Coordination and Continuous Stochastic Choice*

Stephen Morris and Ming Yang
Princeton University and Duke University

this revision: December 2018

Abstract

Players choose a stochastic choice rule, assigning a probability of "investing" as a function of a state. Players receive a return to investment that is increasing in the proportion of others who invest and the state. In addition, they incur a small cost associated with adapting the probability of investment to the state. If costs satisfy infeasible perfect discrimination (discontinuous stochastic choice rules are infinitely costly) and a weak translation insensitivity property, there is a unique equilibrium as costs become small, where play is Laplacian in the underlying complete information game (i.e., actions are a best response to a uniform conjecture over the proportion of others investing). The cost functional on stochastic choice rules may reflect the cost of endogenously acquiring information in order to implement a stochastic choice rule. With this interpretation, our results generalize global game selection results (Carlsson and van Damme (1993) and Morris and Shin (2003)); however, the mechanism works for general information structures and the source of uniqueness is different.

JEL: C72 D82

Keywords: Coordination, endogenous information acquisition, continuous stochastic choice, higher order beliefs

1 Introduction

We study a stochastic choice game where each player chooses a stochastic choice rule, giving a probability of "investing" as a function of the state of the world. Players receive a return to investment which is increasing in the proportion of other players investing and

*An earlier version of this paper was circulated under the title "Coordination and the Relative Cost of Distinguishing Nearby States." We are grateful for the comments of conference discussants Muhamet Yildiz (who discussed the paper at both the "Global Games in Ames" conference in April 2016 and the Cowles Economic Theory conference in June 2016) and Jakub Steiner (who discussed the paper at the AEA winter meetings in January 2018). Morris discussed experimental ideas related to this project with Mark Dean and Isabel Trevino, discussed in Section 7.4. We are also grateful for feedback from Ben Hebert, Filip Matejka, Chris Sims, Philipp Strack, Tomasz Strzalecki, Mike Woodford and many seminar participants. We received valuable research assistance from Ole Agersnap.
the payoff relevant state of the world. Players hence have incentives to coordinate their investment decisions. Their choice of stochastic choice rule incurs a small cost associated with adapting the probability of investment to the state. We study what happens when we fix a cost functional on stochastic choice rules and multiply the cost functional by a constant that becomes small. Two properties of the cost functional are key. Infeasible perfect discrimination requires that discontinuous stochastic choice rules are infinitely costly (and thus infeasible). Translation insensitivity requires that translating a stochastic choice rule has only a small impact on the cost. The main result of the paper is that if the cost functional satisfies these two properties, there is a unique equilibrium of the stochastic choice game in the small cost limit. In particular, the unique equilibrium of the stochastic choice game selects an outcome from the Nash equilibria of each complete information game corresponding to each state of the world. In the limit, the Laplacian equilibrium of the complete information game is played: players invest when invest is a best response to a uniform belief over the proportion of other players choosing to invest in the complete information game. Thus an arbitrarily small friction in players’ ability to fine tune their action to the state gives rise to a natural equilibrium selection with the selected outcome independent of the fine details of the cost functional.

There exist multiple motivations for the cost functional on stochastic choice rules. For example, one could assume that there are "control costs". But our leading motivation for assuming costly stochastic choice rules is informational. The cost functional on stochastic choice rules can be understood as arising from information acquisition.\footnote{We are grateful to an anonymous referee for the excellent suggestion of framing our results in the context of a stochastic choice game, allowing us to abstract from fine details of modelling information acquisition in our main result.} Consider the two stage game where each player first decides to acquire information about the state of the world, given some cost functional on information, and then chooses a strategy mapping signals to actions. The information and strategy together implement a stochastic choice rule and we can understand the cost of a stochastic choice rule to be the cheapest way of implementing it. With this interpretation, infeasible perfect discrimination captures the idea that it is impossible to perfectly discriminate between states that are arbitrarily close together.\footnote{We discuss some alternative foundations for this assumption as well as experimental evidence in Section 7.} Translation insensitivity captures the idea that the cost of information depends on how much discrimination between states occurs but not on where that discrimination is focussed.

With this information acquisition interpretation, our results cleanly embed leading results on global games as a special case. Suppose that each player could observe a signal equal to the true state plus some additive noise. A higher precision of the signal (i.e., lower variance of the noise) is more expensive. But it is infeasible to acquire a perfect signal. The induced cost functional on stochastic choice rules satisfies infeasible perfect discrimi-
nation (because perfect signals are infeasible) and translation insensitivity (because noise is additive) and our main result thus implies leading results about equilibrium selection in global games (we review the relevant literature below).

But the information structure assumed in global games - noisy signals with additive noise - is very special. In particular, it is inflexible. If players acquire information about one region of the state space, they are forced to acquire information about every other region of the state space. We say that a cost function on information is Blackwell-consistent if it has the property that there is always a strict cost saving to acquiring a feasible information structure that is strictly less informative in the sense of Blackwell (1953). A basic but important implication of Blackwell-consistency is that players would always choose to acquire only binary signals about the state of the world. The global game information cost function is not Blackwell-consistent. However, it is very natural to impose Blackwell-consistency as well as infeasible perfect discrimination and translation insensitivity (and we will give examples in the body of the paper).

Our main result is established for stochastic choice games, where it does not matter whether Blackwell-consistency is satisfied or not. However, in order to understand what is driving our main result with an information acquisition interpretation, we also study explicitly an information acquisition game. When we do so, we see a sharp difference between the source of uniqueness in global games and our more general uniqueness result. In global games, players will acquire very accurate signals when signals are very cheap. By doing so, they create an information structure where there is Laplacian uncertainty: a player observing a given signal must always have an approximately uniform belief over the proportion of other players who have observed a higher signal than him. This Laplacian uncertainty property ensures that there is always a unique equilibrium in the coordination game given the information chosen by the players in equilibrium. In particular, there would be a unique equilibrium if these information structures were given to them exogenously.\footnote{This is the intuition for Laplacian selection discussed in Morris and Shin (2003) and formalized in Morris, Shin, and Yildiz (2016).}

But consider the important and natural alternative case where the cost of information is Blackwell-consistent. In this case, players will only acquire information with binary signals. Because players only observe binary signals of the state, it will generally be the case that - given the information that players acquire in equilibrium - there would be multiple equilibria if this information had been given to them exogenously. Limit uniqueness thus arises despite the fact that there would have been multiple equilibria if players’ information had been given to them exogenously.

Our infeasible perfect discrimination property makes sense if the distance between states has a meaning and nearby states are relatively hard to distinguish. We present a converse result showing that if nearby states are relatively easy to distinguish - we say that there is cheap perfect discrimination - then multiplicity is restored. It has become common in the
recent literature to measure the cost of information by the reduction in entropy. Entropy is an information theoretic notion under which the distance between states has no meaning or significance. This implies that the entropy reduction cost functional satisfies our cheap perfect discrimination property and thus gives rise to multiple equilibria.

While infeasible perfect discrimination is a sufficient condition for limit uniqueness, we also provide a weaker sufficient condition. The key to our main result is that players choose continuous stochastic choice rules in equilibrium. Infeasible perfect discrimination directly imposes that they must do so. We also report a weaker condition - expensive perfect discrimination - which implies that players will choose continuous stochastic choice rules in equilibrium, even when it is feasible to choose discontinuous stochastic choice rules that perfectly discriminate between neighboring states at finite cost. We emphasize the stronger sufficient condition (IPD) because it has easy and natural interpretations and justifications.

We present our main results under the maintained assumption that players will choose monotonic stochastic choice rules. This restriction can be relaxed. We say that the cost functional on stochastic choice rules is variation-averse if stochastic choice rules giving rise to the same distribution over mixed strategies are always cheaper if they are monotonic and thus have less variation. This restriction is also consistent with the intuition that it is harder to distinguish nearby states. Under this restriction, one can show that if a player believes that other players will choose monotonic stochastic choice rules, that player will have a monotonic best response. Thus we have that there exists an equilibrium in monotonic stochastic choice rules, and there is uniqueness and Laplacian selection if we restrict attention to such strategies. We also show that under the strong restriction that the cost functional is submodular, we will have non-existence of non-monotonic equilibria.

1.1 Literature and Broader Implications

Carlsson and Damme (1993) introduced global games, where players exogenously observe the true payoffs with a small amount of additive noise, and showed that there is a unique equilibrium played in the limit as noise goes to zero. These results have since been significantly generalized and widely applied. Morris and Shin (2003) provide an early survey of theory and applications; the class of continuum player, binary action, symmetric payoff games studied in this paper is essentially that of Morris and Shin (2003), which embeds most applications of global games. Szkup and Trevino (2015) and Yang (2015) showed that global game uniqueness and selection results will go through essentially unchanged if players endogenously choose the precision of their private signals, as the cost of signals becomes small.

Yang (2015) emphasized that the global game information structure was inflexible - players were constrained to a very restricted parameterized class of information structures,

\footnote{The formal statement of this result appears in the working paper version, Yang (2013).}
with all other information structures being infeasible. Sims (2003) suggested that the ability to process information is a binding constraint, which implies - via results in information theory - that there is a bound on feasible entropy reduction. If information capacity can be bought, this suggests a cost functional that is an increasing function of entropy reduction. An attractive feature of entropy reduction treated as a cost of information is that it is flexible. But Yang (2015) showed that global game uniqueness and selection results are reversed if entropy reduction is used as a cost functional. One contribution of this paper is to reconcile these results. We show that flexible information acquisition is consistent with global game uniqueness results - the key is not the infeasibility of the global game information structure, but rather the natural implicit assumption of infeasible perfect discrimination.

Our paper has implications for the widespread use of entropy reduction in economic applications. Because of its purely information theoretic foundations, this cost function is not sensitive to the labelling of states, and thus it is built in that it is as easy to distinguish nearby states as distant states. Because entropy reduction has a tractable functional form for the cost of information, it has been widely used in economic settings where it does not reflect information processing costs and where the insensitivity to the distance between states does not make sense. While this may not be important in single person decision making, this paper contains a warning about use of entropy as a cost of information in strategic settings. We discuss recent work on related themes in Section 7.

Stochastic choice arises in a variety of contexts in the game theory literature. Stochastic choice arising from payoﬀ perturbations were introduced in Harsanyi (1973) and is used in important literatures on stochastic fictitious play (Fudenberg and Kreps (1993)) and quantal response equilibria (McKelvey and Palfrey (1995)). In these papers, there is upper hemi-continuity (limit points as stochasticity disappears correspond to Nash equilibria of the unperturbed game) and - under stronger assumptions - lower hemi-continuity (strict - or regular - equilibria of the limit game are limits of equilibria of the perturbed game). van Damme (1983) introduced control costs to reduce errors in the complete information game reﬁnement literature; this perturbation selected among irregular Nash equilibria of the complete information game. In the continuous action Nash bargaining game of Carlsson (1991), exogenous noise was added to a player’s choice of action and this selected among the continuum of equilibria. Distinctive conceptual features of our stochastic choice game are that there is uncertainty about a continuous state space and the cost of stochastic choice rules has a simple and natural interpretation as a reduced form for the information acquisition cost. A distinctive technical feature of our stochastic choice game is that we select a unique equilibrium from multiple strict equilibria in the limit game, thus paralleling and generalizing the strong failure of lower hemi-continuity that is known to arise in games of exogenous incomplete information (as in global games or, more generally, Monderer and Samet (1989) and Kajii and Morris (1997)).
Our results address a debate about equilibrium uniqueness without common knowledge. Weinstein and Yildiz (2007) have emphasized that equilibrium selection arguments in the global games literature rely on a particular relaxation of common knowledge (noisy signals of payoffs) and do not go through under other natural exogenous local perturbations from common knowledge. We show that endogenous information acquisition gives rise to uniqueness under some natural and interpretable assumptions about the cost functional.

We maintain the assumption that when players learn about the state, they observe conditionally independent signals. Denti (2018) and Hoshino (2018) consider the case where players may choose to learn about others’ actions as well. In both papers, the role of the distance between states in the cost of information is not modelled: Denti (2018) focuses on an entropy cost function (as well as discussing a class of posterior-separable generalizations, where the labelling of states also does not matter); Hoshino (2018) considers finite states. Denti (2018) nonetheless gets the risk dominant action played, while the richer cost functions in Hoshino (2018) allow any action to be selected. We discuss more detailed connections with these works in Section 7.

While our uniqueness result generalizes the global games literature initiated by Carlsson and Damme (1993), we cannot appeal to the arguments in Carlsson and Damme (1993) and later papers on binary action games because the relevant space of stochastic choice rules cannot be characterized by a threshold. Our results are closer to the argument for uniqueness in general supermodular games in Frankel, Morris, and Pauzner (2003). Here too, translation insensitivity has a crucial role, with contraction like properties giving rise to uniqueness (Mathevet (2008) showed an exact relation to the contraction mapping theorem under slightly stronger assumptions). Mathevet and Steiner (2013) highlighted the role of translation insensitivity in obtaining uniqueness results. All these papers assume noisy information structures and depend on built-in translation insensitivity without highlighting that continuous choice is also obtained for free in these environments. They studied how translation insensitivity helps pin down the equilibrium strategy while the role of continuous choice is not the focus. In contrast, we show that translation insensitivity leads to limit uniqueness (multiplicity) if continuous choice is satisfied (violated), and thus highlight continuous choice as the essential property that leads to the equilibrium uniqueness.

