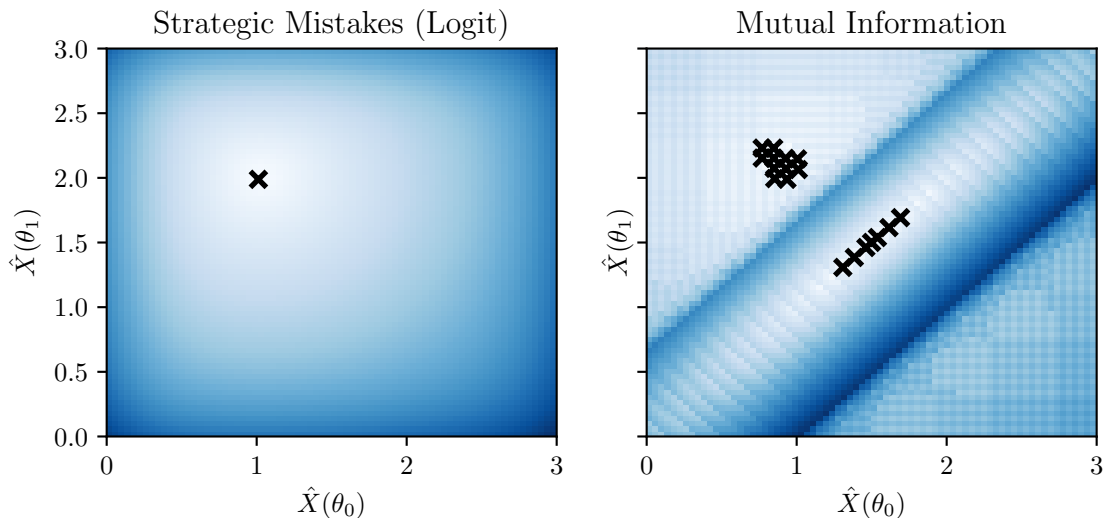
---

<sup>33</sup>This is due to two reasons, in our application with quadratic preferences: the lack of a Gaussian prior and the bounded action space. Moreover, if we had numerically solved a generalized beauty contest with state-dependent costs of mis-optimization, the non-quadratic payoffs would preclude a closed-form mutual-information solution even with a Gaussian prior and unbounded state space.

<sup>34</sup>Note that using logit strategic mistakes will imply that all actions are played with positive probability. To obtain endogenous consideration sets in the strategic mistakes model, we could have instead used a quadratic kernel.

<sup>35</sup>Hellwig and Veldkamp (2009) remark that, for dynamic beauty contests meant to mimic price-setting in New Keynesian models, that  $r = 0.85$  is “commonly used.” Finally, observe that these payoffs are jointly supermodular in  $(x, X, \theta)$  but feature bounded complementarity based on the conditions established in the previous section, provided that  $r \in (0, 1)$ .

**Figure 1:** Equilibria in the Beauty Contest



*Note:* Each plot is a 2-D histogram of  $\|T\hat{X} - \hat{X}\|$ , where  $\|\cdot\|$  indicates the Euclidean norm. Whiter colors indicate smaller values, and hence “closeness to equilibrium.” The cross marks represent equilibria, defined such that  $\|T\hat{X} - \hat{X}\| < 10^{-6}$ .

$\hat{X} = (\hat{X}(\theta_0), \hat{X}(\theta_1))$ .<sup>36</sup> As in the proof of our main results, let us define the operator  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  which constructs essentially the “best response” of aggregates to aggregates by composing the best response with the aggregator:

$$T\hat{X} = \left( X \circ p(\theta_0; \hat{X}), X \circ p(\theta_1; \hat{X}) \right) \quad (32)$$

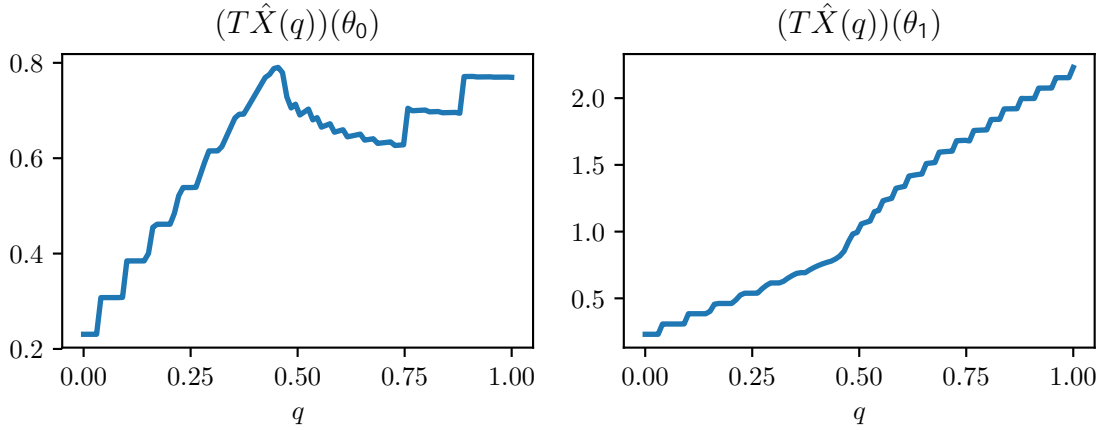
We define equilibria by first searching over a grid covering  $[\underline{x}, \bar{x}]^2$  for approximate fixed points  $\hat{X}$ , or low  $\|T\hat{X} - \hat{X}\|$ , and then using a numerical fixed-point solving algorithm with fine tolerance to confirm equilibria.

### 5.2.2 Equilibrium Uniqueness and the Contraction Map

Figure 1 plots the accuracy of the equilibrium conjecture,  $\|T\hat{X} - \hat{X}\|$ , in a heat map or two-dimensional histogram over the grid of candidate conjectures. Whiter areas denote that the equilibrium conjecture is closer to the aggregate best response, bluer areas indicate the opposite, and crosses identify equilibria. The strategic mistakes model, on the left, features a

<sup>36</sup>For the logit strategic mistakes model, the optimal action profile is known in closed form. For the mutual information model, we apply the Blahut-Arimoto algorithm as described in [Caplin, Dean, and Leahy \(2019\)](#) which iterates over the first-order condition for optimal stochastic choice and updates the marginal distribution over actions until convergence.

**Figure 2:** Partial Equilibrium Comparative Statics With Mutual Information



*Note:* These plots show aggregate best response  $T\hat{X}$  in state  $\theta_0$  (left pane) and  $\theta_1$  (right pane) along the path (33) for the equilibrium conjecture.

single-peaked surface and a single equilibrium. This is consistent with our theoretical results, and with the fixed-point condition (32) being a contraction. The mutual information model, on the right, features a non-monotone surface and 18 confirmed equilibria.

We now deconstruct further the failure of the contraction map argument for the mutual information model. Recall, in our proof of Theorem 1, that establishing monotonicity and discounting for the equilibrium operator  $T$  required first showing monotone and smooth comparative statics for the single-agent decision problem. To “test” this in the mutual information model, we parameterize a path that increases the equilibrium conjecture of  $\hat{X}$  from  $(0,0)$  to one of its equilibrium values.<sup>37</sup> Formally, if we label this chosen equilibrium as  $X_{MI} = (X_{MI}(\theta_0), X_{MI}(\theta_1))$ , we consider points indexed by  $q \in [0, 1]$ :

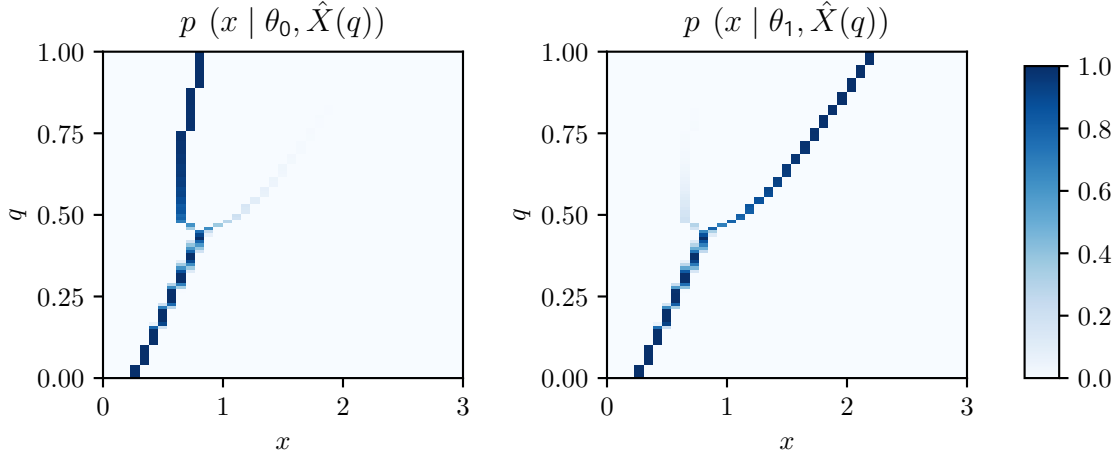
$$\hat{X}(q) = (q \cdot X_{MI}(\theta_0), q \cdot X_{MI}(\theta_1)) \quad (33)$$

and the aggregate best response  $T\hat{X}(q)$ . Figure 2 shows each element of  $T\hat{X}(q)$  as a function of  $q$ . The first element, plotted in the left panel, is (i) non-monotone and (ii) discontinuous in the equilibrium conjecture. In the language of the price-setting application, the mutual information model does not predict that expecting a higher price level increases one’s own price, even though the payoff to setting a higher price has globally increased; and when prices increase, they may jump suddenly.

To better understand the agent’s behavior along this path, we show in Figure 3 a two-

<sup>37</sup>We pick the equilibrium with the largest value of  $\hat{X}(\theta_1) - \hat{X}(\theta_0)$ .

**Figure 3:** Stochastic Choice Strategies With Mutual Information

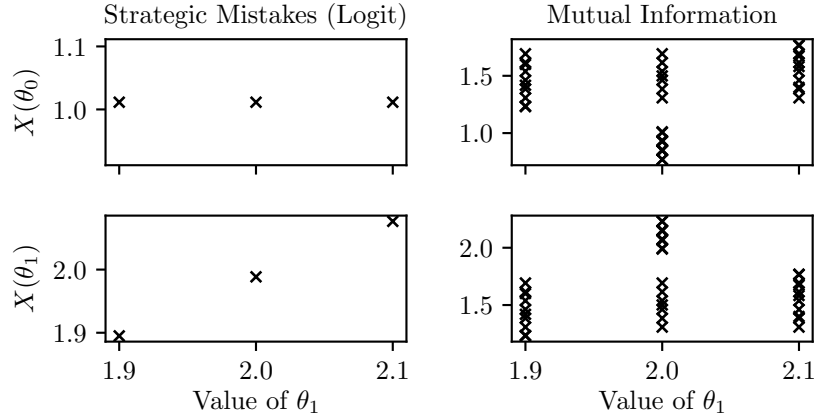


*Note:* Each slice on the the vertical axis ( $q$ ) gives the probability distribution of actions in state 0 (left) or 1 (right), represented via a “heat map” (scale on the right). The path of the equilibrium conjecture corresponds to the same in Figure 2.

dimensional histogram of the stochastic choice patterns conditional on each conjecture indexed by  $q$ . Equilibrium strategies are mostly supported on either one or two points. This sparsity of support is formally described by Jung, Kim, Matějka, and Sims (2019) and Caplin, Dean, and Leahy (2019) in discrete- and continuous-action variants of the mutual information model as a natural consequence of the lowered marginal costs (or, more loosely, “increasing returns to scale”) of allocating probability mass to frequently played actions. Sparse behavior is a characteristic feature of the optimal policy in price-setting applications studied by Matějka (2015) and Stevens (2019). In our example, the optimal policy switches between one and two support points around  $q = 0.45$ . Matějka (2015) refers to such behavior as a *bifurcation* in the optimal policy. As  $q$  increases after the bifurcation point, the optimal policy in Figure 3 pushes the larger and smaller support points away from one another. This violates monotonicity in the sense of first-order stochastic dominance, and therefore can lead to a non-monotone aggregate with respect to some admissible aggregators. Under our chosen aggregator, this behavior causes  $X(\theta_0)$  to decrease, as evident in the left panel of Figure 2.

Our observation is that this force can support multiple equilibria in coordination games because it breaks the contractive properties of the equilibrium map. These multiple equilibria are not, in our reading, very easily interpretable given that choices have an ordinal interpretation, payoffs leverage this interpretation in their definition of complementarity, and agents have a continuum of possible options. This reasoning is quite stark in the price-setting application which Matějka (2015) and Stevens (2019) study with mutual information. While

**Figure 4:** Equilibrium Comparative Statics in the Beauty Contest



*Note:* Each cross mark is an equilibrium, under the strategic mistakes (left) and mutual information (right) models, for different values of  $\theta_1$ . Note the different axis scales in each figure.

it is quite reasonable that a single firm wavers between charging \$1.99 and \$2.99 for its product, and indeed [Stevens \(2019\)](#) provides direct evidence for such behavior, it is a much stronger prediction that an entire (symmetric) economy of firms switches between a coordinated equilibrium of charging (\$1.99, \$2.99), respectively in each of two states of nature, to a different equilibrium of charging (\$1.98, \$3.00).

### 5.2.3 Equilibrium Comparative Statics

A point emphasized in our theory, and in particular the transition from [Theorem 1](#) (existence and uniqueness) to [Theorem 2](#) and [Theorem 3](#) (monotone aggregates and precision), was that the contraction map structure goes hand-in-hand with proving equilibrium comparative statics. We now illustrate the contrast between comparative statics with strategic mistakes and information acquisition in our model. We vary the value of the higher state  $\theta_1$  on the grid  $\{1.90, 2.00, 2.10\}$  and re-solve for all equilibria of each model. Our main results for the strategic mistakes model suggest that  $X(\theta_1)$  should monotonically increase in that model while  $X(\theta_0)$  stays constant, owing to the separability of decisions by state. For the mutual information model, there are no equivalent theoretical results.

[Figure 4](#) plots the equilibria of each model as a function of the chosen  $\theta_1$ . In the strategic mistakes model, we verify the predicted comparative statics across unique equilibria. In the mutual information model, we observe non-monotone comparative statics as equilibria move in and out of the set. Thus, while a mutual-information model may be an appealing



laboratory to study specific behaviors like discrete pricing, it may not lend itself to straightforward comparative statics analysis conditional on this feature outside of specific numerical calibrations.<sup>38</sup>

## 6 Conclusion

This paper introduces a model of strategic mistakes in large games. We provide results on equilibrium existence, uniqueness, efficiency, and monotonicity of equilibrium distributions, aggregates and mistakes. We showed how these results can be applied to make equilibrium predictions in generalized beauty contests that are robust to the extent of the stochastic choice friction. We used these games to model the impact of strategic mistakes on financial speculation, price-setting, the Keynesian cross, business cycles, and Bertrand competition.

In Online Appendix B, we study strategic mistakes in binary-action games, which are used in many applications to capture an extensive margin of adjustment and/or to simplify analysis.<sup>39</sup> We derive sufficient conditions on cognitive costs and payoffs to ensure unique and monotone equilibria and illustrate our results in the context of a simple investment game with linear payoffs (as in Yang, 2015). Unlike the continuous-action games studied in our main analysis, binary-action supermodular games may have multiple equilibria with small stochastic choice frictions. This result hinges on agents' ability to waver between options that have similar payoffs, but are far apart in the action space and induce very different equilibrium externalities. This result offers the following insight for researchers interested in well-posed comparative statics and not multiplicity *per se*: a “more complex” continuous-action model, by smoothing out aggregate best-response functions, may admit simpler analysis than a comparable binary-action model.

In Online Appendix C, we further provide sufficient conditions for a function to satisfy quasi-MLRP. These results may help researchers identify further cost functionals that are consistent with our main analysis and results. Further characterizing this class of functions is an interesting avenue for future work.

---

<sup>38</sup>Of course whether this is a “bug” or instead a “feature,” reflecting the unstable coordinational nature of activities like price-setting, is an open question that merits additional research. Stevens (2019), for instance, uses a model of coarse pricing with mutual-information costs to match micro-level evidence on pricing strategies and macroeconomic dynamics for aggregates. The micro-economic calibration builds the case that non-uniqueness and ambiguous comparative statics may indeed be features of the “correct” descriptive model of this setting.

<sup>39</sup>See Angeletos and Lian (2016) (in particular, Section 5) for a review of this literature.

# Appendices

## A Omitted Proofs

### A.1 Proof of Lemma 1

*Proof.* This follows immediately from the step proving the monotonicity and discounting conditions in Theorem 1. Note that this invokes only Assumptions 1 and 3.  $\square$

### A.2 Proof of Theorem 1

*Proof.* We first study the problem of a single agent  $i$  who is best replying to the conjecture that the law of motion of the aggregate is  $\hat{X} : \Theta \rightarrow \mathbb{R}$ . See that this agent faces the problem:

$$\mathcal{P}(\hat{X}) = \arg \max_{P \in \mathcal{P}} \sum \int_{\mathcal{X}} u(x, \hat{X}(\theta), \theta) dP(x|\theta) \pi(\theta) - c(P) \quad (34)$$

First, let us examine the set of measurable stochastic choice rules:

$$\mathcal{P} = \{p : \Theta \rightarrow \Delta(\mathcal{X})\} = \prod_{\theta \in \Theta} \Delta(\mathcal{X}) \quad (35)$$

See that  $\Delta(\mathcal{X})$  is compact as  $\mathcal{X}$  is compact. It therefore follows by finiteness of  $\Theta$  that  $\mathcal{P}$  is compact.

Define  $k : \mathcal{P} \times \mathcal{B} \rightarrow \bar{\mathbb{R}}$ , where  $\mathcal{B} = \{\hat{X} : \Theta \rightarrow \mathbb{R}\}$  as:

$$k(P, \hat{X}) = \sum \int_{\mathcal{X}} u(x, \hat{X}(\theta), \theta) dP(x|\theta) \pi(\theta) - c(P) \quad (36)$$

As  $\phi$  is strictly convex and  $u$  is bounded, it is without loss of optimality to restrict to optimizing over the set of measurable stochastic choice rules with density bounded above by some  $M \in \mathbb{R}$ ,  $\mathcal{P}_M$ . This is a closed set, which is a subset of a compact set  $\mathcal{P}$ , and therefore also compact. Thus, see that  $k$  is upper-semicontinuous in  $P$ , continuous in  $(P, \hat{X})$  owing to continuity of  $u$  and continuity of  $c$  over  $\mathcal{P}_M$  for any  $M$ . Thus, by Weierstrass' theorem, there exists a maximum. Moreover, by strict convexity of  $k(P, \hat{X})$ , it is unique. It immediately follows that in any equilibrium  $P_i = P = \mathcal{P}(\hat{X})$  for all  $i$  and thus that there cannot exist asymmetric equilibria.

To show existence of an equilibrium it now suffices to show that there exists a  $\hat{X}$  such

that:

$$\hat{X} = X \circ \mathcal{P} (\hat{X}) \quad (37)$$

To this end define the operator  $T : \mathcal{B} \rightarrow \mathcal{B}$  such that:

$$T(\hat{X}) = X \circ \mathcal{P} (\hat{X}) \quad (38)$$

We wish to show that  $T$  has a fixed point. We will moreover prove that this fixed point is unique as under the stated assumptions we can prove that  $T$  is a contraction map. To this, we wish to apply Blackwell's sufficient conditions for an operator to be a contraction. More specifically, if  $T$  operates on the space of bounded functions and is endowed with the sup norm, then the following are sufficient for  $T$  to be a contraction:

1. Monotonicity:  $\hat{X}^\theta \geq \hat{X} \implies T(\hat{X}^\theta) \geq T(\hat{X})$  for any  $\hat{X}^\theta, \hat{X} \in \mathcal{B}$
2. Discounting: there exists  $\beta \in (0, 1)$  such that  $T(\hat{X} + \alpha) \leq T(\hat{X}) + \beta\alpha$  for all  $\alpha \in \mathbb{R}_+$  and any  $\hat{X} \in \mathcal{B}$

Toward proving these properties, we first derive some necessary conditions for optimal stochastic choice. To this end, see that the stochastic choice program under an equilibrium conjecture  $\hat{X}$  is given by:

$$\max_{p \in \mathcal{P}} \sum \int_{\mathcal{X}} u(x, \hat{X}(\theta), \theta) dP(x|\theta) \pi(\theta) - \sum \int_{\mathcal{X}} \phi(p(x|\theta)) dx \pi(\theta) \quad (39)$$