We proceed as follows. Section 2 sets up the model of a stochastic choice game. Section 3 contains our main result about stochastic choice games. Section 4 reports a converse and weaker sufficient conditions for Laplacian selection, in order to deepen our understanding of the main result. In Section 5, we describe explicitly how stochastic choice games can be understood as a reduced form for games with endogenous information acquisition and discuss how this implies a qualitatively distinct understanding of the origins of equilibrium uniqueness from the global games literature. In Section 6, we discuss alternative ways of relaxing the maintained monotonic strategy assumption. Section 7 contains a discussion of
our results.

2 Setting

We first describe a canonical class of games with strategic complementarities, parameterized by a payoff relevant state of the world. This class of games was introduced in the survey of global games of Morris and Shin (2003) and embeds most of the games used in the applied literature on global games. We then describe the stochastic choice game in which each player chooses a stochastic choice rule mapping the state space to the probability simplex of the actions.

2.1 The Complete Information Game

A continuum of players simultaneously choose an action, "not invest" or "invest". The mass of players is normalized to 1 and a generic player is indexed by \( i \in [0,1] \). A player’s return if she invests is \( \pi(l, \theta) \), where \( l \in [0,1] \) is the proportion of players investing and \( \theta \in \mathbb{R} \) is a payoff relevant state. The return to not investing is normalized to 0. Note that while we find it convenient to label actions "invest" and "not invest" to help comprehension, the return of 0 is just a normalization and, modulo this normalization, we are allowing for arbitrary continuum player, binary action, symmetric payoff games.

The following three substantive assumptions on the payoff function \( \pi(l, \theta) \) are the key properties of the game.

**Assumption A1 (Strategic Complementarities):** \( \pi(l, \theta) \) is non-decreasing in \( l \).

**Assumption A2 (State Monotonicity):** \( \pi(l, \theta) \) is non-decreasing in \( \theta \).

**Assumption A3 (Limit Dominance):** There exist \( \theta_{\min} \in \mathbb{R} \) and \( \theta_{\max} \in \mathbb{R} \) such that

\[
\begin{align*}
(i) \quad & \pi(l, \theta) < 0 \quad \text{for all } l \in [0,1] \text{ and } \theta < \theta_{\min}; \\
(ii) \quad & \pi(l, \theta) > 0 \quad \text{for all } l \in [0,1] \text{ and } \theta > \theta_{\max}.
\end{align*}
\]

Assumption A1 states that the incentive to invest is increasing in the proportion of other players who are also investing. Assumption A2 states that the incentive to invest is increasing in the state. Assumption A3 states that players have a dominant strategy to not invest or invest when the state is, respectively sufficiently low or sufficiently high.

We make a number of additional more technical assumptions imposing some strictness in monotonicity assumptions and some continuity.

**Assumption A4 (State Single Crossing):** For any \( l \in [0,1] \), there exists a \( \theta_l \in \mathbb{R} \) such that \( \pi(l, \theta) > 0 \) if \( \theta > \theta_l \) and \( \pi(l, \theta) < 0 \) if \( \theta < \theta_l \).

Given assumption A2, assumption A4 simply rules out the possibility that there is an open interval of \( \theta \) for which \( \pi(l, \theta) = 0 \). Notice that A2 and A4 imply limit dominance. Specifically, we can define \( \theta_{\min} \) and \( \theta_{\max} \) by setting \( \theta_{\min} = \theta_1 \) and \( \theta_{\max} = \theta_0 \) as defined.
in assumption A4. Then for any \( l \in [0, 1] \), \( \pi (l, \theta) \leq \pi (1, \theta) < 0 \) for all \( \theta < \theta_{\text{min}} \), and \( \pi (l, \theta) \geq \pi (0, \theta) > 0 \) for all \( \theta > \theta_{\text{max}} \), i.e., limit dominance holds.

We will be especially concerned about a player’s "Laplacian payoff" when he has a uniform, or "Laplacian", belief about the proportion of opponents who invest in state \( \theta \), or

\[
\Pi (\theta) = \int_{l=0}^{1} \pi (l, \theta) dl.
\]

We impose two assumptions on Laplacian payoffs:

**Assumption A5 (Laplacian Single Crossing):** There exists \( \theta^** \in \mathbb{R} \) such that \( \Pi (\theta) > 0 \) if \( \theta > \theta^** \) and \( \Pi (\theta) < 0 \) if \( \theta < \theta^** \).

We will refer to \( \theta^** \) as the Laplacian threshold. A player with the Laplacian belief who knows the state will invest if the state exceeds \( \theta^** \) and not invest if the state is less than \( \theta^** \).

**Assumption A6 (Laplacian Continuity):** \( \Pi \) is continuous, and \( \Pi^{-1} \) exists on an open neighborhood of \( \Pi (\theta^**) \).

Assumption A5 imposes the same single crossing property we imposed when players were certain about the proportion of opponents investing. Assumption A6 imposes continuity of Laplacian payoffs. Notice that A5 and A6 together imply that \( \theta^** \) uniquely solves \( \Pi (\theta) = 0 \).

Finally, we require:

**Assumption A7 (Bounded Payoffs):** \( |\pi (l, \theta)| \) is uniformly bounded.

Assumption A7 simplifies the proof but could be relaxed.

We will use the following game - widely used in the global games literature\(^5\) - to illustrate our results in the paper.

**Example 1 [Regime Change Game]** Each player has a cost \( t \in (0, 1) \) of investing and gets a gross return of 1 from investing if the proportion of players investing is at least \( 1 - \theta \). Thus

\[
\pi (l, \theta) = \begin{cases} 
1 - t, & \text{if } l \geq 1 - \theta \\
-t, & \text{otherwise}
\end{cases}
\]

This example satisfies assumptions A1 through A7, even though it fails stronger strict monotonicity and continuity properties. In particular, \( \pi (l, \theta) \) is not strictly increasing in \( \theta \) for each \( l \in [0, 1] \); but setting \( \theta_l = 1 - l \), we do have \( \pi (l, \theta) > 0 \) if \( \theta > \theta_l \) and \( \pi (l, \theta) < 0 \) if \( \theta < \theta_l \), and thus we do have the weaker single crossing condition A4. Also, \( \pi (l, \theta) \) is not

continuous in \( \theta \) for each \( l \in [0, 1] \); but the Laplacian payoff is

\[
\Pi (\theta) = \int_0^1 \pi (l, \theta) \, dl = \begin{cases} 
1 - t, & \text{if } \theta \geq 1 \\
\theta - t, & \text{if } 0 \leq \theta \leq 1 \\
-t, & \text{if } \theta \leq 0 
\end{cases}
\]

so Laplacian continuity is easily verified. Finally, observe that the Laplacian threshold solving \( \Pi (\theta) = 0 \) is \( \theta^{**} = t \).

2.2 The Stochastic Choice Game

We now define a stochastic choice game to model a situation where there are costs associated with adapting action choices to the state. Players share a common prior on \( \theta \), denoted by density \( g \), which is continuous and strictly positive on \([\theta_{\min}, \theta_{\max}]\). We also assume that the common prior assigns positive probability to both dominance regions.

A generic player \( i \) chooses a stochastic choice rule \( s_i : \mathbb{R} \to [0, 1] \), with \( s_i (\theta) \) being the player’s probability of investing conditional on the state being \( \theta \). We write \( S \) for the set of all stochastic choice rules, which consists of all Lebesgue measurable functions that map from the real line to \([0, 1]\). Players privately and simultaneously choose their stochastic choice rules so that their actions are independent conditional on the state. As is usual, we adopt the law of large numbers convention that given a strategy profile \( \{s_j\}_{j \in [0,1]} \), the proportion of players that invest when the state is \( \theta \) is \( \int s_j (\theta) \, dj \).

The first component of a player’s payoff will then be her expected return given by

\[
u (s, \{s_j\}_{j \in [0,1]}) = \int s_i (\theta) \pi (\int s_j (\theta) \, dj, \theta) g (\theta) \, d\theta ,
\]

which is the expectation of the complete information game payoffs.

The second component of a player’s payoff depends only on his own stochastic choice rule. This component reflects the cost of adapting action choices to the state. A cost functional \( c : S \to \mathbb{R}_+ \cup \{\infty\} \) maps stochastic choice rules to the extended positive real line. Here \( c (s) = \infty \) just means that \( s \) is not feasible. We equip the strategy space with the \( L^1 \)-metric, so that the distance between stochastic choice rules \( s_1 \) and \( s_2 \) is given by

\[
\|s_1, s_2\| = \int_\mathbb{R} |s_1 (\theta) - s_2 (\theta)| g (\theta) \, d\theta ;
\]

and write \( B_\delta (s) \) for the open set of stochastic choice rules within distance \( \delta \) of \( s \) under this metric. A player incurs cost \( \lambda_c (s) \) if she chooses \( s \in S \). We will hold the cost functional \( c \)

---

6The law of large numbers is not well defined for a continuum of random variables (Sun (2006)). Our convention is equivalent to assuming that opponents’ play is the limit of play of finite selections from the population.

7All \( L^p \)-metrics are equivalent in the Hilbert space of stochastic choice rules. Here we choose the \( L^1 \)-
fixed in our analysis and vary $\lambda \geq 0$, a parameter that represents the difficulty of controlling the state contingency of actions; we will refer to the resulting stochastic choice game as the $\lambda$-game. The payoff of player $i$ in the $\lambda$-game are thus given by

$$u \left( s_i, \{ s_j \}_{j \in [0,1]} \right) - \lambda \cdot c(s_i).$$

When $\lambda = 0$, the players can choose actions fully contingent on $\theta$ at no cost and the stochastic choice game reduces to a continuum of complete information games parameterized by $\theta$. We will perturb these complete information games by letting $\lambda$ be strictly positive but close to zero. Focussing on small but positive $\lambda$ sharpens the statement and intuition of our results.

We restrict attention to monotonic (non-decreasing) stochastic choice rules in the body of the paper. This is consistent with many applications (e.g., the stochastic choice rule is always monotone in global game models) and allows us to highlight key insights. Section 6 provides conditions under which this restriction does not weaken the results given our research purpose. We write $S_M$ for the set of monotonic stochastic choice rules, which is the strategy space for all players.

Thus we have the following definition of equilibrium.

**Definition 2 (Nash Equilibrium)** A strategy profile $\{ s_j \}_{j \in [0,1]}$ is a Nash equilibrium of the $\lambda$-game if

$$s_i \in \arg \max_{s \in S_M} \left[ u \left( s, \{ s_j \}_{j \in [0,1]} \right) - \lambda \cdot c(s) \right]$$

for each $i \in [0,1]$.

Because the strategy space $S_M$ is compact according to Helly’s selection theorem, the best responses to any profile $\{ s_j \}_{j \in [0,1]}$ exist and so do the equilibria.

We have one important maintained assumption on the cost functional. An important stochastic choice rule will be the (discontinuous) step function $1_{\{ \theta \geq \psi \}}$, where a player invests if and only if the state exceeds a threshold. We will assume that it is possible to approximate this step function arbitrarily closely (perhaps by a continuous stochastic choice rule) at some finite cost.\(^8\)

**Assumption A8:** For any $\psi \in \mathbb{R}$ and $\delta > 0$, there exists a stochastic choice rule $s \in S_M$ such that $s \in B_\delta \left( 1_{\{ \theta \geq \psi \}} \right)$ and $c(s) < \infty$.

We will use the following example of a cost functional to illustrate our results. It also illustrate Assumption A8.

\(^8\)The step function is the ideal stochastic choice rule for a decision problem when there is complete information. If it were infeasible to approximate step functions, the limit would not approach complete information.
Example 3 [Max Slope Cost Functional] The cost \( c \) of a stochastic choice rule \( s \in S_M \) is \( f(k) \), where \( k \) is the maximum slope of \( s \) and \( f : \mathbb{R}_+ \cup \{ \infty \} \to \mathbb{R}_+ \cup \{ \infty \} \) is weakly increasing with \( f(0) = 0 \) and \( f(k) < \infty \) for all \( k \in \mathbb{R}_+ \). Thus

\[
c(s) = f \left( \sup_{\theta} s'(\theta) \right).
\]  

If \( s \) is discontinuous at \( \theta \), then \( s'(\theta) \) is understood to be infinity, and the cost \( c(s) = f(\infty) \). Discontinuous stochastic choice rules are infeasible if \( f(\infty) = \infty \).

An intuition for this cost functional is that it is costly to adapt the stochastic choice rule to the state and the derivative is a measure of how finely it is adapted. In Section 5, we show that this cost functional is closely related to the global game information structure with uniform noise. In a very different class of evolutionary models, Robson (2001), Rayo and Becker (2007) and Netzer (2009), slope is used as a measure of attention in an analogous way. In Section 7.1, we very briefly discuss natural generalizations of this cost functional.

To see that the max slope cost functional satisfies Assumption A8, note that we can always approximate \( 1_{\{s \geq \psi\}} \) by

\[
s(\theta) = \begin{cases} 
0, & \text{if } \theta \leq \psi - \frac{1}{2k} \\
\frac{1}{2} + k(\theta - \psi), & \text{if } \psi - \frac{1}{2k} \leq \theta \leq \psi + \frac{1}{2k} \\
1, & \text{if } \theta \geq \psi + \frac{1}{2k}
\end{cases}
\]

with \( k \in \mathbb{R}_+ \) large enough, no matter whether \( f(\infty) = \infty \).

3 Main Result

Our main result establishes that in the low cost limit (i.e., when \( \lambda \to 0 \)), players choose the Laplacian action, i.e., a best response to the Laplacian conjecture that the proportion of others investing is uniformly distributed between 0 and 1. Thus they invest when the state exceeds the Laplacian threshold \( \theta^* \), as defined in Assumption A5, and they do not invest when the state is less than \( \theta^* \). Thus for small \( \lambda \), equilibrium stochastic choice rules are well approximated by the step function \( 1_{\{s \geq \psi\}} \). The following definition gives the relevant formal statement of this approximation.

**Definition 4 (Laplacian Selection)** Laplacian selection occurs if, for any \( \delta > 0 \), there exists \( \lambda > 0 \) such that \( \|s, 1_{\{s \geq \psi\}}\| \leq \delta \) whenever \( s \) is an equilibrium strategy in the \( \lambda \)-game and \( \lambda \leq \lambda \).

\(^9\)If \( s(\theta) \) is continuous but not differentiable at \( \theta \), we can take it to equal the maximum of the left and right derivatives.

\(^{10}\)We are grateful to a referee for pointing out this connection.
We have two key sufficient conditions for Laplacian selection. First:

**Definition 5 (IPD)** Cost functional \( c(\cdot) \) satisfies infeasible perfect discrimination if it assigns infinite cost to all \( s \in S_M \) that are not absolutely continuous.

If the cost functional satisfies IPD, all stochastic choice rules with jumps are infinitely costly so that they are not optimal as long as \( \lambda > 0 \). For example, the max slope cost functional satisfies IPD if and only if \( f(\infty) = \infty \).

IPD is our key assumption and is a weak and natural assumption under most interpretations of the cost functional on stochastic choice rules. If the cost represents control costs, then the assumption says that perfect control is infeasible. Under our leading informational interpretation of the cost (discussed in depth in Section 5), the assumption maintains that it is infeasible to perfectly discriminate nearby states. Heinemann, Nagel, and Ockenfels (2004) provide a form of direct evidence in favor of the assumption: as part of an experimental analysis of global games, they consider a complete information treatment and show that population behavior takes the form of a stochastic choice rule which "looks" continuous.

Our second condition concerns how costs vary as we translate the stochastic choice rule. Let \( T_\Delta : S_M \to S_M \) be a translation operator: that is, for any \( \Delta \in \mathbb{R} \) and \( s \in S_M \),

\[
(T_\Delta s)(\theta) = s(\theta + \Delta).
\]

**Definition 6 (Translation Insensitivity)** Cost functional \( c(\cdot) \) satisfies translation insensitivity if there exists \( K > 0 \) such that, for all \( s \),

\[
|c(T_\Delta s) - c(s)| < K \cdot |\Delta|.
\]

This property requires that the cost responds at most linearly to translations of the stochastic choice rules. In the context of information acquisition, translation insensitivity captures the idea that the cost of information acquisition reflects the cost of paying attention to some neighborhood of the state space, but is not too sensitive to where attention is paid. It is straightforward to see that the max slope cost functional satisfies translation insensitivity, as the maximal slope of the stochastic choice rule remains the same after translation. Now we have our main result:

**Proposition 7 (IPD and Laplacian Selection)** If the cost functional satisfies infeasible perfect discrimination and translation insensitivity, then there is Laplacian selection.

Thus when \( c(\cdot) \) satisfies infeasible perfect discrimination and translation insensitivity, and the cost multiplier \( \lambda \) is small, all equilibria are close to the Laplacian switching strategy.\(^{11}\)

---

\(^{11}\)The proposition would continue to hold if translation insensitivity was relaxed to a local version, where it holds only in a small neighborhood of the step functions. In particular, it is enough to have that for any \( \psi \in [\theta_{\min}, \theta_{\max}] \), there exist \( \delta > 0 \) and \( K > 0 \) such that

\[
|c(T_\Delta \tilde{s}) - c(\tilde{s})| < K \cdot |\Delta| \quad \text{as long as } \tilde{s}, T_\Delta \tilde{s} \in B_\delta \{1_{\{\psi \geq \psi\}}\}.
\]
A first step in the proof involves establishing implications of infeasible perfect discrimination in a decision problem. In particular, in a monotonic decision problem that will arise in solving for equilibrium, optimal strategies must approach a step function. Let

\[ V_\lambda (\tilde{s}|s) = \int \pi (s(\theta), \theta) \cdot \tilde{s}(\theta) g(\theta) d\theta - \lambda \cdot c(\tilde{s}) \]

denote a player’s expected payoff from playing stochastic choice rule \( \tilde{s} \) if all other players choose strategy \( s \). Consider the player’s decision problem

\[ \max_{\tilde{s} \in \mathcal{S}_\lambda} V_\lambda (\tilde{s}|s) . \] (3)

We call this problem the \((s, \lambda)\)-decision problem and write \( \mathcal{S}_\lambda (s) \) for the set of optimal stochastic choice rules, i.e.,

\[ \mathcal{S}_\lambda (s) = \arg \max_{\tilde{s} \in \mathcal{S}_\lambda} V_\lambda (\tilde{s}|s) . \]

Since \( s \) is nondecreasing in \( \theta \), Assumptions A1 and A4 imply that there exists a threshold \( \theta_s \in \mathbb{R} \) such that \( \pi (s(\theta), \theta) > 0 \) if \( \theta > \theta_s \) and \( \pi (s(\theta), \theta) < 0 \) if \( \theta < \theta_s \). We will show that it is optimal for players to choose strategies that are close to a step function jumping at \( \theta_s \) when the cost of information is small.

**Lemma 8 (Optimal Strategies in the Decision Problems)** The essentially unique optimal stochastic choice rule if \( \lambda = 0 \) is a step function at \( \theta_s \), i.e.,

\[ \mathcal{S}_0 (s) = \{ 1_{\{ \theta \geq \theta_s \}} \} . \]

Moreover, for any \( \rho > 0 \), there exists a \( \bar{\lambda} > 0 \) such that \( \mathcal{S}_\lambda (s) \subset B_\rho (1_{\{ \theta \geq \theta_s \}}) \) for all \( s \in \mathcal{S}_\lambda \) and \( \lambda < \bar{\lambda} \).

The fact that the decision maker’s optimal stochastic choice rules approximate \( 1_{\{ \theta \geq \theta_s \}} \) as \( \lambda \to 0 \) reflects her motive to sharply identify event \( \{ \theta \geq \theta_s \} \) from its complement. In a decision problem, whether this is achieved by a continuous or discontinuous \( s_\lambda \in \mathcal{S}_\lambda (s) \) is not important, since the loss caused by deviating from \( 1_{\{ \theta \geq \theta_s \}} \) is of the order of magnitude of \( \| 1_{\{ \theta \geq \theta_s \}}, s_\lambda \| \) in either way. In contrast, in the game considered here, we will see that the continuity of \( s_\lambda \) is crucial in determining the equilibrium outcomes.

We report the remainder of the proof in the Appendix. To give some intuition for the result, and to sketch the idea of the proof, we consider the case of regime change payoffs (Example 1) and the max slope cost functional (Example 3).

---

12 Equivalently, \( s(\theta) \) can be interpreted as the aggregate stochastic choice rule, which is the proportion of the players that invest when the state is \( \theta \).
3.1 Sketch of Proof for Regime Change Game and Max Slope Cost Function

Let $s$ be a symmetric equilibrium of the game with regime change payoffs and the max slope cost functional. Let $\psi$ denote the threshold above which the regime changes (i.e., $l \geq 1 - \theta$) in this equilibrium. When $f(\infty) = \infty$, the max slope cost functional satisfies IPD and $s$ is continuous. Assuming a continuum law of large numbers, $s(\theta)$ is also the proportion of the players that invest when the state is $\theta$. Hence the threshold $\psi$ is implicitly defined by the following equation

$$\psi = 1 - s(\psi),$$

resulting in a payoff gain $1_{(\theta \geq \psi)} - t$ to the players. Then under the max slope cost functional, the optimal stochastic choice rules always take the form

$$s_{\xi,k}(\theta) = \begin{cases} 
0, & \text{if } \theta \leq \xi - \frac{1}{2k} \\
\frac{1}{2} + k(\theta - \xi), & \text{if } \xi - \frac{1}{2k} \leq \theta \leq \xi + \frac{1}{2k} \\
1, & \text{if } \theta \geq \xi + \frac{1}{2k} 
\end{cases}.$$

(4)

We illustrate a typical payoff gain function $1_{(\theta \geq \psi)} - t$ and stochastic choice rule $s_{\xi,k}$ in this form in Figure 3.1 below.

![Figure 3.1: optimal stochastic choice rules](image-url)

Thus the equilibrium stochastic choice rule will take values 0 or 1 except in a region centered around a "cutoff" $\xi$ where it will be linearly increasing with slope $k$. As an implication of Lemma 8, when $\lambda$ is close to zero, $\xi$ is sufficiently close to $\psi$ and $k$ is sufficiently large.
Note that as a best response, the equilibrium \( s \) itself takes the above form \( s_{\xi,k} \) for some \( \xi \) and \( k \). Now suppose a player considers deviating by slightly translating the equilibrium stochastic choice rule \( s \) to \( T_\Delta s \). He should not benefit from this deviation. Since the max slope cost functional is invariant to translation, this requires that the expected return should not improve after translation, i.e.,

\[
\int \left[ 1_{\{\theta \geq \psi\}} - t \right] \left[ s(\theta + \Delta) - s(\theta) \right] g(\theta) \, d\theta \leq 0 .
\]

Taking derivative with respect to \( \Delta \) at \( \Delta = 0 \), the above inequality implies

\[
\int \left[ 1_{\{\theta \geq \psi\}} - t \right] g(\theta) s'(\theta) \, d\theta = 0 . \tag{5}
\]

By Lemma 8, when \( \lambda \) is close to zero, \( s(\theta) \) varies continuously from 0 to 1 within a small neighborhood \([\psi - \eta, \psi + \eta]\) of \( \psi \), where \( \eta \) is of the order of \( O(\lambda) \). Since \( g \) is continuous, \( g(\theta) = g(\psi) + O(\lambda) \) in this small neighborhood.\(^{13}\) Hence, (5) becomes

\[
0 = \int_{\psi - \eta}^{\psi + \eta} \left[ 1_{\{\theta \geq \psi\}} - t \right] g(\theta) s'(\theta) \, d\theta
\]

\[
= g(\psi) \int_{\psi - \eta}^{\psi + \eta} \left[ 1_{\{s(\theta) \geq 1 - \psi\}} - t \right] s'(\theta) \, d\theta + O(\lambda) ,
\]

where the second equality follows the fact that in equilibrium \( \{\theta \geq \psi\} = \{s(\theta) \geq 1 - \psi\} \).

Since \( g \) is continuous and strictly positive over \([\theta_{\text{min}}, \theta_{\text{max}}]\), \( g(\psi) \) has a strictly positive lower bound for \( \psi \in [\theta_{\text{min}}, \theta_{\text{max}}] \). Hence, we obtain

\[
\int_{\psi - \eta}^{\psi + \eta} \left[ 1_{\{s(\theta) \geq 1 - \psi\}} - t \right] s'(\theta) \, d\theta = O(\lambda) .
\]

Since \( s(\cdot) \) is absolutely continuous, \( s'(\cdot) \) exists and we can change the variable of integration from \( \theta \) to \( s \), resulting in

\[
\int_{0}^{1} \left[ 1_{\{s \geq 1 - \psi\}} - t \right] ds = O(\lambda) . \tag{6}
\]

Simple calculus shows that the left hand side of this Laplacian equation equals \( \psi - t \). Hence, as \( \lambda \to 0 \), \( \psi \) converges to \( t \), which is the Laplacian threshold \( \theta^{**} \) of the regime change game. Therefore, as an immediate implication of Lemma 8, any equilibrium stochastic choice rule converges to the Laplacian selection \( 1_{\{\theta \geq \theta^{**}\}} \) as \( \lambda \to 0 \).

\(^{13}\)Note that \( \psi \in [\theta_{\text{min}}, \theta_{\text{max}}] \) by definition, and \( g \) is continuous over \([\theta_{\text{min}}, \theta_{\text{max}}]\) and thus uniformly continuous over this closed interval. Then \( g(\theta) = g(\psi) + O(\lambda) \) holds no matter where \( \psi \) is in \([\theta_{\text{min}}, \theta_{\text{max}}]\).
3.2 Why Laplacian Selection?