Take the optimal policy  $p$  and now consider a perturbation for actions  $x, x^\theta \in \mathcal{X}$  in state  $\theta \in \Theta$  such that  $p(x|\theta; \hat{X}), p(x^\theta|\theta; \hat{X}) > 0$  by  $\varepsilon > 0$ :

$$\begin{aligned} \tilde{p}(x^\theta|\theta; \hat{X}) &= p(x^\theta|\theta; \hat{X}) + \varepsilon \\ \tilde{p}(x|\theta; \hat{X}) &= p(x|\theta; \hat{X}) - \varepsilon \end{aligned} \quad (40)$$

Taking the FOC with respect to  $\varepsilon$  and evaluating at  $\varepsilon = 0$  yields the following necessary optimality condition:

$$u(x^\theta, \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) = \phi^\theta(p(x^\theta|\theta; \hat{X})) - \phi^\theta(p(x|\theta; \hat{X})) \quad (41)$$

which must hold for almost all pairs  $x, x^\theta \in \mathcal{X}$ .<sup>40</sup> By the previous necessary conditions and the supermodularity assumption (Assumption 1) we have that (for all  $x^\theta \geq x$  in the support

---

<sup>40</sup>A more rigorous justification is given by the following. Suppose for any  $x' > x$ , we choose  $\delta > 0$  so that

of both stochastic choice rules and all  $\theta$ ):

$$\begin{aligned}
\phi^\theta(p(x^\theta|\theta; \hat{X}^\theta)) - \phi^\theta(p(x|\theta; \hat{X}^\theta)) &= u(x^\theta, \hat{X}^\theta(\theta), \theta) - u(x, \hat{X}^\theta(\theta), \theta) \\
&\geq u(x^\theta, \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) \\
&= \phi^\theta(p(x^\theta|\theta; \hat{X})) - \phi^\theta(p(x|\theta; \hat{X}))
\end{aligned} \tag{46}$$

We now need to check the case where the stochastic choice rules do not have full support. Moreover, if  $p(x^\theta|\theta) > 0$  and  $p(x|\theta) = 0$  then:

$$u(x^\theta, \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) \leq \phi^\theta(p(x^\theta|\theta; \hat{X})) - \phi^\theta(p(x|\theta; \hat{X})) = \phi^\theta(p(x^\theta|\theta; \hat{X})) - \phi^\theta(0) \tag{47}$$

Now take any  $x, x^\theta$  in the support of  $p(\theta; \hat{X})$  such that  $x^\theta \geq x$ . By the complementary slackness conditions, and supermodularity assumption, one of two cases holds: either both  $x$  and  $x^\theta$  are in the support of  $p(\theta; \hat{X}^\theta)$ ; or  $x$  is not in the support of  $p(\theta; \hat{X}^\theta)$  and  $x^\theta$  is. Under the first case, the given argument goes through exactly. In the second case, simply supplement Equation 46 with Equation 47.

In either of the above cases, if  $\phi^\theta$  satisfies quasi-MLRP (Assumption 3), then we have that  $p(\theta; \hat{X}^\theta) \succeq_{FOSD} p(\theta; \hat{X})$  for all  $\theta$ . It then follows by the monotonicity property of the aggregator (Assumption 2) that  $X(p(\theta; \hat{X}^\theta)) \geq X(p(\theta; \hat{X}))$  for all  $\theta$  and therefore that  $T(\hat{X}^\theta) \geq T(\hat{X})$ , which establishes the required monotonicity property.

We now prove discounting. To this end, we will show that when we take  $\hat{X}^\theta = \hat{X} + \alpha$  for  $\alpha \in \mathbb{R}_+$  that the resulting stochastic choice is dominated by an  $\alpha$  right translation of  $x' - \delta > x$ . Now consider a perturbation for  $\varepsilon > 0$  such that:

$$\begin{aligned}
\tilde{p}(\tilde{x}|\theta; \hat{X}) &= p(\tilde{x}|\theta; \hat{X}) + \varepsilon, \quad \tilde{x} \in [x', x' - \delta] \\
\tilde{p}(\tilde{x}|\theta; \hat{X}) &= p(\tilde{x}|\theta; \hat{X}) - \varepsilon, \quad \tilde{x} \in [x, x - \delta]
\end{aligned} \tag{42}$$

for  $p$  that has full support on  $[x', x' - \delta], [x, x - \delta]$ , we have that  $\tilde{p} \in \mathcal{P}$ . Taking the derivative of the value of  $\tilde{p}$  in  $\varepsilon$  and evaluating at  $\varepsilon = 0$ , we obtain the necessary condition:

$$\int_{x' - \delta}^{x'} \left[ u(\tilde{x}, \hat{X}(\theta), \theta) - \phi'(p(\tilde{x}|\theta; \hat{X})) \right] d\tilde{x} = \int_{x - \delta}^x \left[ u(\tilde{x}, \hat{X}(\theta), \theta) - \phi'(p(\tilde{x}|\theta; \hat{X})) \right] d\tilde{x} \tag{43}$$

Normalizing both sides by  $\delta > 0$ , we obtain:

$$\frac{\int_{x' - \delta}^{x'} \left[ u(\tilde{x}, \hat{X}(\theta), \theta) - \phi'(p(\tilde{x}|\theta; \hat{X})) \right] d\tilde{x}}{\delta} = \frac{\int_{x - \delta}^x \left[ u(\tilde{x}, \hat{X}(\theta), \theta) - \phi'(p(\tilde{x}|\theta; \hat{X})) \right] d\tilde{x}}{\delta} \tag{44}$$

Taking the limit of both sides as  $\delta \rightarrow 0$ , by Leibniz rule we obtain:

$$u(x', \hat{X}(\theta), \theta) - \phi'(p(x'|\theta; \hat{X})) = u(x, \hat{X}(\theta), \theta) - \phi'(p(x|\theta; \hat{X})) \tag{45}$$

Thus yielding the claimed necessary condition.

the original stochastic choice under  $\hat{X}$ . Under this transformation, observe by the necessary condition for optimality and the sufficient concavity condition on utility (Assumption 1) that:

$$\begin{aligned}\phi^\theta(p_\alpha(x^\theta|\theta; \hat{X})) - \phi^\theta(p_\alpha(x|\theta, \hat{X})) &= u(x^\theta - \alpha, \hat{X}(\theta), \theta) - u(x - \alpha, \hat{X}(\theta), \theta) \\ &\geq u(x^\theta, \hat{X}(\theta) + \alpha, \theta) - u(x, \hat{X}(\theta) + \alpha, \theta) \quad (48) \\ &= \phi^\theta(p(x^\theta|\theta; \hat{X} + \alpha)) - \phi^\theta(p(x|\theta; \hat{X} + \alpha))\end{aligned}$$

Moreover, by quasi-MLRP of  $\phi^\theta$  (Assumption 3), we have that  $p_\alpha(\theta, \hat{X}) \succeq_{FOSD} p(\theta, \hat{X} + \alpha)$  where  $p_\alpha$  is the described right translation by  $\alpha$  of  $p$ . Moreover, by the discounting property of the aggregator (Assumption 2), we then have that:

$$T(\hat{X} + \alpha) \leq X \circ p_\alpha(\hat{X}) \leq T(\hat{X}) + \beta\alpha \quad (49)$$

which establishes the discounting property. We have now shown that  $T$  satisfies Blackwell's sufficient conditions and is a contraction map. By the Banach fixed point theorem, there then exists a unique equilibrium  $\Omega$ .  $\square$

### A.3 Proof of Theorem 2

*Proof.* To show that the unique equilibrium aggregate law of motion of monotone in  $\theta$ , we use Corollary 1 from Chapter 3 of [Stokey, Lucas, and Prescott \(1989\)](#).

Define the set of monotone increasing and bounded functions  $\mathcal{M} = \{\hat{X} \in \mathcal{B} | \hat{X}(\theta^\theta) \geq \hat{X}(\theta) \quad \forall \theta, \theta^\theta \in \Theta : \theta^\theta \geq \theta\}$ . See that this set is closed. If we can show that  $T(\hat{X}) \in \mathcal{M}$  for any  $\hat{X} \in \mathcal{M}$ , then we know that the unique fixed point of  $T$  is in  $\mathcal{M}$  and therefore that the unique equilibrium law of motion is in  $\mathcal{M}$  according to Corollary 1 of [Stokey, Lucas, and Prescott \(1989\)](#). To this end, we wish to show that:

$$\hat{X}(\theta^\theta) \geq \hat{X}(\theta) \quad \forall \theta, \theta^\theta \in \Theta : \theta^\theta \geq \theta \implies T(\hat{X})(\theta^\theta) \geq T(\hat{X})(\theta) \quad \forall \theta, \theta^\theta \in \Theta : \theta^\theta \geq \theta \quad (50)$$

This follows immediately from the necessary condition used in the proof of Theorem 1. More precisely, we use the necessary optimality condition and Assumption 4 by taking  $\hat{X}(\theta^\theta) \geq \hat{X}(\theta)$ :

$$\begin{aligned}\phi^\theta(p(x^\theta|\theta^\theta, \hat{X})) - \phi^\theta(p(x|\theta^\theta, \hat{X})) &= u(x^\theta, \hat{X}(\theta^\theta), \theta^\theta) - u(x, \hat{X}(\theta^\theta), \theta^\theta) \\ &\geq u(x^\theta, \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) \quad (51) \\ &= \phi^\theta(p(x^\theta|\theta, \hat{X})) - \phi^\theta(p(x|\theta, \hat{X}))\end{aligned}$$

By the quasi-MLRP property of  $\phi^\theta$  (Assumption 3) we then have that  $p(\theta^\theta; \hat{X}) \succeq_{FOSD} p(\theta; \hat{X})$  and thus by the monotonicity of the aggregator (Assumption 2) that  $T(\hat{X})(\theta^\theta) \geq T(\hat{X})(\theta)$ .  $\square$

## A.4 Proof of Theorem 3

*Proof.* Recall also by Theorem 1, that the unique symmetric stochastic choice rule consistent with the unique equilibrium  $\hat{X}$  solves the following program:

$$p \in \arg \max_{p \in \mathcal{P}} \sum \int_X u(x, \hat{X}(\theta), \theta) dP(x|\theta) \pi(\theta) - \sum \int_X \phi(p(x|\theta)) dx \pi(\theta) \quad (52)$$

where we will suppress the dependence of the optimal policy on  $\hat{X}$  as it is unique. We first derive a useful necessary first-order condition for our calculation of precision. Consider a perturbation of the equilibrium optimal policy  $p$  for a given  $\theta$  and  $x$  such that  $p(x|\theta) > 0$ :

$$\begin{aligned} \tilde{p}(\gamma(\hat{X}(\theta), \theta)|\theta) &= p(\gamma(\hat{X}(\theta), \theta)|\theta) + \varepsilon \\ \tilde{p}(x|\theta) &= p(x|\theta) - \varepsilon \end{aligned} \quad (53)$$