The above sketch of the proof gives an accurate idea of how the proof works in the general case. However, it is an argument by contradiction that fails to give much intuition for why the limit equilibrium corresponds to Laplacian play. Recall that the Laplacian threshold $\theta^{**} = t$ is determined by the Laplacian equation (6), which requires that a player with the Laplacian belief, the uniform belief about the proportion of others who invest, is indifferent between investing and not investing. It is worth providing an intuition to see why the Laplacian belief arises in determining the limit unique equilibrium. In contrast to the usual global game intuition which relies on the interim beliefs derived from the signal structures extra to the game, the intuition here will be solely based on the structure of the stochastic choice game.

A player should not benefit from slightly deviating from the equilibrium stochastic choice rule $s$ to its translation $T_\Delta s$. Since translation insensitivity holds, the translation does not have a significant impact on the cost when $\Delta$ is small. We will then derive the intuition from the impact of the translation on the expected return.

Suppose the translation $\Delta > 0$. Then $s$ and $T_\Delta s$ result in different outcomes only when the player does not invest under $s$ but invest under $T_\Delta s$. This event is indicated as the yellow-shaded region in Figure 3.2. When $\Delta$ is sufficiently close to zero, $s$ and $T_\Delta s$ vary continuously from 0 to 1 within a sufficiently small interval $[\psi - \eta, \psi + \eta]$ in which the density $g(\theta)$ is approximately equal to $g(\psi)$. Then the probability that the mass of players investing falls into $[l, l + dl]$ when this event occurs is approximately equal to $g(\psi) \cdot \Delta \cdot dl$. Note that this probability does not depend on $l$ (up to the approximation).

Hence the distribution of $l$ in the event that $s$ and $T_\Delta s$ induce different outcomes is a uniform one, resulting in a marginal impact on the expected return approximately equal to $\int_{0}^{1} \pi(l, \psi) \, dl$. Finally, the equilibrium condition requires the player being (approximately) indifferent between $s$ and $T_\Delta s$, so that $\psi$ approximates the Laplacian threshold $\theta^{**}$ defined by $\int_{0}^{1} \pi(l, \theta^{**}) \, dl = 0$.

![Figure 3.2: translation leads to Laplacian belief](image-url)
It is worth noting that the Laplacian belief, which is uniform over all $l \in [0, 1]$, reflects a player’s strategic uncertainty about others’ decisions. This strategic uncertainty stems from the continuity of the equilibrium strategy $s$, in the sense that all $l \in [0, 1]$ are possible. In particular, all $l \in [0, 1]$ take place with (approximately) equal probability in a small neighborhood of the Laplacian threshold $\theta^{**}$ and there is no sharp distinction between any pair of states close to $\theta^{**}$. To see the importance of continuity, consider an example $s = 1_{\{y \geq \psi\}}$, which jumps from 0 to 1 at some threshold $\psi$. Then when comparing $s$ and $T_\Delta s$, in the event that they induce different outcomes the player is pretty sure that $l$ does not belong to $(0, 1)$ and takes distinct values at any pair of states on different sides of $\psi$ no matter how close they are. This lack of strategic uncertainty results in the indeterminacy of the threshold $\psi$. We will further illustrate this intuition in the next section.

4 Tightening Results and Continuous Stochastic Choice

Our main result gave natural and interpretable sufficient conditions for Laplacian selection. In this section, in order to deepen our understanding of the main result, we first identify general conditions under which Laplacian selection fails and there are multiple equilibria. We then identify weaker sufficient conditions for Laplacian selection. While our characterizations are not tight, it turns out that our general results pin down in general when there is limit uniqueness in our leading max slope cost functional example. Our results highlight that the key to Laplacian selection is whether the stochastic choice rules played in equilibrium when $\lambda > 0$ are continuous or not. Our natural infeasible perfect discrimination condition simply assumed that discontinuous stochastic choice rules were infeasible. The analysis of this section establishes that even when all stochastic choice rules are feasible, Laplacian selection still occurs if and only if continuous stochastic choice rules are sufficiently cheaper to be played in equilibrium.

4.1 A Converse

In order to appreciate the importance of the conditions for limit uniqueness and Laplacian selection, this subsection contrasts these conditions with a sufficient condition for limit multiplicity, so that multiple equilibria exist in the limit.

Yang (2015) showed that there is limit multiplicity if the cost of information is given by entropy reduction, which has been widely used as a cost of information following the work of Sims (2003). In particular, for any stochastic choice rule $s$, the associated entropy

14Sims (2003) proposed using entropy reduction to model information processing constraints, but Yang (2015) and many others have used it as a cost of information.
reduction is
\[ c(s) = E[h(s(\theta))] - h(E(s(\theta))] \]
where \( h : [0, 1] \to \mathbb{R} \) is given by
\[ h(x) = x \ln x + (1 - x) \ln (1 - x) \].

A key feature of entropy reduction is that it is information theoretic and therefore is independent of the labelling of states. Thus it is no harder to distinguish nearby states than to distinguish far away states. Yang (2015) reported another class of cost functionals where there is limit multiplicity. Say that the cost functional is \( \text{Lipschitz} \) if there exists a \( K > 0 \) such that
\[ |c(s_1) - c(s_2)| < K \cdot \|s_1, s_2\| \]
for all \( s_1, s_2 \in S \).

This condition also builds in the feature that it is easy to distinguish nearby states because any change in the stochastic choice rule at a small set of states results in a cost change of the same order, even if discontinuities are introduced. The entropy reduction cost is not Lipschitz. This is because \( \lim_{x \to 1} h_0(x) = 1 \) and \( \lim_{x \to 0} h_0(x) = 1 \), so that the marginal cost of letting \( s(\theta) \) approach 1 or 0 may tend to infinity. We will report a sufficient condition for multiplicity that covers both Lipschitz and entropy reduction.

We first introduce an operation on any stochastic choice rule that makes it sharper in discriminating the threshold events. In particular, for any \( \psi \in (\theta_{\min}, \theta_{\max}) \) and \( \varepsilon \in (0, 1/2) \), define an operator \( L_{\psi}^\varepsilon : S_M \to S_M \) such that
\[
(L_{\psi}^\varepsilon s)(\theta) = \begin{cases} 
\max(1 - \varepsilon, s(\theta)) & \text{if } \theta \geq \psi \\
\min(\varepsilon, s(\theta)) & \text{if } \theta < \psi.
\end{cases}
\]

Note that \( L_{\psi}^\varepsilon s \) does a weakly better job than \( s \) in discriminating event \( \{ \theta \geq \psi \} \) from its complement, since \( (L_{\psi}^\varepsilon s)(\theta) \geq s(\theta) \) on \( \{ \theta \geq \psi \} \) and \( (L_{\psi}^\varepsilon s)(\theta) \leq s(\theta) \) on \( \{ \theta < \psi \} \). It does a strictly better job if \( s \) is continuous, in the sense that for any pair of states \( \theta_1 \) and \( \theta_2 \) on different sides of \( \psi \), \( s(\theta_1) \) and \( s(\theta_2) \) converge to each other when \( \theta_1 \) and \( \theta_2 \) get closer, while \( L_{\psi}^\varepsilon s \) jumps in a magnitude of at least \( 1 - 2\varepsilon > 0 \) between \( \theta_1 \) and \( \theta_2 \) no matter how close they are. Whether this operation is preferable depends on the cost-benefit analysis, as captured by the following condition.

**Definition 9 (CPD)** The cost functional satisfies cheap perfect discrimination if for any \( \psi \in \mathbb{R} \) and \( \varepsilon \in (0, 1/2) \), there exists a \( \rho > 0 \) and \( K > 0 \) such that
\[
|c(L_{\psi}^\varepsilon s) - c(s)| \leq K \cdot \|L_{\psi}^\varepsilon s, s\|
\]
for all \( s \in B_p(1_{\theta \geq \psi}) \).

Note that if \( s(\theta) \geq 1 - \varepsilon \) for \( \theta \geq \psi \) and \( s(\theta) \leq \varepsilon \) for \( \theta < \psi \), \( L_\psi^s = s \) and the inequality automatically holds. Otherwise, the cost change caused by the operation is \( \lambda \frac{c \left( L_\psi^s \right) - c(s)}{c(s)} \) and the incremental expected return is of the order \( \left\| L_\psi^s, s \right\| \). The CPD condition requires that the cost responds at most linearly to the operation in a neighborhood of \( 1_{\theta \geq \psi} \). In other words, once CPD holds, it is inexpensive to sharply discriminate nearby states. Since CPD only requires Lipschitz property to hold for a special operation within a small neighborhood of the step function, it is implied by the Lipschitz property. It is straightforward to verify that the entropy reduction cost functional satisfies CPD, but the proof is tedious. We relegate the formal result and the proof to Lemma 18 in the appendix.

Proposition 10 If cost functional \( c(\cdot) \) satisfies cheap perfect discrimination, then for any threshold \( \lambda(\cdot) \in (\theta_{\min}, \theta_{\max}) \) and \( \varepsilon \in (0, 1/2) \), there exists \( \lambda > 0 \) such that whenever \( \lambda \in [0, \lambda] \), there is an equilibrium strategy profile \( \left\{ s^*_i, \lambda \right\} \) in which any \( s^*_i, \lambda \) satisfies

\[
\begin{align*}
\left\{ \begin{array}{ll}
1 - \varepsilon & \text{if } \theta \geq \theta^* \\
\varepsilon & \text{if } \theta < \theta^*
\end{array} \right. 
\end{align*}
\]

The proposition states that if CPD holds, for any threshold \( \theta^* \in (\theta_{\min}, \theta_{\max}) \), as \( \lambda \) vanishes, there is a sequence of equilibria uniformly converging to \( 1_{\theta \geq \theta^*} \), which is the ideal stochastic choice rule that perfectly discriminates event \{ \theta \geq \theta^* \} from its complement. Hence, the \( \lambda \)-game has infinitely many equilibria when \( \lambda \) is sufficiently small, a multiplicity result in sharp contrast to the limit unique equilibrium obtained in Proposition 7.

The key to understand this difference is again the (dis)continuity of the strategies. When CPD holds, the incremental cost of choosing a discontinuous stochastic choice rule \( L_\psi^s \) over a continuous rule \( s \) is at most proportional to the incremental expected return and thus is negligible at small \( \lambda \), making \( L_\psi^s \) a better choice rule than \( s \). Knowing that every other’s and hence the aggregate stochastic choice rule jumps at \( \theta^* \) radically reduces the strategic uncertainty a player faces, especially when \( \varepsilon \) is small (so that the jump is large). In particular, now a player is pretty sure that \( l \), the fraction of others investing, exceeds \( 1 - \varepsilon \) for states above threshold \( \theta^* \) and otherwise falls below \( \varepsilon \), making his payoff-gain cross zero at exactly \( \theta^* \). Since CPD holds, he then also prefers such a stochastic choice rule that jumps at \( \theta^* \) from below \( \varepsilon \) to above \( 1 - \varepsilon \), confirming the equilibrium. This logic applies to all thresholds within \( (\theta_{\min}, \theta_{\max}) \) and results in multiple equilibria.

To see a concrete example, consider symmetric equilibria of the regime change game with the max slope cost functional. For a threshold \( \psi \) and a stochastic choice rule \( s \) with maximal slope \( k \), the cost increment is \( c \left( L_\psi^s \right) - c(s) = f(\psi) - f(k) \) and \( \left\| L_\psi^s, s \right\| \) is of the order \( k^{-1} \). Then CPD holds if \( f(\psi) < \infty \) and \( \lim_{k \to \infty} [f(\psi) - f(k)] \cdot k < \infty \). Suppose the regime changes at threshold \( \theta^* \in (0, 1) \). Under the max slope cost functional, a player’s
optimal stochastic choice rule \(s^*\) always takes the form of (4). When \(\lambda\) is sufficiently small, CPD further implies \(k = \infty\) and \(s^* = 1_{\{\theta \geq \theta^*\}}\). Then the aggregate stochastic choice rule is also \(s^* = 1_{\{\theta \geq \theta^*\}}\) and does result in regime change at \(\theta^*\), confirming the equilibrium.

The max slope cost functional is special in the sense that the equilibrium stochastic choice rules in its CPD case actually achieve the step function \(1_{\{\theta \geq \theta^*\}}\) when \(\lambda\) is sufficiently small rather than "really" uniformly converging to it. The entropy reduction cost functional is a different special case of CPD, where the equilibrium stochastic choice rules never achieve the step function but uniformly converge to it. We can generalize the entropy reduction cost by allowing \(h\) to be any convex function on \([0, 1]\) and CPD still holds, as proved in Lemma 18 in the Appendix. Entropy reduction is special because \(\lim_{x \to 1} h'(x) = \infty\) and \(\lim_{x \to 0} h'(x) = -\infty\) so that the equilibrium stochastic choice rules never achieve the step function. If the two limits exist, the equilibrium stochastic choice rules achieve the step function for \(\lambda\) sufficiently small.

### 4.2 Continuous Stochastic Choice and a Strengthening of the Main Result

The comparison between the IPD and CPD cases shows that it is the continuity of the stochastic choice rules plays a key role in determining the equilibrium. Infeasible perfect discrimination is a natural property that immediately implies that continuous stochastic choice rules will be chosen in equilibrium. However, it is enough that continuous stochastic choice rules be chosen in equilibrium even if discontinuous stochastic choice rules are feasible. In this section, we report a condition - expensive perfect discrimination (EPD) - which is weaker than IPD but also sufficient for Laplacian selection. This result thus helps close the significant gap between IPD and CPD.

**Definition 11 (EPD)** Cost functional \(c(\cdot)\) satisfies expensive perfect discrimination, if for any stochastic choice rule \(s_1 \in S_M\) that is not absolutely continuous, any \(K > 0\), and any \(\delta > 0\), there exists an absolutely continuous \(s_2 \in B_\delta(s_1)\) such that \(c(s_1) - c(s_2) > K'\|s_1, s_2\|\).

Instead of precluding discontinuous stochastic choice rules by assigning infinite costs, EPD requires that it is cheap to approximate such choice rules with absolutely continuous ones relative to the degree of approximation. To see the intuition, note that in the example of max slope cost functional, EPD is equivalent to \(\lim_{k \to \infty} [f(\infty) - f(k)] \cdot k = \infty\). This could be true either because \(f(\infty) = \infty\), so IPD holds, or if \(f(\infty) < \infty\), so all stochastic choice rules have finite cost, and IPD fails, but CPD also fails. To see why this condition is important, consider approximating \(s_1 = 1_{\{\theta \geq \xi\}}\) with \(s_2 = s_{\xi,k}\) for some \(\xi \in \mathbb{R}\) and \(k\) large (where \(s_{\xi,k}\) was defined in equation (4)). Assuming a uniform common prior over \([\underline{\theta}, \overline{\theta}]\),
simple calculation shows
\[
\frac{c(s_1) - c(s_2)}{\|s_1, s_2\|} = \frac{f(\infty) - f(k)}{[4(\theta - \theta') k]^{-1}},
\]
which is unbounded since \([f(\infty) - f(k)] \cdot k\) is unbounded. In this example, \(c(s_1) = f(\infty) < \infty\) so that \(s_1\) is feasible, but \(f(\infty) - f(k)\), the cost saving from using \(s_2\) over \(s_1\), converges slower to zero than does \(\|s_1, s_2\| = [4(\theta - \theta') k]^{-1}\), the degree of approximation, which is of the order \(k^{-1}\).

In general, the sacrificed expected return from choosing \(s_2\) over \(s_1\) is of the order \(\|s_1, s_2\|\). If EPD holds, the sacrificed expected return is dominated by the cost saving and a player would be better off from choosing an absolutely continuous stochastic choice rule such as \(s_2\) rather than \(s_1\). We formalize this result in the following lemma.

**Lemma 12 (EPD implies continuous choice)** If cost functional \(c(\cdot)\) satisfies expensive perfect discrimination, then for any \(s \in S_M\), \(S_\lambda(s)\) consists only of absolutely continuous stochastic choice rules if \(\lambda > 0\).

**Proof.** Suppose \(s_1 \in S_\lambda(s)\) is not absolutely continuous. Then we can find an absolutely continuous \(s_2\) such that
\[
\|s_1, s_2\| < \frac{\lambda}{\pi} \cdot [c(s_1) - c(s_2)],
\]
where \(\pi\) is the uniform bound on \(|\pi(l, \theta)|\). Then, the gain from replacing \(s_1\) by \(s_2\) is
\[
V_\lambda(s_2|s) - V_\lambda(s_1|s)
= \int [s_2(\theta) - s_1(\theta)] \cdot \pi(s(\theta), \theta) \cdot g(\theta) d\theta + \lambda \cdot [c(s_1) - c(s_2)]
> -\int |s_2(\theta) - s_1(\theta)| \cdot \pi \cdot g(\theta) d\theta + \pi \cdot \|s_1, s_2\|
= 0,
\]
which contradicts the optimality of \(s_1\). ■

When the cost functional satisfies EPD, even though the step functions could be feasible, they are too expensive (relative to their continuous approximations) to be optimal. It is then a corollary that EPD and translation insensitivity imply the Laplacian selection.

**Proposition 13 (EPD and Laplacian Selection)** If \(c(\cdot)\) satisfies expensive perfect discrimination and translation insensitivity, then there is Laplacian selection.

This proposition strengthens Proposition 7, showing the same conclusion under a weaker assumption, highlighting that the continuity of the optimal choice rule is key to the limit
5 Information Acquisition

We use stochastic choice games as a reduced form description of games with information acquisition. In this section, we describe an explicit information acquisition game. Players first decide what experiment to acquire about the state of the world. They then decide whether to invest as a function of their realized signal, without observing either which experiments other players have chosen or the realizations of others’ experiments.

Because players do not observe other players’ information choice, the analysis of this information acquisition game reduces to the analysis of the stochastic choice game. One purpose of this section is to spell out how this reduction works in order to provide a full information acquisition interpretation of our main results about stochastic choice games. A player’s choice of experiment and an investment rule jointly determine a stochastic choice rule. We will illustrate how an arbitrary cost functional on experiments will translate into a reduced form cost functional on the smaller domain of stochastic choice rules. In particular, we will discuss how important examples of cost functionals on experiments translate into reduced form cost functionals on stochastic choice rules. Suppose players observe the true state with a uniformly distributed additive noise term and must pay to reduce the size of the noise. This cost functional on experiments reduces exactly to the max slope cost functional on stochastic choice rules. However, the assumption of additive noise is extremely inflexible, requiring that a player becomes equally informed about all regions of the state space. We will also discuss when we impose the natural Blackwell ordering on the cost functional on experiments, requiring that experiments that are less informative in the sense of Blackwell (1953) are cheaper.

A second purpose of this section is to discuss the reason for the unique equilibrium selection in our setting and to compare our results with existing results for coordination games with exogenous information.

5.1 The Information Acquisition Game

Before selecting an action, players can simultaneously and privately acquire information about $\theta$. Players observe conditionally independent real-valued signals that are informative about $\theta$; as always, the labelling of signals does not matter and we are using $\mathbb{R}$ as a signal space to economize on notation. Each player can pick an experiment $q$, where $q(\cdot|\theta) \in \Delta(\mathbb{R})$ is a probability measure on $\mathbb{R}$ conditional on $\theta$.

Footnote 11: The proposition would continue to hold if we replaced translation insensitivity with the local version of footnote 11 and required the following local stochastic continuous choice property: for all $\psi \in [\theta_{\min}, \theta_{\max}]$, there exists $\delta > 0$ such that $S_\lambda(\hat{s})$ consists only of absolutely continuous functions for all $\hat{s} \in B_{\delta}(1_{\{\theta \geq \psi\}})$ and $\lambda \in \mathbb{R}_{++}$. 
Let $Q$ denote the set of all experiments. An information cost functional $C : Q \rightarrow \mathbb{R}_+ \cup \{\infty\}$ assigns a cost to each experiment. A player incurs a cost $\lambda \cdot C(q)$ if she chooses experiment $q \in Q$. We assume that each player observes an independent experiment. A player’s strategy corresponds to an experiment $q$ together with an investment rule $\sigma : \mathbb{R} \rightarrow [0, 1]$, with $\sigma(x)$ being the probability of investing upon signal realization $x$. We can then study Nash equilibria of the information acquisition game.

The experiment $q$ and decision rule $\sigma$ then jointly determine the player’s stochastic choice rule,

$$s_{q, \sigma}(\theta) = \int q(x|\theta) \cdot \sigma(x) \, dx.$$  

For any given stochastic choice rule $s$, we can define the set of experiments $\tilde{Q}(s)$ that would be sufficient to implement the stochastic choice rule $s$:

$$\tilde{Q}(s) = \{ q \in Q | \text{there exists a } \sigma \text{ such that } s = s_{q, \sigma} \}.$$

Now we can define a cost functional for stochastic choice rules:

$$c(s) = \inf_{q \in \tilde{Q}(s)} C(q). \tag{9}$$

Now every equilibrium of the information acquisition game will correspond to an equilibrium of the stochastic choice game, and every equilibrium of the stochastic choice game will correspond to one or more equilibria of the information acquisition game. The stochastic choice game can be seen as a reduced form representation of the information acquisition game (where we abstract from information signals that are not used in equilibrium).

**Lemma 14** If $\{q_i, \sigma_i\}_{i \in [0,1]}$ is an equilibrium of the information acquisition game, then $\{s_i\}_{i \in [0,1]}$ is an equilibrium of the stochastic choice game, where each $s_i$ is the stochastic choice rule induced by $(q_i, \sigma_i)$.

**Lemma 15** Suppose $\{s_i\}_{i \in [0,1]}$ is an equilibrium of stochastic choice game. Then $\{q_i, \sigma_i\}_{i \in [0,1]}$ is an equilibrium of the information acquisition game, if $(q_i, \sigma_i)$ is a minimum cost way of attaining the stochastic choice rule $s_i$ for each player $i \in [0,1]$.

Blackwell (1953) introduced an informativeness ordering on experiments: one experiment is more informative than another if the latter can be obtained from the former from adding noise. A natural and important restriction that we may want to impose on the cost functional on experiments is that it respects Blackwell’s ordering. Formally, we will say that a cost functional on experiments is Blackwell-consistent if whenever an experiment $q$ is strictly less informative in the sense of Blackwell (1953) than a feasible (i.e., finite cost) experiment $\tilde{q}$, experiment $q$ is strictly less expensive than $\tilde{q}$. Blackwell-consistency puts no
restrictions on how many experiments are feasible and puts no restrictions on the relative informativeness of infeasible experiments.

As a leading example, consider the experiment corresponding to the global game information structure, where a player observes the true state plus additive uniformly distributed noise. Thus the player observes signal \( x = \theta + \frac{1}{k} \varepsilon \), and \( \varepsilon \) is uniformly distributed on the interval \([-\frac{1}{2}, \frac{1}{2}]\), so that the conditional probability density is

\[
q^k(x|\theta) = \begin{cases} 
    k, & \text{if } \theta - \frac{1}{2k} \leq x \leq \theta + \frac{1}{2k} \\
    0, & \text{otherwise}
\end{cases}
\]

We will call this class \( \{q^k\}_{k \in [0, \infty]} \) the global game information structures. It is parameterized by \( k \) which measures the accuracy of the signal. If a nondecreasing function \( f(k) \) represents the cost of accuracy \( k \), we obtain the (uniform noise) global game information cost functional

\[
C^k_{GG}(q) = \begin{cases} 
    f(k), & \text{if } q = q^k \\
    \infty, & \text{otherwise}
\end{cases}
\]

Here a \( q \) which does not correspond to a (uniform noise) global game information structure has an infinite cost, or is infeasible.

In terms of equilibrium predictions, the information acquisition game with the global game information cost functional is equivalent to a stochastic choice game equipped with the max slope cost functional introduced in Section 2.2. To see why, note that on the one hand any possible equilibrium strategy \( s \) of the stochastic choice game takes the piecewise linear form of (4) as discussed in Subsection 3.1, which has maximum slope \( k \in [0, \infty] \) and incurs cost \( f(k) \). On the other hand, any possible equilibrium strategy \( (\bar{q}^k, \sigma) \) of the information acquisition game induces a stochastic choice rule \( s \) given by

\[
s(\theta) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \sigma \left( \theta + \frac{\varepsilon}{k} \right) d\varepsilon, \tag{10}
\]

with slope

\[
s'(\theta) = k \left[ \sigma \left( \theta + \frac{1}{2k} \right) - \sigma \left( \theta - \frac{1}{2k} \right) \right].
\]

For any signal realization \( x \), we consider the investment rules \( \sigma(x) \in \{0, 1\} \) since the knife-edge case \( \sigma(x) \in (0, 1) \) is always weakly dominated by either \( \sigma(x) = 1 \) or \( \sigma(x) = 0 \). Hence the maximum slope of the induced stochastic choice rule is \( k \cdot [1 - 0] = k \). The induced stochastic rule takes the same piecewise linear form of (4) and incurs an information cost of \( f(k) \). Therefore, the information acquisition game with the global game information cost and the stochastic choice game with the max slope cost effectively lead to the same best
response and thus the same equilibrium outcomes.

For simplicity, we referred to this uniform noise cost functional as the global game cost functional even though we assumed a particular (uniform) distribution of the noise. If we allowed an arbitrary distribution of the noise, it would affect the shape of the stochastic choice rule and decreasing the precision of the signal would correspond to a stretch of the stochastic choice rule.