See that if  $p \in \mathcal{P}$ , then  $\tilde{p} \in \mathcal{P}$ . Taking the first-order condition with respect to  $\varepsilon$  and evaluating at  $\varepsilon = 0$  reveals that a necessary condition for optimality of  $p$  is that:

$$u(\gamma(\hat{X}(\theta), \theta), \hat{X}(\theta), \theta) - u(x, \hat{X}(\theta), \theta) = \phi^\theta(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi^\theta(p(x|\theta)) \quad (54)$$

Under Assumption 5, we moreover have that

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \quad (55)$$

Thus our necessary condition simplifies to:

$$\beta(\hat{X}(\theta), \theta)(x - \gamma(\hat{X}(\theta), \theta))^2 = \phi^\theta(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi^\theta(p(x|\theta)) \quad (56)$$

Now consider any  $\theta, \theta^\theta$  such that  $\beta(\theta^\theta, \hat{X}(\theta^\theta)) \geq \beta(\theta, \hat{X}(\theta))$ . Note that, by Theorem 2, the aggregate  $\hat{X}$  is monotone increasing in the state  $\theta$ . Thus if  $\beta(\theta, X)$  is decreasing in both arguments, the stated case corresponds to  $\theta^\theta \leq \theta$ . If instead  $\beta(\theta, X)$  is increasing in both arguments, the stated case corresponds to  $\theta^\theta \geq \theta$ . Therefore, to verify the desired result, we now prove that the action distribution in state  $\theta^\theta$  is more precise about  $\gamma(\theta^\theta, \hat{X}(\theta^\theta))$  than the action distribution in state  $\theta$  is about  $\gamma(\theta, \hat{X}(\theta))$ , with respect to  $\phi^\theta$ .

To that end, we take  $x, x^\theta$  such that:

$$|x - \gamma(\hat{X}(\theta), \theta)| = |x^\theta - \gamma(\hat{X}(\theta^\theta), \theta^\theta)| \quad (57)$$

It follows that:

$$\begin{aligned} \phi^\theta(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi^\theta(p(x|\theta)) &= \beta(\hat{X}(\theta), \theta)(x - \gamma(\hat{X}(\theta), \theta))^2 \\ &\geq \beta(\hat{X}(\theta^\theta), \theta^\theta)(x^\theta - \gamma(\hat{X}(\theta^\theta), \theta^\theta))^2 \\ &= \phi^\theta(p(\gamma(\hat{X}(\theta^\theta), \theta^\theta)|\theta^\theta)) - \phi^\theta(p(x^\theta|\theta)) \end{aligned} \quad (58)$$

We now take care of those points that have no density. To this end consider the first-order condition for  $p(x|\theta)$ :

$$u(x, \hat{X}(\theta), \theta) - \phi^\theta(p(x|\theta)) - \lambda(\theta) - \kappa(x, \theta) = 0 \quad (59)$$

where  $\lambda(\theta)$  is the Lagrange multiplier on the constraint that  $\int_{\mathcal{X}} p(x|\theta) = 1$  and  $\kappa(x|\theta)$  is the Lagrange multiplier on the constraint that  $p(x|\theta) \geq 0$ . When  $p(x|\theta) = 0$ , we have that  $\kappa(x, \theta) \leq 0$ . Given our assumption on utility, this is given by:

$$\kappa(x, \theta) = -\beta(\hat{X}(\theta), \theta)(x - \gamma(\hat{X}(\theta), \theta))^2 + \alpha(\theta) - \lambda(\theta) \quad (60)$$

which is monotonically decreasing in  $|x - \gamma(\hat{X}(\theta), \theta)|$ . Thus, if there is an  $x$  such that  $p(x|\theta) = 0$ , then there exists an  $\bar{x}(\theta)$  such that  $p(x|\theta) = 0$  if and only if  $|x - \gamma(\hat{X}(\theta), \theta)| \geq |\bar{x}(\theta) - \gamma(\hat{X}(\theta), \theta)|$ . Moreover, by monotonicity of  $\beta(\hat{X}(\theta), \theta)$  in  $\theta$ , we have that  $p(x|\theta) = 0$  implies that  $p(x^\theta|\theta^\theta) = 0$ . Hence, we have always that:

$$\phi^\theta(p(\gamma(\hat{X}(\theta), \theta)|\theta)) - \phi^\theta(p(x|\theta)) \geq \phi^\theta(p(\gamma(\hat{X}(\theta^\theta), \theta^\theta)|\theta^\theta)) - \phi^\theta(p(x^\theta|\theta)) \quad (61)$$

for all  $x \in \mathcal{X}$ . It follows then by the definition of precision that  $p(\theta)$  is more precise about  $\gamma(\hat{X}(\theta), \theta)$  than  $p(\theta^\theta)$  about  $\gamma(\hat{X}(\theta^\theta), \theta^\theta)$  under  $\phi^\theta$ .  $\square$

## A.5 Proof of Proposition 1

*Proof.* Recall that the planner's problem is given by:

$$P^E \in \arg \max_{P \in \mathcal{P}} \sum \int_{\mathcal{X}} u(x, X(P(\theta)), \theta) dP(x|\theta) \pi(\theta) - c(P) \quad (62)$$

and that the aggregator is a linear function of the density function:

$$X(g) = \int_{\mathcal{X}} f(x) dG(x) = \int_{\mathcal{X}} f(x)g(x) dx \quad (63)$$

for some non-constant function  $f$ .

If the efficient allocation is obtained as an interior solution, then we have the the following first-order condition describing the optimality of allocating probability density between any two points  $x$  and  $x^\theta$ , written in some abuse of notation:

$$\begin{aligned} \frac{\partial c(P)}{\partial p^E(x^\theta|\theta)} - \frac{\partial c(P)}{\partial p^E(x|\theta)} = & u(x^\theta, X(p^E(\theta)), \theta) - u(x, X(p^E(\theta)), \theta) \\ & + \int_{\mathcal{X}} [f(x^\theta) - f(x)] u_X(\tilde{x}, X(p^E(\theta)), \theta) p^E(\tilde{x}|\theta) d\tilde{x} \end{aligned} \quad (64)$$

Let us now consider the equilibrium stochastic choice rule  $P$ . In any equilibrium, there are two possibilities in each state  $\theta$ . Either two or more actions are played in that state, or one action is played in that state. Suppose that two or more actions are played with positive density in state  $\theta$  and denote either of these by  $x, x^\theta$ . We have that  $p(x|\theta), p(x^\theta|\theta) > 0$ . Now consider the perturbation of the equilibrium stochastic choice  $P$  to  $\tilde{P}$  which shifts density from point  $x$  to  $x^\theta$ . Formally, in terms of density functions, the perturbation satisfies

$$\begin{aligned} \tilde{p}(x^\theta|\theta) &= p(x^\theta|\theta) + \varepsilon \\ \tilde{p}(x|\theta) &= p(x|\theta) - \varepsilon \end{aligned} \quad (65)$$

Note that if  $P \in \mathcal{P}$ , then we also have that  $\tilde{P} \in \mathcal{P}$ . Taking the first-order condition with respect to  $\varepsilon$  and evaluating at  $\varepsilon = 0$ , we obtain a necessary condition for equilibrium:

$$\frac{\partial c(P)}{\partial p(x^\theta|\theta)} - \frac{\partial c(P)}{\partial p(x|\theta)} = u(x^\theta, X(P(\theta)), \theta) - u(x, X(P(\theta)), \theta) \quad (66)$$

By the fact that there exists a unique efficient allocation  $P^E$ , we require that  $P = P^E$  for efficiency. Thus, we require that:

$$\frac{\partial c(P)}{\partial p^E(x^\theta|\theta)} - \frac{\partial c(P)}{\partial p^E(x|\theta)} = \frac{\partial c(P)}{\partial p(x^\theta|\theta)} - \frac{\partial c(P)}{\partial p(x|\theta)} \quad (67)$$

This equality implies that:

$$\int_{\mathcal{X}} [f(x^\theta) - f(x)] u_X(\tilde{x}, X(p^E(\theta)), \theta) p^E(\tilde{x}|\theta) d\tilde{x} = 0 \quad (68)$$



This, combined with the first-order conditions (64) and (66), yields the condition

$$[f(x^\theta) - f(x)] \int_{\mathcal{X}} u_X(\tilde{x}, X(P^E(\theta)), \theta) p^E(\tilde{x}|\theta) d\tilde{x} = 0 \quad (69)$$

As  $f$  is non-constant, we have that  $f(x^\theta) \neq f(x)$  for some  $x, x^\theta \in \mathcal{X}$ . Thus, we require for efficiency that:

$$\int_{\mathcal{X}} u_X(\tilde{x}, X(P^E(\theta)), \theta) p^E(\tilde{x}|\theta) d\tilde{x} = \int_{\mathcal{X}} u_X(\tilde{x}, X(P(\theta)), \theta) p(\tilde{x}|\theta) d\tilde{x} = 0 \quad (70)$$

where this is required for all  $\theta \in \Theta$ . If one action is played in equilibrium, then as  $P^E$  is unique and interior, we know that  $P \neq P^E$ . In this case, the integral condition above is vacuous and therefore remains necessary.  $\square$

## A.6 Proof of Corollary 1

*Proof.* We have directly assumed that Assumptions 2, 3 and 5 hold. The first claim follows so long as condition 22 implies Assumption 1. Observe that it suffices for us to show that  $|u_{xx}(X, \theta)| > u_{xX}(x, X, \theta) > 0$  for all  $x, X, \theta$ . This is exactly the claimed condition in the text given our payoff function from Assumption 5. The second claim follows so long as condition 23 implies Assumption 4. To see this, as we have already that  $u_{xX}(x, X, \theta) > 0$  for all  $x, X, \theta$ , it is sufficient to check that  $u_{x\theta}(x, X, \theta) > 0$  for all  $x, X, \theta$ . This is exactly the claimed condition in the text given our payoff function from Assumption 5.