The global game cost functional on experiments trivially fails to be Blackwell-consistent, because most information structures are infeasible and thus it is infinitely costly to throw away information. However, there is a natural variation on the global game cost functional which will be Blackwell-consistent. Suppose that a player first had to pay a perception cost to acquire access to a signal of a certain precision. But then a player had to pay an attention cost to remember the signal that he acquired. Suppose that the attention cost was modelled as proportional to the entropy reduction, as formally described in Section 4.1. The total cost of an experiment was the sum of these two costs (we omit the tedious formal definition: one was given in the working paper version of this paper, Morris and Yang (2016)). This cost functional would clearly satisfy Blackwell-consistency. We will see that the implied cost saving to a player from not paying attention to certain signals will have important strategic implications.

5.2 Endogenous versus Exogenous Information Case and the Intuition for Uniqueness

One reason for explicitly describing the information acquisition game (rather than the reduced form stochastic choice game) is that it allows us to make an explicit comparison between the information acquisition game and a counterfactual coordination game with exogenous information. In particular, in the information acquisition game, players endogenously choose some information in equilibrium. Even though players do not observe the experiments that others have acquired, there will be common certainty (in equilibrium) what experiments others have acquired. If players had been exogenously endowed with those experiments, would there have been a unique equilibrium? The propositions below address this counterfactual question.

This question matters for the interpretation of our results. Suppose that there is a unique equilibrium and Laplacian selection in the stochastic choice game. If we interpret the stochastic choice game as a reduced form of the information acquisition game, what explains the uniqueness and selection? Is it because players will endogenously choose information that necessarily gives rise to a unique equilibrium? Proposition 16 shows that this is true with the global game cost functional on experiments. Or is there a unique equilibrium and Laplacian selection despite the fact that players will endogenously end up with information that is consistent with multiple equilibria? Proposition 17 shows that this
is true if the cost functional on experiments is Blackwell-consistent.

The latter proposition will hold if the probability assigned to dominant strategy regions of the state space is not too large. In particular, if players were told that there was symmetric information that all players knew was that invest was the Laplacian action, or that not invest was the Laplacian action, this would still be consistent with multiple equilibria. Formally, assume $g$, the density of the common prior, satisfies

$$\int_{-\infty}^{\theta^{**}} \pi(1, \theta) g(\theta) d\theta > 0 \quad (11)$$

and

$$\int_{\theta^{**}}^{\infty} \pi(0, \theta) g(\theta) d\theta < 0 \quad (12)$$

Now we have:

**Proposition 16** Suppose that the cost functional on experiments is the global game cost functional which satisfies that $f(k) \in [0, \infty]$ is weakly increasing, $f(0) = 0$ and $\lim_{k \to \infty} [f(\infty) - f(k)]$.

Then, for any $\delta > 0$, we can choose $\lambda > 0$, such that the following holds for all $\lambda \leq \lambda^*$. Suppose $k^\lambda$ is the precision of the global game information structures chosen in an equilibrium of the information acquisition game with cost parameter $\lambda$. The exogenous information coordination game in which each player is endowed with the global game information structure with precision $k^\lambda$ has an equilibrium that induces a stochastic choice rule $s^*$ with $\|s^*, 1_{\theta \geq \theta^{**}}\| \leq \delta$ for all players.

This result is a generalization of results that appear in Szkup and Trevino (2015) and Yang (2013). With global game information acquisition, uniqueness and Laplacian selection arise because players necessarily acquire information in a way that gives rise to Laplacian interim beliefs.

Recall that under Blackwell-consistency, players always acquire experiments with binary signals in equilibrium. When Laplacian selection holds, it is natural to label signals according to the actions that will be played in the unique equilibrium of the stochastic choice game.

**Proposition 17** Suppose that the cost functional on experiments is Blackwell-consistent and satisfies Laplacian selection. Suppose the common prior satisfies (11) and (12). Then there exists a $\lambda > 0$, such that the following holds for any $\lambda \in (0, \lambda^*)$. Let $q^\lambda$ be the experiment chosen by each player in equilibrium of the information acquisition game with cost parameter $\lambda$. The exogenous information coordination game in which each player is endowed with experiment $q^\lambda$ has four equilibria: i) each player always invests regardless of the recommendation of her signal; ii) each player never invests regardless of her signal; iii)
each player follows the recommendation of her signal; iv) each player chooses the action opposite to the recommendation of her signal.

Since players will necessarily acquire binary signal experiments in equilibrium, if they had exogenously been given such experiments, they would never form the Laplacian interim beliefs and there would have been multiple equilibria. These "equilibria", except the one corresponding to the Laplacian selection (i.e., (iii)) in the statement of Proposition 17, are not equilibria of the information acquisition game, because the players would have not acquired such information in the first place given the prescribed strategies of these equilibria. Therefore, there is a deeper and more general reason for uniqueness and Laplacian selection in our formalization of stochastic choice games. The key is the continuity of the stochastic choice rules regardless of the details of the information structure or information acquisition. The global game reasoning is a special case in the sense that such continuity is a "built-in" feature of the global game information structure.

Recall that our definition of Blackwell-consistency was: whenever an experiment \( q \) is strictly less informative in the sense of Blackwell (1953) than experiment \( \bar{q} \), then experiment \( q \) is strictly less expensive than \( \bar{q} \) (i.e., \( C(q) < C(\bar{q}) \)).\(^{17}\) A weaker definition of Blackwell-consistency would be that whenever an experiment \( q \) is weakly less informative in the sense of Blackwell (1953) than experiment \( \bar{q} \), then experiment \( q \) is weakly less expensive than \( \bar{q} \) (i.e., \( C(q) \leq C(\bar{q}) \)). In this case, the proposition would remain true if we restricted attention to equilibria of the information acquisition game where binary signal information structures were chosen.

### 6 Monotonicity

We made the simplifying assumption that players were restricted to choose monotonic stochastic choice rules. In this section, we report two alternative ways in which we can relax this restriction. Because each result has some different drawbacks which we will describe, it is convenient to separate our relaxation of monotonicity from the main statements of our results.

We first introduce a weak restriction on the cost functional - variation-aversion - under which players will always have a monotonic best response if they anticipate that others will choose monotonic stochastic choice rules. Thus existence of monotonic equilibria is guaranteed even if players are not restricted to choose monotonic stochastic choice rules. Our main result then ensures limit uniqueness under monotonic equilibria. We will see that variation-aversion also captures the idea that it is harder to distinguish nearby states, and in this sense is a very weak property. It is satisfied by our leading max slope cost functional.

\(^{17}\)We adopted the convention that \( C(q) < C(\bar{q}) \) if \( C(q) = C(\bar{q}) = \infty \).
example. However, one weakness is that the definition depends on the prior on states, which some may not think is a natural property.

Next, we consider a stronger submodularity restriction on the cost functional that ensures that the stochastic choice game is supermodular, so that there are largest and smallest equilibria that are monotonic by standard arguments, so that our uniqueness of monotonic equilibria results implies the non-existence of non-monotonic equilibria, which is the ideal result. However, submodularity does not have a compelling motivation and it is not satisfied by our leading max slope cost functional.

6.1 Existence of Monotonic Equilibria under a Variation-Averse Cost Functional

A stochastic choice rule \( s \) induces a probability distribution over investment probabilities \( F_s \). Formally, for any stochastic choice rule \( s \) we define the c.d.f. \( F_s \) as

\[
F_s(p) = \int_{\{\theta: s(\theta) \leq p \}} g(\theta) \, d\theta.
\]

We want to capture the idea that, holding fixed the induced distribution of investment probabilities, it is always cheaper to choose a monotonic stochastic choice rule since this implies less variability in investment probabilities. For any distribution over investment probabilities \( F \), we write \( \hat{s}[F] \) for the unique monotonic (nondecreasing) stochastic choice rule inducing \( F \), so

\[
\hat{s}[F](\theta) = F^{-1}(G(\theta)),
\]

where \( F^{-1}(x) = \inf\{p \in [0,1] : F(p) \geq x\} \).

**Assumption A9 (Variation-Averse):** The cost functional \( c(\cdot) \) is variation averse if \( c(\hat{s}[F]) \leq c(s) \) for all \( s \in S \).

Now observe that a player’s expected payoff

\[
V_\lambda(\tilde{s}|s) = \int \pi(s(\theta), \theta) \cdot \tilde{s}(\theta) g(\theta) \, d\theta - \lambda \cdot c(\tilde{s})
\]

will be maximized by a monotonic \( \tilde{s} \) whenever \( s \) is monotonic. To see why, observe that \( \int \pi(s(\theta), \theta) \cdot \tilde{s}(\theta) g(\theta) \, d\theta \) will always be increased by replacing \( \tilde{s} \) with \( \hat{s}[F_s] \), while the variation-aversion of \( c \) ensures that \( c(\tilde{s}) \) will be decreased by replacing \( \tilde{s} \) with \( \hat{s}[F_s] \).
6.2 Non-Existence of Non-Monotonic Equilibria under a Submodular Cost Functional

We have been focusing on monotonic stochastic choice rules so far. This subsection provides conditions under which this restriction does not weaken the results given our research purpose. We first define a partial order $\succeq$ on $S$, the set of all stochastic choice rules. In particular, for any $s_1$ and $s_2$ in $S$, $s_2 \succeq s_1$ if and only if $s_2(\theta) \geq s_1(\theta)$ almost surely under the common prior. Accordingly, $\lor$ and $\land$, the join and meet operators, take the form $[s_2 \lor s_1](\theta) = \max\{s_2(\theta), s_1(\theta)\}$ and $[s_2 \land s_1](\theta) = \min\{s_2(\theta), s_1(\theta)\}$, respectively. It is straightforward to see that for any $s_1$ and $s_2$ in $S$, both their join and meet belong to $S$, so that $(S, \succeq)$ forms a complete lattice. We next introduce the following assumption on the cost functional $c(\cdot)$:

**Assumption A10 (Submodularity):** The cost functional $c(\cdot)$ is submodular on $(S, \succeq)$; i.e., $c(s_2 \lor s_1) + c(s_2 \land s_1) \leq c(s_1) + c(s_2)$ for all $s_1$ and $s_2$ in $S$.

Intuitively, the meet and the join are "flatter" than the original choice rules so that they are jointly cheaper. Since the order $\succeq$ is defined in a coordinate-wise manner, this assumption amounts to a decreasing-difference property of $c(s)$ (i.e., the increasing-difference property of $-c(s)$) for any pair of states $\theta_1$ and $\theta_2$.

Note that each player's expected payoff

$$V_\lambda(\bar{s}|s) = \int \pi(s(\theta), \theta) \cdot \bar{s}(\theta) g(\theta) d\theta - \lambda \cdot c(\bar{s})$$

is supermodular in her own choice rule $\bar{s}$, because the first term is linear in $\bar{s}$ and the second term $-\lambda \cdot c(\bar{s})$ is supermodular. In addition, the "Strategic Complementarities" Assumption A1 implies that $V_\lambda(\bar{s}|s)$ has increasing difference over $\bar{s}$ and other players' aggregate choice rule $s$. Hence, the stochastic choice game is a supermodular game. Its equilibria form a sublattice of $(S, \succeq)$ with both the largest and the smallest equilibria being monotonic stochastic choice rules. According to the results in the previous sections, as the cost vanishes, these two extreme equilibria converge to the unique equilibrium and so do the equilibria between them. Therefore, Assumption A10 helps us obtain the result of Laplacian selection without requiring the choice rules to be monotonic.

7 Discussion

7.1 Foundations for Cost Functionals

7.1.1 Learning from Sampling a Diffusion

One specific model of information acquisition is that players learn by observing a drift diffusion process that is informative about the state. Any stopping rule will then give rise
to an experiment. It is natural to identify the (ex ante) cost of the experiment with the expected stopping time. We do not in general know which experiments could be derived this way, nor - in this continuum state case - if there is going to be a simple expression for the cost functional. However, Strack (2016) shows that - under the weak assumption that the stopping time is uniformly bounded - any feasible stochastic choice rule will be continuous. Under this foundation, we know that infeasible perfect discrimination is satisfied, even though we do neither have a characterization of which stochastic choice rules are feasible, nor do we have a representation of the costs of stochastic choice rules which are feasible.

For purposes of this paper, it is not necessary to establish the cost of feasible stochastic choice rules once one has shown that the discontinuous stochastic choice rules are infeasible. However, it is of independent interest to derive foundations for the cost functional more generally. For finite state spaces, Morris and Strack (2018) do characterize the cost functional on feasible stochastic choice rules derived from the Wald problem, where players observe a fixed diffusion and the cost is the expected stopping time. With two states, all experiments are feasible. With more than two states, they provide a characterization of which experiments are feasible. Hebert and Woodford (2016) consider a case where an individual can control what diffusion to observe, so that any experiment is feasible. They describe a natural procedure for taking a limit from finite to infinite state spaces, and identify conditions under which infeasible perfect discrimination holds.

### 7.1.2 Axiomatic

Pomatto, Strack, and Tamuz (2019) propose four natural axioms for the cost of information: (1) Blackwell-consistency, (2) linearity in independent experiments, (3) linearity in probability, and (4) continuity. We have already motivated the Blackwell-consistency condition. The linearity assumptions capture the idea that there is a constant cost of information. They provide a representation for cost functionals satisfying those four axioms. The cost of information can be written as the weighted sum - across pairs of states - of the Kullback-Leibler divergences of the distributions over signals for those two states. The weight on a pair of states is naturally interpreted as the difficulty of distinguishing that pair of states. They then put extra structure on the state space, letting states be points on the real line. Now if one assumes, in addition to the initial four axioms, that (5) the cost of distinguishing a pair of states depends only on the distance between them, and (6) the cost of observing states plus Gaussian noise of fixed variance does not depend on the state space, then the distance between states that appears in the representation is the reciprocal of squared distance. This provides an elegant foundation for a cost functional reflecting the distance between states. They then show that this cost functional implies that stochastic choice rules satisfy a Lipschitz continuity condition, in their setting where the state space is a finite subset of the real line. This property implies that a cost functional on stochastic
choice rules on the real line would satisfy infeasible perfect discrimination if it was the limit
finite state cost functionals satisfying their six axioms.

7.2 Functional Forms

Cost functions used in economic analysis are often rather ad hoc and it is hard to know how
results depend on the exact modelling choice. This is true, for example, of the additive
noise used in global games. One reason why entropy reduction is used as a cost functional is
because it is both tractable and flexible. But we have argued that the entropic cost function
builds in cheap perfect discrimination and this is rarely a natural assumption. It would be
nice to have tractable functional forms that are natural and sufficiently rich to encompass
both cheap perfect discrimination and infeasible perfect discrimination. We introduced one
equation that did this: the max slope cost functional. We showed that this cost functional
on stochastic choice rules corresponds in terms of equilibrium analysis to the global game
cost functional on experiments. In the working paper version of this paper, Morris and
Yang (2016), we discussed a rich parameterized class of cost functions that were increasing
in some average of the slope of the derivative of the stochastic choice rule, where we could
give explicit expressions for when the cost function satisfied CPD giving limit uniqueness of
limit multiplicity.

Alternative paths to well behaved cost functionals would be to take the limits of the
sequential sampling and axiomatic cost functionals described in the previous sub-section.

7.3 Learning about Others’ Actions

A maintained assumption in our analysis is that players acquire information about the state
only. Hence, their signals are conditionally independent given the state. Denti (2018) and
Hoshino (2018) consider the problem when (finitely many) players can acquire information
about others’ information, which essentially allows the players’ signals to be correlated even
if conditioned on the state.

Denti (2018) focusses on entropy reduction, which prevents the stochastic choice rules
from attaining 0 or 1 as the marginal cost of doing so is infinite. Consequently, the players’
actions contain residual uncertainty other than that originated from the uncertain state,
which allows the players to correlate their actions through acquiring others’ information.
This gives rise to smoother best responses and a different answer for us for the entropy cost
function - limit uniqueness - in this case. Note that if he assumed an information cost
with the Lipschitz property, he too would get limit multiplicity. This is because the players
would choose step functions like $1_{\{a \geq \psi\}}$ so that their actions are deterministic functions of
the state. Hence, acquiring information about others’ information amounts to acquiring
information about the state. Therefore, the game reduces to the one studied in Section 4.1
and limit multiplicity follows.

Hoshino (2018) shows that for some assumptions on the cost of information, there is limit uniqueness but equilibrium behavior from the underlying complete information game can be selected. There are finite states in Hoshino (2018), so there is no role for building in the idea that nearby states are hard to distinguish. In this sense, infeasible perfect discrimination is ruled out.

7.4 Evidence on Informational Costs

The property that nearby states are harder to distinguish than distant states seems natural in any setting where states have a natural metric and correspond to physical outcomes. Jazayeri and Movshon (2007) examine decision makers' ability to discriminate the direction of dots on the screen when they face a threshold decision problem. There is evidence that subjects are better at discriminating states on either side of the threshold, consistent with optimal allocation of scarce resources to discriminate. However, the ability to discriminate between states on either side of the threshold disappears as we approach the threshold, giving rise to continuous choice in our sense in this setting. The allocation of resources in this case is presumably at the unconscious neuro level.

Caplin and Dean (2015) introduce a natural laboratory experiment to identify the cost functional on information in a finite state setting, where the allocation of resources is presumably a conscious choice (e.g., how much time to devote to the task). Subjects observe a screen with a mixture of balls of different colors, say red and blue. On one screen there are more red balls than the other, and subjects are asked to distinguish between the number of balls. Dewan and Neligh (2017) establish that it is easier to distinguish between states that are further apart in the sense of the number of balls. Thus in a finite state problem, they exhibit the key property that matters in this paper: it is harder to distinguish nearby states than distant states. On the other hand, Dean and Neligh (2017) consider an alternative treatment where players are asked to distinguish which letter appears the most on a screen full of letters. There is arguably no natural order on states in this setting and there is less evidence that any pair of states are easier to distinguish than any other pair.\footnote{This treatment was originally developed in work of Dean, Morris and Trevino on "Endogenous information structures and play in global games."} This is thus a finite state analogue of the cheap perfect discrimination property.

7.5 Potential Games

We established our results for symmetric binary action continuum player games. In the global games literature, all these assumptions have been relaxed. In particular, Frankel, Morris, and Pauzner (2003) examine global games where the underlying coordination game...
has arbitrary numbers of players and actions, and asymmetric payoffs, maintaining the assumption that payoffs are supermodular in actions and satisfy increasing differences with respect to the state. They show two kinds of results. The first result is that there is limit uniqueness in general: if players observe additive noisy signals of payoffs, then there is a unique equilibrium in the limit as the noise goes to zero. However, in general, the equilibrium selected depends on the distribution of the noise. The second result gives sufficient conditions for noise independent selection (so the limit equilibrium does not depend on the shape of the noise). Frankel, Morris, and Pauzner (2003) report some sufficient conditions based on generalized potential games; Monderer and Shapley (1996) introduced potential games and Morris and Ui (2005) analyze the relevant generalizations. While we have not appealed to potential arguments in this paper, the Laplacian selection in the symmetric binary action continuum player games is the potential maximizing Nash equilibrium and we conjecture that generalizations of our results would go through for the generalized potential games discussed in Frankel, Morris, and Pauzner (2003), but this extension is beyond the scope of this paper.

References


8 Appendix

Proof of Lemma 8.

Proof. When \( \lambda = 0 \), the player can choose any choice rule for free and the optimal choice rule is \( 1_{\{\theta \geq s_\ast\}} \).

Now consider the case \( \lambda > 0 \). For any \( \delta > 0 \), define

\[
\alpha (\delta) = \inf_{l \in [0, 1]} \min \left( \pi (l, \theta_l + \delta), -\pi (l, \theta_l - \delta) \right).
\]

Note that given \( \delta > 0 \), \( \min \left( \pi (l, \theta_l + \delta), -\pi (l, \theta_l - \delta) \right) \) is a function of \( l \) on a compact set \([0, 1] \). By Assumption A4, this function is always strictly positive. Hence, its infimum on \([0, 1] \) exists and is strictly positive. That is, \( \alpha (\delta) > 0 \) for all \( \delta > 0 \). In addition, for any \( s \in S_M \) and \( \theta \notin [\theta_s - \delta, \theta_s + \delta] \), we have

\[
\{\pi (s (\theta_l), \theta) \geq \alpha (\delta) \}
\]

where the first inequality follows Assumptions A1 and A4, and the second inequality follows the definition of \( \alpha (\delta) \).

Let \( \bar{g} = \sup_{\theta \in \mathbb{R}} g (\theta) < \infty \) and choose \( \delta = \frac{\rho}{2 \bar{g}} \). For any \( s_\lambda \in S_\lambda (s) \), note that

\[
\int_{-\infty}^{\infty} \pi (s (\theta_l), \theta) \cdot \left[ 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \right] g (\theta) d\theta
\]

\[
\geq \int_{\theta_s - \delta}^{\theta_s + \delta} \pi (s (\theta_l), \theta) \cdot \left[ 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \right] g (\theta) d\theta
\]

\[
\geq \int_{-\infty}^{\theta_s + \delta} z (\delta) \cdot \left| 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \right| g (\theta) d\theta - \int_{\theta_s - \delta}^{\theta_s + \delta} z (\delta) \cdot \left| 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \right| g (\theta) d\theta
\]

\[
\geq z (\delta) \cdot \left[ \| 1_{\{\theta \geq \theta_l\}}, s_\lambda \| - 2 \cdot \bar{g} \cdot \delta \right] = z (\delta) \cdot \left[ \| 1_{\{\theta \geq \theta_l\}}, s_\lambda \| - \frac{\rho}{2} \right]. \tag{14}
\]

The first inequality holds since \( \pi (s (\theta_l), \theta) \) and \( 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \) always have the same sign and thus

\[
\int_{\theta_s - \delta}^{\theta_s + \delta} \pi (s (\theta_l), \theta) \cdot \left[ 1_{\{\theta \geq \theta_l\}} - s_\lambda (\theta) \right] g (\theta) d\theta > 0 ,
\]

and the second inequality follows (13).

Let \( \varepsilon = z \left( \frac{\rho}{2 \bar{g}} \right) \cdot \frac{\rho}{2 \bar{g}} \), where \( \pi \) is the uniform upper bound of \( \pi (l, \theta) \). For each \( \theta_l \in [\theta_{\min}, \theta_{\max}] \), according to Assumption A8, there exists a choice rule \( \tilde{s}_l \in B_\varepsilon \left( 1_{\{\theta \geq \theta_l\}} \right) \) such that \( c (\tilde{s}_l) < \infty \). Note that \( \{1_{\{\theta \geq \theta_l\}} : \theta \in [\theta_{\min}, \theta_{\max}] \} \) is sequentially compact and thus
compact in \( S \) under the \( L^1 \)-metric. Since \( \{ B_\varepsilon (1_{\{\theta \geq \theta_i\}}) : \theta_i \in [\theta_{\min}, \theta_{\max}] \} \) is an open cover of \( \{1_{\{\theta \geq \theta_i\}} : \theta_i \in [\theta_{\min}, \theta_{\max}]\} \), it has a finite sub-cover. Let

\[
\left\{ B_\varepsilon (1_{\{\theta \geq \theta^1\}}), B_\varepsilon (1_{\{\theta \geq \theta^2\}}), \ldots, B_\varepsilon (1_{\{\theta \geq \theta^N\}}) \right\}
\]

denote the finite sub-cover, and \( \tilde{s}^1 \in B_\varepsilon (1_{\{\theta \geq \theta^1\}}) \), \( \tilde{s}^2 \in B_\varepsilon (1_{\{\theta \geq \theta^2\}}) \), \ldots, \( \tilde{s}^N \in B_\varepsilon (1_{\{\theta \geq \theta^N\}}) \) be the corresponding choice rules with finite costs. Define \( \bar{c} = \max \left( c \left( \tilde{s}^1 \right), c \left( \tilde{s}^2 \right), \ldots, c \left( \tilde{s}^N \right) \right) \), which is finite. By definition, any \( s \in S_M \) induces a cutoff \( \theta_s \in [\theta_{\min}, \theta_{\max}] \) so that \( 1_{\{\theta \geq \theta_s\}} \) belongs to some member \( B_\varepsilon (1_{\{\theta \geq \theta^1\}}) \) of the sub-cover and thus \( \| 1_{\{\theta \geq \theta_s\}}, \tilde{s}^n \| < \varepsilon \). Consider the binary decision problem when the aggregate stochastic choice rule of all other players is given by \( s \). Absent the cost, the ideal strategy is \( 1_{\{\theta \geq \theta_s\}} \). The sacrificed expected return from using \( \tilde{s}^n \) instead of \( 1_{\{\theta \geq \theta_s\}} \) is

\[
\int_{-\infty}^{\infty} \pi (s (\theta), \theta) \cdot \left[ 1_{\{\theta \geq \theta_s\}} - \tilde{s}^n (\theta) \right] g (\theta) d\theta \\
\leq \pi \cdot \| 1_{\{\theta \geq \theta_s\}}, \tilde{s}^n \| < z \left( \frac{\rho}{4g} \right) \cdot \frac{\rho}{4}, \tag{15}
\]

Combining (14) and (15) leads to

\[
\int_{-\infty}^{\infty} \pi (s (\theta), \theta) \cdot \left[ \tilde{s}^n (\theta) - s_\lambda (\theta) \right] g (\theta) d\theta > z \left( \frac{\rho}{4g} \right) \cdot \left[ \| 1_{\{\theta \geq \theta_s\}}, s_\lambda \| - \frac{3}{4} \rho \right],
\]

where the left hand side is the sacrificed expected return from using \( \tilde{s}^n \) instead of \( s_\lambda (\theta) \).