The third claim follows by Proposition 1. Recall from Proposition 1 that a necessary condition for efficiency of an equilibrium  $P$  is that:

$$\int_{\mathcal{X}} u_X(\tilde{x}, X(P(\theta)), \theta) dP(\tilde{x}|\theta) = 0 \quad (71)$$

for all  $\theta \in \Theta$ . We now have that utility is given by:

$$u(x, X, \theta) = \alpha(X, \theta) - \beta(X, \theta)(x - \gamma(X, \theta))^2 \quad (72)$$

We therefore have that:

$$u_X(x, X, \theta) = \alpha_X(X, \theta) - \beta_X(X, \theta)(x - \gamma(X, \theta))^2 + 2\gamma_X(X, \theta)\beta(X, \theta)(x - \gamma(X, \theta)) \quad (73)$$

Plugging this into the necessary condition and evaluating at the equilibrium aggregate

$\hat{X}(\theta) = X(P(\theta))$ , we obtain:

$$0 = \int_{\mathcal{X}} \left[ \alpha_X(X(P(\theta)), \theta) - \beta_X(X(P(\theta)), \theta) (\tilde{x} - \gamma(X(P(\theta)), \theta))^2 + 2\gamma_X(X(P(\theta)), \theta) \beta(X(P(\theta)), \theta) (\tilde{x} - \gamma(X(P(\theta)), \theta)) \right] dP(\tilde{x}|\theta) \quad (74)$$

Which can be rewritten in terms of the equilibrium bias and variance with respect to  $\gamma$  as:

$$0 = \alpha_X(X(P(\theta)), \theta) - \beta_X(X(P(\theta)), \theta) (\text{Disp}[P(\theta), \theta])^2 + 2\gamma_X(X(P(\theta)), \theta) \beta(X(P(\theta)), \theta) \text{Bias}[P(\theta), \theta] \quad (75)$$

as desired.

That all of the above hold for any  $\lambda > 0$  follows by the observation that if  $\hat{\phi}$  satisfies Assumption 3, then so too does  $\lambda\hat{\phi}$  for any  $\lambda > 0$ .  $\square$

## A.7 Proof of Proposition 2

*Proof.* To establish that  $\Omega$  is an equilibrium of the mutual information model, it is sufficient to establish that  $P$  solves each individual's optimization problem when they take  $\hat{X}$  as given. By Corollary 2 in Matějka and McKay (2015), all interior unconditional choice probabilities  $p(x) = \sum_{\theta \in \Theta} p(x|\theta)\pi(\theta)$  in the mutual information model satisfy the following first-order condition:

$$p(x|\theta) = \frac{p(x) \exp\{u(x, \hat{X}(\theta), \theta)/\lambda\}}{\sum_{\tilde{x} \in \mathcal{X}} p(\tilde{x}) \exp\{u(\tilde{x}, \hat{X}(\theta), \theta)/\lambda\}} \quad (76)$$

and the following additional constraint:

$$\sum_{\theta \in \Theta} \frac{\exp\{u(x, \hat{X}(\theta), \theta)/\lambda\}}{\sum_{\tilde{x} \in \mathcal{X}} p(\tilde{x}) \exp\{u(\tilde{x}, \hat{X}(\theta), \theta)/\lambda\}} \pi^\theta(\theta) = 1 \quad (77)$$

Observe that, if and only if  $p(x) = p(x^\theta)$  for all  $x, x^\theta \in \mathcal{X}$ , then the choice probabilities that solve (76) are

$$p(x|\theta) = \frac{\exp\{u(x, \hat{X}(\theta), \theta)/\lambda\}}{\sum_{\tilde{x} \in \mathcal{X}} \exp\{u(\tilde{x}, \hat{X}(\theta), \theta)/\lambda\}} \quad (78)$$

This would verify that the stochastic choice rule  $P$  is a unique, interior solution to agents' choice problem. Hence it remains only to verify that  $p(x) = p(x^\theta)$  for all  $x, x^\theta \in \mathcal{X}$ , or exchangeability, in the agent's optimal program.

It is straightforward to derive such a condition using (77). Stacking equation (77) over

all interior  $x \in \mathcal{X}$ , we obtain the system:

$$\tilde{U}(\{p(x)\}_{x \in \mathcal{X}})\pi^\theta = 1 \quad (79)$$

where:

$$\tilde{u}_{x_i, \theta_j}(\{p(x)\}_{x \in \mathcal{X}}) = \frac{\exp\{u(x_i, \hat{X}(\theta_j), \theta_j)/\lambda\}}{\sum_{x_k \in \mathcal{X}} p(x_k) \exp\{u(x_k, \hat{X}(\theta_j), \theta_j)/\lambda\}} \quad (80)$$

and 1 is a  $|\Theta|$  length vector. Thus, there exists a prior consistent with uniform unconditional choice  $p(x) = \frac{1}{|\mathcal{X}|}$  if and only if the following linear system has a solution probability vector  $\pi^\theta \in \Delta(\Theta)$ :

$$\tilde{U}\pi^\theta = |\mathcal{X}|^{-1} \mathbf{1} \quad (81)$$

where  $\mathbf{1}$  is a  $|\Theta|$  length vector, and  $\tilde{U}$  is as stated in the result. This completes the proof, with  $\pi^\theta$  solving the given system supporting the equilibrium under the mutual information model.  $\square$

## References

- ACEMOGLU, D., AND M. K. JENSEN (2010): “Robust comparative statics in large static games,” in *49th IEEE Conference on Decision and Control (CDC)*, pp. 3133–3139. IEEE.
- (2013): “Aggregate comparative statics,” *Games and Economic Behavior*, 81, 27–49.
- (2015): “Robust comparative statics in large dynamic economies,” *Journal of Political Economy*, 123(3), 587–640.
- ALLEN, F., S. MORRIS, AND H. S. SHIN (2006): “Beauty contests and iterated expectations in asset markets,” *The Review of Financial Studies*, 19(3), 719–752.
- ALLEN, R., AND J. REHBECK (2021): “A Generalization of Quantal Response Equilibrium via Perturbed Utility,” *Games*, 12(1), 20.
- ANGELETOS, G.-M., AND J. LA’O (2010): “Noisy business cycles,” *NBER Macroeconomics Annual*, 24(1), 319–378.
- (2013): “Sentiments,” *Econometrica*, 81(2), 739–779.
- ANGELETOS, G.-M., AND C. LIAN (2016): “Incomplete information in macroeconomics: Accommodating frictions in coordination,” in *Handbook of Macroeconomics*, vol. 2, pp. 1065–1240. Elsevier.
- ANGELETOS, G.-M., AND C. LIAN (2018): “Forward guidance without common knowledge,” *American Economic Review*, 108(9), 2477–2512.

- ANGELETOS, G.-M., AND A. PAVAN (2007): “Efficient use of information and social value of information,” *Econometrica*, 75(4), 1103–1142.
- ANGELETOS, G.-M., AND K. A. SASTRY (2019): “Inattentive Economies,” Working Paper 26413, National Bureau of Economic Research.
- BENHABIB, J., P. WANG, AND Y. WEN (2015): “Sentiments and aggregate demand fluctuations,” *Econometrica*, 83(2), 549–585.
- BERGEMANN, D., AND S. MORRIS (2013): “Robust predictions in games with incomplete information,” *Econometrica*, 81(4), 1251–1308.
- BERGER, D., AND J. VAVRA (2019): “Shocks versus Responsiveness: What Drives Time-Varying Dispersion?,” *Journal of Political Economy*, 127(5), 2104–2142.
- CAPLIN, A., AND M. DEAN (2015): “Revealed Preference, Rational Inattention, and Costly Information Acquisition,” *American Economic Review*, 105(7), 2183–2203.
- CAPLIN, A., M. DEAN, AND J. LEAHY (2017): “Rationally inattentive behavior: Characterizing and generalizing Shannon entropy,” Working Paper 23652, National Bureau of Economic Research.
- (2019): “Rational inattention, optimal consideration sets, and stochastic choice,” *The Review of Economic Studies*, 86(3), 1061–1094.
- CARLSSON, H., AND E. VAN DAMME (1993): “Global games and equilibrium selection,” *Econometrica*, pp. 989–1018.
- CHAHROUR, R., AND R. ULBRICHT (2019): “Robust predictions for DSGE models with incomplete information,” Working Paper 3292550, SSRN.
- DENTI, T. (2019): “Posterior-separable cost of information,” Working paper, Cornell University.
- DENTI, T., M. MARINACCI, AND A. RUSTICHINI (2019): “Experimental cost of information,” Working paper, Cornell University.
- FARHI, E., AND I. WERNING (2019): “Monetary Policy, Bounded Rationality, and Incomplete Markets,” *American Economic Review*, 109(11), 3887–3928.
- FLYNN, J. P., AND K. A. SASTRY (2021): “Attention Cycles,” Working Paper 3592107, SSRN.
- FRANKEL, D. M., S. MORRIS, AND A. PAUZNER (2003): “Equilibrium selection in global games with strategic complementarities,” *Journal of Economic Theory*, 108(1), 1–44.
- FUDENBERG, D., R. IJIMA, AND T. STRZALECKI (2015): “Stochastic choice and revealed perturbed utility,” *Econometrica*, 83(6), 2371–2409.
- GABAIX, X. (2014): “A sparsity-based model of bounded rationality,” *The Quarterly Journal of Economics*, 129(4), 1661–1710.

- (2019): “Behavioral inattention,” in *Handbook of Behavioral Economics: Applications and Foundations 1*, vol. 2, pp. 261–343. Elsevier.
- (2020): “A Behavioral New Keynesian Model,” *American Economic Review*, 110(8), 2271–2327.
- GARCIA-SCHMIDT, M., AND M. WOODFORD (2019): “Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis,” *American Economic Review*, 109(1), 86–120.
- HARSANYI, J. C. (1973): “Oddness of the number of equilibrium points: a new proof,” *International Journal of Game Theory*, 2(1), 235–250.
- HÉBERT, B. M., AND J. LA’O (2020): “Information Acquisition, Efficiency, and Non-Fundamental Volatility,” Discussion paper, National Bureau of Economic Research.
- HÉBERT, B. M., AND M. WOODFORD (2020): “Neighborhood-Based Information Costs,” Discussion paper, National Bureau of Economic Research.
- HELLWIG, C., AND L. VELDKAMP (2009): “Knowing what others know: Coordination motives in information acquisition,” *The Review of Economic Studies*, 76(1), 223–251.
- HICKS, J. R. (1937): “Mr. Keynes and the ‘classics’; a suggested interpretation,” *Econometrica: journal of the Econometric Society*, pp. 147–159.
- HUO, Z., AND M. PEDRONI (2020): “A single-judge solution to beauty contests,” *American Economic Review*, 110(2), 526–68.
- ILUT, C. L., AND R. VALCHEV (2020): “Economic Agents as Imperfect Problem Solvers,” Working Paper 27820, National Bureau of Economic Research.
- JENSEN, M. K. (2018): “Aggregative games,” in *Handbook of Game Theory and Industrial Organization, Volume 1*. Edward Elgar Publishing.
- JUNG, J., J. H. J. KIM, F. MATĚJKA, AND C. A. SIMS (2019): “Discrete Actions in Information-Constrained Decision Problems,” *The Review of Economic Studies*, 86(6), 2643–2667.
- KACPERCZYK, M., S. VAN NIEUWERBURGH, AND L. VELDKAMP (2016): “A rational theory of mutual funds’ attention allocation,” *Econometrica*, 84(2), 571–626.
- KEYNES, J. M. (1936): *The General Theory of Employment, Interest, and Money*. Macmillan, London.
- KIMBALL, M. (1990): “Precautionary Saving in the Small and in the Large,” *Econometrica*, 58(1), 53–73.
- MACAULAY, A. (2020): “Cyclical attention to saving,” Working paper, Oxford University.
- MAĆKOWIAK, B., F. MATEJKA, AND M. WIEDERHOLT (2020): “Rational Inattention: A Review,” Discussion Paper 15408, CEPR.
- MAĆKOWIAK, B., AND M. WIEDERHOLT (2009): “Optimal sticky prices under rational inattention,” *American Economic Review*, 99(3), 769–803.

- MATĚJKA, F. (2015): “Rationally Inattentive Seller: Sales and Discrete Pricing,” *The Review of Economic Studies*, 83(3), 1125–1155.
- MATĚJKA, F., AND A. MCKAY (2015): “Rational inattention to discrete choices: A new foundation for the multinomial logit model,” *American Economic Review*, 105(1), 272–98.
- MATHEVET, L. (2010): “A contraction principle for finite global games,” *Economic Theory*, 42(3), 539.
- MATTSSON, L.-G., AND J. W. WEIBULL (2002): “Probabilistic choice and procedurally bounded rationality,” *Games and Economic Behavior*, 41(1), 61–78.
- McFADDEN, D. L. (1973): “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, ed. by P. Zarembka, pp. 105–142. Academic Press, New York.
- MCKELVEY, R. D., AND T. R. PALFREY (1995): “Quantal response equilibria for normal form games,” *Games and Economic Behavior*, 10(1), 6–38.
- MELO, E. (2020): “On the Uniqueness of Quantal Response Equilibria,” Working Paper 3631575, SSRN.
- MORRIS, S., R. ROB, AND H. S. SHIN (1995): “p-Dominance and belief potential,” *Econometrica*, pp. 145–157.
- MORRIS, S., AND H. S. SHIN (1998): “Unique equilibrium in a model of self-fulfilling currency attacks,” *American Economic Review*, pp. 587–597.
- (2002): “Social value of public information,” *American Economic Review*, 92(5), 1521–1534.
- MORRIS, S., AND M. YANG (2019): “Coordination and Continuous Stochastic Choice,” Working Paper 2889861, SSRN.
- MUNRO, D. R. (2021): “Consumer Behavior and Firm Volatility,” *Journal of Money, Credit and Banking*, 53(4), 845–873.
- MYATT, D. P., AND C. WALLACE (2012): “Endogenous information acquisition in coordination games,” *The Review of Economic Studies*, 79(1), 340–374.
- MYERSON, R. B. (1978): “Refinements of the Nash equilibrium concept,” *International Journal of Game Theory*, 7(2), 73–80.
- PALFREY, T. R., C. A. HOLT, AND J. GOEREE (2016): *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press.
- POMATTO, L., P. STRACK, AND O. TAMUZ (2018): “The cost of information,” Working paper, arXiv:1812.04211.
- RAITH, M. (1996): “A general model of information sharing in oligopoly,” *Journal of Economic Theory*, 71(1), 260–288.
- SELTEN, R. (1975): “Reexamination of the perfectness concept for equilibrium points in extensive games,” *International Journal of Game Theory*, 4, 25–55.

- SIMS, C. A. (2003): “Implications of rational inattention,” *Journal of Monetary Economics*, 50(3), 665–690.
- STAHL, D. O. (1990): “Entropy control costs and entropic equilibria,” *International Journal of Game Theory*, 19(2), 129–138.
- STEVENS, L. (2019): “Coarse Pricing Policies,” *The Review of Economic Studies*, 87(1), 420–453.
- STOKEY, N. L., R. E. LUCAS, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.
- STROEBEL, J., AND J. VAVRA (2019): “House prices, local demand, and retail prices,” *Journal of Political Economy*, 127(3), 1391–1436.
- TIROLE, J. (2015): “Cognitive Games and Cognitive Traps,” Working paper, Toulouse School of Economics.
- VAN DAMME, E. (1991): *Stability and perfection of Nash equilibria*, vol. 339. Springer.
- VAN ZANDT, T., AND X. VIVES (2007): “Monotone equilibria in Bayesian games of strategic complementarities,” *Journal of Economic Theory*, 134(1), 339–360.
- VIVES, X. (1984): “Duopoly information equilibrium: Cournot and Bertrand,” *Journal of Economic Theory*, 34(1), 71–94.
- WEINSTEIN, J., AND M. YILDIZ (2007): “Impact of higher-order uncertainty,” *Games and Economic Behavior*, 60(1), 200–212.
- WOODFORD, M. (2003): “Imperfect Common Knowledge and the Effects of Monetary Policy,” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*.
- (2012): “Inattentive valuation and reference-dependent choice,” Working paper, Columbia University.
- (2020): “Modeling imprecision in perception, valuation, and choice,” *Annual Review of Economics*, 12, 579–601.
- YANG, M. (2015): “Coordination with flexible information acquisition,” *Journal of Economic Theory*, 158, 721–738.

# Online Appendix to *Strategic Mistakes* by Flynn and Sastry

## B Strategic Mistakes in Binary-Action Games

In this Appendix, we adapt our analysis to study binary-action games, which are also common for modeling coordination phenomena in macroeconomics and finance. We first provide results ensuring existence, uniqueness and monotone comparative statics. We next apply our results to study the “investment game,” introduced by Carlsson and Van Damme (1993) and studied recently by Yang (2015) and Morris and Yang (2019). Bridging our continuous-action and binary-action analyses, we finally discuss how the the action space can have a large bearing on our model’s uniqueness predictions. This may be an important consideration for researchers when the choice of action space is primarily based on analytical convenience and not descriptive realism regarding adjustment on an extensive margin.

### B.1 Existence, Uniqueness, and Comparative Statics

We now study the same environment as Section 2 with the sole change that agents now have a binary action set  $\mathcal{X} = \{0, 1\}$ .<sup>41</sup> Let  $p(\theta)$  denote the probability that a given agent plays action 1 in state  $\theta$ . It is without loss of generality to restrict to the aggregator  $X(p(\theta)) = p(\theta)$ , since transformations of this aggregate can be applied within payoffs, and we adopt this convention throughout. Given a conjecture for the law of motion of the aggregate  $\hat{p}$  and state  $\theta \in \Theta$ , we define the benefit of playing action 1 over action 0 as:

$$\Delta u(\hat{p}(\theta), \theta) \equiv u(1, \hat{p}(\theta), \theta) - u(0, \hat{p}(\theta), \theta) \quad (82)$$

We let  $\Delta u_X$  denote this function’s derivative in the first argument.

We now provide an existence and uniqueness result. To do so, we place the following regularity condition on the stochastic choice functional:<sup>42</sup>

**Assumption 6.** *The kernel of the cost functional satisfies the Inada condition  $\lim_{x \rightarrow 0} \phi^0(x) = -\infty$ . Moreover,  $\phi^0$  is globally strictly convex.*<sup>43</sup>

---

<sup>41</sup>Naturally, all integrals are now replaced with summations and density functions by mass functions.

<sup>42</sup>For existence, this can be weakened in the obvious way: the objective need only be continuous. We present results with this stronger assumption for brevity.

<sup>43</sup>Note that, in view of the Inada condition, it is impossible for  $\phi''$  to be globally strictly concave.



This rules out stochastic choice rule's being concentrated on only one of the two actions in any state. The result follows:<sup>44</sup>

**Proposition 3.** *Suppose that  $\phi$  satisfies assumption 6 and  $\Delta u(p, \theta)$  is continuously differentiable in its first argument. There exists an equilibrium. All equilibria are symmetric. A sufficient condition for there to be a unique  $p(\theta)$  is that:*

$$\max_{p \in [0,1]} \Delta u_X(p, \theta) < 2\phi''\left(\frac{1}{2}\right) \quad (83)$$

A sufficient condition for there to be a unique  $p$  is that (83) holds for all  $\theta \in \Theta$ .

*Proof.* Under Assumption 6, for any  $\theta$ , we have that  $p(\theta) \in (0, 1)$ . Thus equilibrium is characterized by the first-order condition obtained by moving probability of playing zero to playing one. Thus, the condition characterizing equilibrium is given by:

$$\Delta u(p(\theta), \theta) = \phi'(p(\theta)) - \phi'(1 - p(\theta)) \quad (84)$$

To prove uniqueness for a given  $\theta$  it is sufficient to prove that the minimal slope of the RHS exceeds the maximal slope of the LHS:

$$\max_{p \in [0,1]} \Delta u_X(p, \theta) < \min_{p \in [0,1]} \phi''(p) + \phi''(1 - p) \quad (85)$$

If  $\phi''$  is strictly convex, then the problem is solved by solving the FOC:

$$\phi'''(p) = \phi'''(1 - p) \quad (86)$$

As  $\phi''$  is strictly convex,  $\phi'''$  is strictly increasing and is therefore invertible. Thus the unique solution is  $p = \frac{1}{2}$  and the minimized value is  $2\phi''(\frac{1}{2})$ . Applying this argument state by state yields the global condition.  $\square$

Condition (83) checks the maximum value of complementarity (left-hand-side) against the lowest value for the slope of the marginal cognitive cost of investing (right-hand-side), which is realized at  $p = \frac{1}{2}$ .<sup>45</sup> We will provide a simple graphical intuition for this condition in the upcoming example.

---

<sup>44</sup>One can extend this result in the obvious way beyond the differentiability assumption to allow for Lipschitz continuous  $\Delta u(p, \theta)$ . Naturally, the key property being ruled out is a sudden threshold around which the gains from playing action 1 change discontinuously.

<sup>45</sup>That  $p = \frac{1}{2}$  is such a point can be derived by noting the symmetry of the likelihood-separable cost around  $p = \frac{1}{2}$  and the convexity of  $\phi$ .

It is moreover simple to establish when the aggregate  $p(\theta)$  increases in  $\theta$ . As in our main analysis, this simply requires supermodularity of payoffs in  $(x, p, \theta)$ , or that higher actions by others and states are complementary with playing  $x = 1$ :

**Assumption 7** (Joint Supermodularity). *The benefit of playing action 1 over action 0 satisfies, for all  $p^\theta \geq p, \theta^\theta \geq \theta$ :*

$$\Delta u(p^\theta, \theta^\theta) \geq \Delta u(p, \theta) \quad (87)$$

**Proposition 4.** *Suppose that Assumptions 6 and 7 hold, and that the inequality in Equation 83 holds for all  $\theta \in \Theta$  so that there is a unique equilibrium  $p$ . The unique equilibrium  $p(\theta)$  is monotone increasing in  $\theta$ .*

*Proof.* Under Assumption 6, the equilibrium is characterized by Equation 84. Under the assumption that the inequality in Equation 83 holds, there is a unique solution  $p(\theta)$  for all  $\theta \in \Theta$ . Note that that this unique equilibrium occurs when  $\Delta u(p, \theta)$  intersects  $\phi^\theta(p) - \phi^\theta(1-p)$  from above. Moreover, by Assumption 7 we know that  $\Delta u(p, \theta)$  is increasing in  $(p, \theta)$ . Thus, when we take  $\theta^\theta \geq \theta$ , we know that the unique intersection occurs for  $p(\theta^\theta) \geq p(\theta)$ .  $\square$

Analogous results with general information acquisition or stochastic choice, by contrast, require more extensive analysis (see, e.g., Yang, 2015; Morris and Yang, 2019).

## B.2 Application: The Investment Game

We now apply these results in a variant of the binary-action investment game introduced by Carlsson and Van Damme (1993), which like our example in Section 4.1 models coordination motives in financial speculation. Each agent chooses an action  $x \in \{0, 1\}$ , or “not invest” and “invest.” The state of nature  $\theta \in \Theta \subseteq \mathbb{R}$  scales the desirability of investing independent of other conditions. Agents’ payoffs depend on the action, the total fraction of investing agents, and the state of nature separably and linearly:

$$u(x, p, \theta) = x(\theta - r(1 - p)) \quad (88)$$

where  $r \geq 0$  scales the degree of strategic complementarity between investment decisions.

It is straightforward to derive the following fixed-point equation that describes the equilibria of the model when  $\phi$  satisfies the Inada condition in Assumption 6:

$$\theta + rp(\theta) - r = \phi^\theta(p(\theta)) - \phi^\theta(1 - p(\theta)) \quad (89)$$

Equilibrium is guaranteed to be unique by Proposition 4 provided that the following condition holds relating strategic complementarity  $r$  with the second derivative of the kernel  $\phi$ :

$$r < 2\phi''\left(\frac{1}{2}\right) \quad (90)$$

This condition is independent of the state space  $\Theta$  or the prior. But it does depend on the scale and character of cognitive costs through  $\phi''\left(\frac{1}{2}\right)$ .

Condition (90) admits the following interpretation about uniqueness with vanishing costs under arbitrary functional forms. For any positive (but arbitrarily small) level of strategic complementarity, and with a sufficiently rich state space, there will be multiple equilibria for a sufficiently small cost of stochastic choice:

**Corollary 2.** *Consider a family of investment games  $\{\mathcal{G}_\lambda : \lambda \in (0, L]\}$  with fixed payoff  $s$ , action space, and state space, each with the re-scaled cost functional for some common  $\hat{\phi}$  that satisfies Assumption 6, i.e.  $\phi_\lambda = \lambda\hat{\phi}$ . Then, for all*

$$\lambda > L := \frac{r}{2\hat{\phi}''\left(\frac{1}{2}\right)} \quad (91)$$

*game  $\mathcal{G}_\lambda$  has a unique action profile  $(p(\theta))_{\theta \in \Theta}$ . Conversely, when  $\lambda < L$ , there exists at least some  $\theta \in \Theta$  such that the equilibrium of  $\mathcal{G}_\lambda$  is not unique if  $\theta \in \Theta$ .*

*Proof.* Recall that for any  $\phi_\lambda$  (owing to  $\hat{\phi}$  satisfying Assumption 6), we have that:

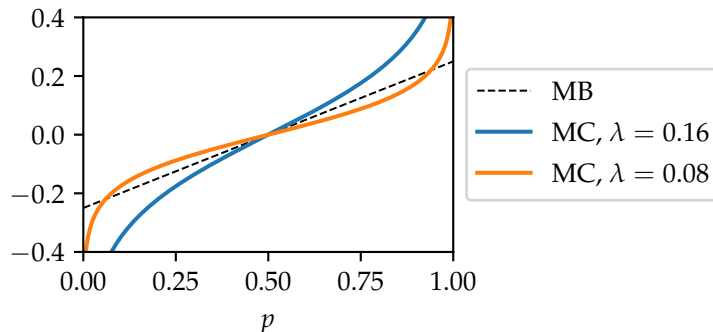
$$\theta + rp(\theta) - r = \phi_\lambda^l(p(\theta)) - \phi_\lambda^l(1 - p(\theta)) \quad (92)$$

Consider state  $\theta = \frac{r}{2}$ . In this state, we have that  $p(\theta) = \frac{1}{2}$  is an equilibrium. Moreover, see that the slope of the LHS in  $p$  is given by  $r$  and the slope of the RHS in  $p$  at  $p = \frac{1}{2}$  is given by  $2\lambda\hat{\phi}''\left(\frac{1}{2}\right)$ . Hence, when  $\lambda < \frac{r}{2\hat{\phi}''\left(\frac{1}{2}\right)}$ , we have that the slope of the LHS exceeds the slope of the RHS. But we know that the RHS is continuous on  $(0, 1)$  and that  $\lim_{p \rightarrow 1} \phi^l(p) - \phi^l(p) = \infty$ . Thus, the RHS must intersect the LHS from below for some other  $p \in \left(\frac{1}{2}, 1\right)$ . Thus, in state  $\theta$ , if  $\lambda < \frac{r}{2\hat{\phi}''\left(\frac{1}{2}\right)}$  there are multiple  $p(\theta)$  that can arise in equilibrium. Consequently, if  $\theta \in \Theta$  and  $\lambda < \frac{r}{2\hat{\phi}''\left(\frac{1}{2}\right)}$ , we have that equilibrium is not globally unique. The final claim that we have global uniqueness for  $\lambda > \frac{r}{2\hat{\phi}''\left(\frac{1}{2}\right)}$  follows immediately from Theorem 3.  $\square$

The result contrasts with Corollary 1 which showed limit uniqueness in the generalized beauty contest. We will further discuss this issue in Section B.4.

To illustrate the uniqueness result, we consider a specialization of the model in which the kernel function is  $\hat{\phi}(x) = x \log x$ . In this case,  $\phi''(0.5) = 2$  and the cost threshold

**Figure 5:** Multiplicity in the Investment Game



*Note:* The dotted line is the marginal benefits of investing more often as a function of others’ investment probability, or the right-hand side of (89). The blue and orange lines are the marginal costs of investing more often under respectively more and less severe costs of stochastic choice. Each intersection is an equilibrium.

for uniqueness is  $L = \frac{\tau}{4}$ .<sup>46</sup> Figure 5 illustrates the scope for multiplicity in a benchmark parameter case of this logit model. We fix  $r = 0.50$ , and  $\theta = 0.25$ , the state such that a 50% aggregate investment corresponds with zero payoff. The dotted black line is the “Marginal Benefit,” which corresponds with the left-hand-side of (89). The blue and orange lines are the the “Marginal Cost” of increasing the investing probability, or the right-hand-side of (89), with respectively higher and lower values of  $\lambda$  or costs of attention. By construction, there is an equilibrium with  $p = \frac{1}{2}$  for any value of  $\lambda$ . Whether or not there are additional equilibria corresponding to more “confident” play, or  $p$  closer to 0 or 1, depends on the *slope* of these marginal costs. When  $\lambda$  is high (blue line), it is costly to play more certainly and hence there is only one intersection with the dotted line. When  $\lambda$  is low (orange line), marginal costs cross marginal benefits from above at  $p = 0.5$ . This visualizes a violation of the condition in Proposition 3. As a result there are two more “confident” equilibria near  $p = 0$  and  $p = 1$ .

The right-hand-side of the confident-wavering condition (90) is a well-defined moment which researchers may try to calibrate via laboratory experiments and could interpret in our model without taking a stand on the entire  $\phi$  function. In this way, (90) can be read as a sufficient statistic gauge of the potential for multiplicity and fragility that relies only on one informative aspect of the underlying stochastic choice model.

<sup>46</sup>This is exactly the condition obtained by Yang (2015) for this game with information acquisition costs proportional to mutual information. This foreshadows a deeper connection which we will explore in the next subsection.

### B.3 Strategic Mistakes *vs.* Mutual Information

In the vein of our main analysis' comparison of beauty contests with strategic mistakes and mutual information, we now compare the investment game under logit strategic mistakes with the equivalent game under mutual information, as studied by Yang (2015). Observe first that the mutual information model does not always admit an interior solution. Intuitively, if agents place an arbitrarily high prior weight on fundamentals always being very high or very low, they may decide to unconditionally invest or dis-invest without learning anything. These scenarios are ruled out by respectively assuming  $E_\pi[\exp\{\lambda^{-1}\theta\}] > \exp\{\lambda^{-1}r\}$  and  $E_\pi[\exp\{-\lambda^{-1}\theta\}] > 1$ . No analogue of either is possible in the strategic mistakes model with logistic choice which always features positive probability of playing both actions in all states, so these conditions *a fortiori* rule out an application of Proposition 2. Nonetheless, after ruling out these cases, we can show the following:

**Corollary 3.** *Compare identical investment games  $\mathcal{G}^{LSM}$  and  $\mathcal{G}^{MI}$ , distinguished by their costs of stochastic choice, scaled by a common scalar  $\lambda$ . Assume*

1. (Interiority)  $E_\pi[\exp\{\lambda^{-1}\theta\}] > \exp\{\lambda^{-1}r\}$  and  $E_\pi[\exp\{-\lambda^{-1}\theta\}] > 1$
2. (Global uniqueness)  $r < 4\lambda$

Each game has a unique equilibrium  $(p^{LSM}(\cdot), p^{MI}(\cdot))$ . Moreover,

$$\begin{cases} p^{LSM}(\theta) = p^{MI}(\theta), \forall \theta & \text{if } \sum p^{MI}(\theta) \pi(\theta) = 1/2, \\ p^{LSM}(\theta) < p^{MI}(\theta), \forall \theta & \text{if } \sum p^{MI}(\theta) \pi(\theta) > 1/2, \\ p^{LSM}(\theta) > p^{MI}(\theta), \forall \theta & \text{if } \sum p^{MI}(\theta) \pi(\theta) < 1/2. \end{cases} \quad (93)$$

*Proof.* It follows from Proposition 2 of Yang (2015), that when  $E_\pi[\exp\{\lambda^{-1}\theta\}] > \exp\{\lambda^{-1}r\}$  and  $E_\pi[\exp\{-\lambda^{-1}\theta\}] > 1$ , the equilibria of the game with mutual information cost are characterized by:

$$\theta + rp^{MI}(\theta) - r = \lambda \left[ \ln \left( \frac{p^{MI}(\theta)}{1 - p^{MI}(\theta)} \right) - \ln \left( \frac{\bar{p}^{MI}}{1 - \bar{p}^{MI}} \right) \right] \quad (94)$$

for all  $\theta \in \Theta$  where  $\bar{p}^{MI} = \sum p^{MI}(\theta)\pi(\theta)$ . It moreover follows from Proposition 3 of Yang (2015) that when  $r < 4\lambda$ , this model features a unique equilibrium. Recall that when  $r < 4\lambda$  our model with entropic stochastic choice also features a unique equilibrium and this is characterized by:

$$\theta + rp^L(\theta) - r = \lambda \left[ \ln \left( \frac{p^L(\theta)}{1 - p^L(\theta)} \right) \right] \quad (95)$$

Moreover, when  $\bar{p}^{\text{MI}} > \frac{1}{2}$ , we have that  $\ln\left(\frac{p^{\text{MI}}}{1-p^{\text{MI}}}\right) > 0$ , when  $\bar{p}^{\text{MI}} = \frac{1}{2}$ , we have that  $\ln\left(\frac{p^{\text{MI}}}{1-p^{\text{MI}}}\right) = 0$  and when  $\bar{p}^{\text{MI}} < \frac{1}{2}$ , we have that  $\ln\left(\frac{p^{\text{MI}}}{1-p^{\text{MI}}}\right) < 0$ . It is then immediate that  $p^L(\theta) < p^{\text{MI}}(\theta)$  when  $\bar{p}^{\text{MI}} > \frac{1}{2}$ ,  $p^L(\theta) = p^{\text{MI}}(\theta)$  when  $\bar{p}^{\text{MI}} = \frac{1}{2}$ , and  $p^L(\theta) > p^{\text{MI}}(\theta)$  when  $\bar{p}^{\text{MI}} < \frac{1}{2}$ .  $\square$

Conditional on interiority, anchoring in the mutual information model distorts the choice probabilities but perhaps more surprisingly is completely separable from the game’s uniqueness properties. More formally, in binary-action games with mutual information, the only difference between the strategic mistakes model with entropy is that log-odds ratio  $\log\left(\frac{p(\theta)}{1-p(\theta)}\right)$  in state  $\theta \in \Theta$  differs across the models by a state-independent additive constant. In our earlier graphical analysis, this can be seen as a vertical shift of the marginal cost curve. Thus, our confident wavering argument applies directly to the mutual information model and offers an alternative window into the main result of Yang (2015). This separability of anchoring from uniqueness properties with binary actions may be an independently useful insight in other models with mutual information cost.

## B.4 Discussion: Global vs. Local Mistakes

Binary-action settings are sometimes used as a convenient metaphor for underlying environments with many possible actions—for instance, simplifying financial speculation as the choice between extremes of investing and dis-investing instead of a continuous portfolio choice. Our analysis reveals that, in models of stochastic choice, the restriction to two extreme actions may significantly change the character of the game because it removes the possibility of local substitution of actions. The binary-action game allows for “global mistakes,” like fully investing when fully disinvesting is instead optimal, that impose discontinuously different externalities and can support multiple equilibria. Our benchmark continuous-action model implies by contrast that agents make “local mistakes” like substituting an optimal action with an alternative that is sub-optimal but nearby in the action space. Whether an analyst should use the binary-action or continuous-action model then depends on the problem at hand and how seriously they take the prediction of global substitution relative to the potential loss in tractability.

Our results also contrast with those in the global games literature in which there is, instead of stochastic choice, vanishing private measurement error in observing the fundamental (Carlsson and Van Damme, 1993; Morris, Rob, and Shin, 1995; Frankel, Morris, and Pauzner, 2003). When combined with the earlier observation linking strategic mistakes with cross-sectional heterogeneity in payoff functions (Section 2.1), our results draw a sharp distinction between measurement errors for payoffs (studied here, which do not yield limit uniqueness)

and measurement errors for fundamentals (studied in the aforementioned literature, which do yield limit uniqueness). One way of thinking about the difference is that the “contagion” argument formalized in the above references, which shows that having dominant actions in specific states iteratively implies unique rationalizable actions in neighboring states, has no analogue in the present model with no interim beliefs or cross-state reasoning. A different interpretation is that the mere observation that agents have trembling hands is not sufficient to imply the sharp and specific predictions of canonical global games, a point also made by Yang (2015) and Morris and Yang (2019).

## C Quasi-MLRP Kernels

In this Appendix, we state and prove one main result and one supporting result which help describe the class of stochastic choice kernels satisfying quasi-MLRP.

**Lemma 2.**  *$\mathcal{F}$ , the class of functions satisfying quasi-MLRP, is non-empty and strictly larger than  $\mathcal{F}^{MLRP} = \{\log(\cdot)\}$ , the class of functions satisfying MLRP.*

*Proof.* To establish both claims, it is sufficient to show  $f(x) = \log(x)$  and  $f(x) = x$  satisfy quasi-MLRP. To see that quasi-MLRP is satisfied for  $f(x) = \log x$  is immediate as the quasi-MLRP property is just MLRP. Indeed for quasi-MLRP we require for any two distributions  $g^\theta, g \in \Delta(\mathcal{X})$ :

$$\left( f(g^\theta(x^\theta)) - f(g^\theta(x)) \geq f(g(x^\theta)) - f(g(x)) \forall x^\theta \geq x \right) \implies g^\theta \succeq_{FOSD} g \quad (96)$$

With  $f(x) = \log(x)$  this becomes:

$$\left( \frac{g^\theta(x^\theta)}{g^\theta(x)} \geq \frac{g(x^\theta)}{g(x)} \forall x^\theta \geq x \right) \implies g^\theta \succeq_{FOSD} g \quad (97)$$

The left hand side of this implication is simply the MLRP property. MLRP implies FOSD. Thus, one sees immediately that quasi-MLRP is a weakening of MLRP.

That this weakening is strict is shown by proving that  $f(x) = x$  satisfies quasi-MLRP. This requires us to prove that for any two distributions  $g^\theta, g \in \Delta(\mathcal{X})$ :

$$\left( g^\theta(x^\theta) - g^\theta(x) \geq g(x^\theta) - g(x) \forall x^\theta \geq x \right) \implies g^\theta \succeq_{FOSD} g \quad (98)$$

To do this, we first prove a technical lemma which may be of use for characterizing other functions that satisfy quasi-MLRP:

**Lemma 3.** For any two distributions  $g^\theta, g \in \Delta(\mathcal{X})$ , the following holds:

$$\left( f(g^\theta(x^\theta)) - f(g^\theta(x)) \geq f(g(x^\theta)) - f(g(x)) \quad \forall x^\theta \geq x \right) \implies \left( \frac{\int_{\underline{x}}^{\bar{x}} [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{\bar{x} - x} \geq \frac{\int_{\underline{x}}^x [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{x - \underline{x}} \quad \forall x \in \mathcal{X} \right) \quad (99)$$

*Proof.* To prove the required implication, we begin with the hypothesis:

$$f(g^\theta(x^\theta)) - f(g^\theta(x)) \geq f(g(x^\theta)) - f(g(x)) \quad \forall x^\theta \geq x \quad (100)$$

Which can be rewritten as:

$$f(g^\theta(x^\theta)) + f(g(x)) \geq f(g(x^\theta)) + f(g^\theta(x)) \quad \forall x^\theta \geq x \quad (101)$$

We now integrate from  $\underline{x}$  to  $x^\theta$  with respect to  $x$  to obtain the inequality:

$$(x^\theta - \underline{x})f(g^\theta(x^\theta)) + \int_{\underline{x}}^{x^\theta} f(g(x)) dx \geq (x^\theta - \underline{x})f(g(x^\theta)) + \int_{\underline{x}}^{x^\theta} f(g^\theta(x)) dx \quad (102)$$

Imposing  $x^\theta = x$  we obtain:

$$(x - \underline{x}) [f(g^\theta(x)) - f(g(x))] \geq \int_{\underline{x}}^x [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x} \quad (103)$$

Applying the same procedure but this time integrating from  $x$  to  $\bar{x}$  with respect to  $x^\theta$  and evaluate at  $x^\theta = x$  to obtain this inequality:

$$\int_x^{\bar{x}} [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x} \geq (\bar{x} - x) [f(g^\theta(x)) - f(g(x))] \quad (104)$$

Combining our two inequalities we obtain the required one:

$$\frac{\int_x^{\bar{x}} [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{\bar{x} - x} \geq \frac{\int_{\underline{x}}^x [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{x - \underline{x}} \quad \forall x \in \mathcal{X} \quad (105)$$

Which completes the proof. □



If it can be established that:

$$\left( \frac{\int_{\underline{x}}^{\bar{x}} [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{\bar{x} - x} \geq \frac{\int_{\underline{x}}^x [f(g^\theta(\tilde{x})) - f(g(\tilde{x}))] d\tilde{x}}{x - \underline{x}} \quad \forall x \in \mathcal{X} \right) \quad (106)$$

$$\implies g^\theta \succeq_{FOSD} g$$

then we will have established that function  $f$  satisfies quasi-MLRP.

We now use this to prove that  $f(x) = x$  satisfies quasi-MLRP. Plugging in to the derived integral condition, we obtain:

$$\frac{G(x) - G^\theta(x)}{\bar{x} - x} \geq \frac{G^\theta(x) - G(x)}{x - \underline{x}} \quad \forall x \in \mathcal{X} \quad (107)$$

Re-arranging this:

$$G(x) \geq G^\theta(x) \quad \forall x \in \mathcal{X} \quad (108)$$

which is the definition that  $g^\theta \succeq_{FOSD} g$ . This completes the proof and establishes that quasi-MLRP is a strict weakening of MLRP.  $\square$