The optimality of \( s_\lambda \) implies

\[
\int_{-\infty}^{\infty} \pi (s (\theta), \theta) \cdot \left[ \tilde{s}^n (\theta) - s_\lambda (\theta) \right] g (\theta) d\theta \leq \lambda \cdot \left[ c \left( \tilde{s}^n \right) - c \left( s_\lambda \right) \right] \leq \lambda \cdot \bar{c},
\]

where the second inequality follows the definition of \( \bar{c} \). The above two inequalities imply

\[
z \left( \frac{\rho}{4g} \right) \cdot \left[ \| 1_{\{\theta \geq \theta_s\}}, s_\lambda \| - \frac{3}{4} \rho \right] < \lambda \cdot \bar{c},
\]

i.e.,

\[
\| 1_{\{\theta \geq \theta_s\}}, s_\lambda \| < \frac{\lambda \cdot \bar{c}}{z \left( \frac{\rho}{4g} \right)} + \frac{3}{4} \rho.
\]

Let \( \bar{\lambda} = \frac{z \left( \frac{\rho}{4g} \right)}{\bar{c}} \). Therefore, for all \( s \in S_M \) and \( \lambda < \bar{\lambda} \), we have \( \| 1_{\{\theta \geq \theta_s\}}, s_\lambda \| < \rho \). ■

**Proof of Proposition 7.**

38
Proof. By Lemma 8, we have

$$\lim_{\lambda \to 0} \sup_{\hat{s} \in S_{\lambda}(s) \text{ and } s \in S_M} \left\| \hat{s}, 1_{\{\theta \geq \theta_s\}} \right\| = 0. \quad (16)$$

Let \( \{\hat{s}_{i,\lambda}\}_{i \in [0,1]} \) denote an equilibrium of the \( \lambda \)-game. Then the aggregate stochastic choice rule is given by

$$\hat{s}_{\lambda}^* (\theta) = \int_{i \in [0,1]} s_{i,\lambda}^* (\theta) \, di,$$

which by Assumption A4 induces a threshold \( \theta_{\lambda}^* \) such that \( \pi (\hat{s}_{\lambda}^* (\theta), \theta) > 0 \) if \( \theta > \theta_{\lambda}^* \) and \( \pi (\hat{s}_{\lambda}^* (\theta), \theta) < 0 \) if \( \theta < \theta_{\lambda}^* \). By (16),

$$\lim_{\lambda \to 0} \left\| s_{i,\lambda}^*, 1_{\{\theta \geq \theta_{\lambda}^*\}} \right\| = 0.$$

Since

$$\left\| s_{i,\lambda}^*, 1_{\{\theta \geq \theta_{\lambda}^*\}} \right\| \leq \left\| s_{i,\lambda}^*, 1_{\{\theta \geq \theta_{\lambda}^*\}} \right\| + \left\| 1_{\{\theta \geq \theta_{\lambda}^*\}}, 1_{\{\theta \geq \theta^{**}\}} \right\|,$$

it suffices to show that \( \theta_{\lambda}^* \) becomes arbitrarily close to \( \theta^{**} \) as \( \lambda \to 0 \).

We next show that \( \int_{-\infty}^{\infty} \pi (\hat{s}_{\lambda}^* (\theta), \theta) \cdot g (\theta) \, d\hat{s}_{\lambda}^* (\theta) \) is arbitrarily close to zero when \( \lambda \) is small enough. Consider player \( i \)'s expected payoff from slightly shifting her equilibrium strategy \( s_{i,\lambda}^* \) to \( T_{\Delta} s_{i,\lambda}^* \), which is given by

$$W (\Delta) = \int_{-\infty}^{\infty} \pi (\hat{s}_{\lambda}^* (\theta), \theta) \cdot s_{i,\lambda}^* (\theta + \Delta) \cdot g (\theta) \, d\theta - \lambda \cdot c \left( T_{\Delta} s_{i,\lambda}^* \right).$$

The player should not benefit from this deviation, which implies \( W' (0) = 0 \), i.e.,

$$\int_{-\infty}^{\infty} \pi (\hat{s}_{\lambda}^* (\theta), \theta) \cdot \frac{ds_{i,\lambda}^* (\theta)}{d\theta} \cdot g (\theta) \, d\theta - \lambda \cdot \frac{d \left( T_{\Delta} s_{i,\lambda}^* \right)}{d\Delta} \bigg|_{\Delta = 0} = 0.$$

Here \( W' (0) \) takes this form because \( s_{i,\lambda}^* \) is absolutely continuous. In addition, translation insensitivity implies \( -K < \frac{dc (T_{\Delta} s_{i,\lambda}^*)}{d\Delta} \bigg|_{\Delta = 0} < K \) for some \( K > 0 \). Hence, for any small
\[ \varepsilon > 0, \text{ by choosing } \lambda \in (0, \varepsilon) \text{ we obtain} \]

\[ -K\varepsilon < \int_{-\infty}^{\infty} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \cdot g (\theta) \, ds^*_\lambda (\theta) < K\varepsilon . \]

The above inequality holds for all \( i \in [0, 1] \), and thus implies

\[ -K\varepsilon < \int_{-\infty}^{\infty} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \cdot g (\theta) \, ds^*_\lambda (\theta) < K\varepsilon , \]

i.e.,

\[ \left| \int_{-\infty}^{\infty} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \cdot g (\theta) \, ds^*_\lambda (\theta) \right| < K\varepsilon . \quad (17) \]

Since the density function \( g (\theta) \) is continuous on \([\theta_{\min}, \theta_{\max}]\), it is also uniformly continuous on \([\theta_{\min}, \theta_{\max}]\). For the same reason, \( \Pi (\theta) \) is also uniformly continuous on \([\theta_{\min}, \theta_{\max}]\). Hence, for any \( \varepsilon > 0 \), we can find an \( \eta > 0 \) such that \( |g (\theta) - g (\theta')| < \varepsilon \) and \( |\Pi (\theta) - \Pi (\theta')| < \varepsilon \) for all \( \theta, \theta' \in [\theta_{\min}, \theta_{\max}] \) and \( |\theta - \theta'| < 2\eta \). Without loss of generality, we can choose \( \eta < \varepsilon \). By (16), for all \( i \), the effective strategy \( s^*_{i, \lambda} \) converges to \( 1_{\{\theta \geq \theta^*_i\}} \) in \( L^1 \)-norm, so does the aggregate effective strategy \( \tilde{s}^*_\lambda \). Together with the monotonicity of \( \tilde{s}^*_\lambda \), this implies the existence of a \( \lambda_1 > 0 \) such that for all \( \lambda \in (0, \lambda_1) \), \( \left| \tilde{s}^*_\lambda (\theta) - 1_{\{\theta \geq \theta^*_i\}} \right| < \varepsilon \) for all \( \theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty) \). Choosing \( \lambda \in (0, \min (\lambda_1, \varepsilon)) \), by (17), we obtain

\[
\left| \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \cdot g (\theta) \, ds^*_\lambda (\theta) \right| < 2L\varepsilon + K\varepsilon ,
\]

where \( L > 0 \) is the uniform bound for \( |\pi (l, \theta)| \) and \( \eta = \sup_{\theta \in \mathbb{R}} |g (\theta)| < \infty \). By the definition of \( \eta \), \( |g (\theta) - g (\theta^*_\lambda)| < \varepsilon \) for all \( \theta \in [\theta^*_\lambda - \eta, \theta^*_\lambda + \eta] \). Hence,

\[
\left| g (\theta^*_\lambda) \cdot \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \, ds^*_\lambda (\theta) - \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi \left( \tilde{s}^*_\lambda (\theta) , \theta \right) \cdot g (\theta) \, ds^*_\lambda (\theta) \right| < L\varepsilon . \quad (19)
\]
Inequalities (18) and (19) imply
\[
\left| \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi(\hat{\lambda}(\theta), \theta) d\hat{\lambda}(\theta) \right| < \frac{2L\bar{g} + K + L}{g} \varepsilon,
\]
(20)
where \( \bar{g} = \inf_{\theta \in [\theta_{\min}, \theta_{\max}]} g(\theta) > 0 \) since \( g \) is continuous and strictly positive on \( [\theta_{\min}, \theta_{\max}] \) by assumption.

Next note that
\[
\int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda - \eta) ds - \int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda + \eta) ds \leq \left| \Pi(\theta^*_\lambda + \eta) - \Pi(\theta^*_\lambda - \eta) \right| + 4L \varepsilon < \varepsilon + 4L \varepsilon,
\]
(21)
where the first inequality follows the fact that \( \hat{\lambda}(\theta) - 1_{\{\theta \geq \theta^*_\lambda\}} < \varepsilon \) for all \( \theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty) \), and the second inequality follows the uniform continuity of \( \Pi(\theta) \) on \( [\theta_{\min}, \theta_{\max}] \).

Further note that Assumption A2 implies
\[
\int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda - \eta) ds \leq \int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda + \eta) ds < \left( \frac{2L\bar{g} + K + L}{g} + 4L + 1 \right) \varepsilon.
\]
(22)
which together with (20) and (21) implies
\[
- \left( \frac{2L\bar{g} + K + L}{g} + 4L + 1 \right) \varepsilon < \left| \int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda - \eta) ds \right| < \left( \frac{2L\bar{g} + K + L}{g} + 4L + 1 \right) \varepsilon.
\]

By Assumption A2, the monotonicity of \( \pi(s, \theta) \) in \( \theta \) implies
\[
\left| \int_{\hat{\lambda}(\theta^*_\lambda - \eta)}^{\hat{\lambda}(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda) ds \right| < \left( \frac{2L\bar{g} + K + L}{g} + 4L + 1 \right) \varepsilon.
\]
Again, using the fact that \( \hat{\lambda}(\theta) - 1_{\{\theta \geq \theta^*_\lambda\}} < \varepsilon \) for all \( \theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty) \), the
above inequality implies
\[
\left| \int_0^1 \pi(s, \theta_\lambda^*) \, ds \right| < \left( \frac{2Lg + K + L}{g} + 6L + 1 \right) \varepsilon .
\]
Therefore, we have
\[
\lim_{\lambda \to 0} \Pi(\theta_\lambda^*) = 0,
\]
which implies
\[
\lim_{\lambda \to 0} \theta_\lambda^* = \theta^{**}
\]
according to Assumptions A5 and A6. \( \blacksquare \)

Lemma 18 The entropy reduction information cost satisfies CPD for all \( \psi \in (\theta_{\min}, \theta_{\max}) \).

Proof. For any stochastic choice rule \( s \), the associated entropy reduction is
\[
c(s) = E[h(s(\theta))] - h(E(s(\theta))) ,
\]
where \( h : [0, 1] \to \mathbb{R} \) is given by
\[
h(x) = x \ln x + (1 - x) \ln (1 - x) .
\]
The exact functional form of \( h \) is not important for the proof. The proof below works for any convex function \( h : [0, 1] \to \mathbb{R} \).

Let \( p_1(s) = E(s(\theta)) \) denote the unconditional probability that action 1 is chosen under stochastic choice rule \( s \).

Since \( \psi \in (\theta_{\min}, \theta_{\max}) \) and the prior density \( g \) is positive over \([\theta_{\min}, \theta_{\max}]\), we have \( E(1_{\{\theta \geq \psi\}}) \in (0, 1) \). Choose \( \xi > 0 \) such that \( E(1_{\{\theta \geq \psi\}}) \in (\xi, 1 - \xi) \). Then choose \( \rho > 0 \) small enough such that for all \( s \in B_\rho(1_{\{\theta \geq \psi\}}) \), \( p_1(s) \in (\xi, 1 - \xi) \). Note that for small \( \varepsilon > 0 \), \( s \in B_\rho(1_{\{\theta \geq \psi\}}) \) implies \( L_{\psi,1}^s \in B_\rho(1_{\{\theta \geq \psi\}}) \). Let \( A(s) = \{ \theta : L_{\psi,1}^s(\theta) \neq s(\theta) \} \).

By the convexity of \( c(s) \), we have
\[
c(L_{\psi,1}^s) - c(s) \leq \int_{A(s)} \left[ h'(L_{\psi,1}^s(\theta)) - h'(p_1(L_{\psi,1}^s(s))) \right] \left( L_{\psi,1}^s(s(\theta)) - s(\theta) \right) dG(\theta)
\]
and
\[
c(L_{\psi,1}^s) - c(s) \geq \int_{A(s)} \left[ h'(s(\theta)) - h'(p_1(s)) \right] \left( L_{\psi,1}^s(s(\theta)) - s(\theta) \right) dG(\theta) ,
\]
where
\[
h'(s(\theta)) - h'(p_1(s))
\]
and
\[ h' \left( L_{\psi}^s(\theta) \right) - h' \left( p_1 \left( L_{\psi}^s \right) \right) \]
are the Fréchet derivatives of \( c(\cdot) \) at \( s \) and \( L_{\psi}^s \), respectively.\(^{19}\) Hence,
\[
|c \left( L_{\psi}^s \right) - c(\cdot)| \leq \max \left( \left| \int_{A(s)} \left[ h' \left( s(\theta) \right) - h' \left( p_1 \left( s \right) \right) \right] dG(\theta) \right|, \\
\left| \int_{A(s)} \left[ h' \left( L_{\psi}^s(\theta) \right) - h' \left( p_1 \left( L_{\psi}^s \right) \right) \right] \left( L_{\psi}^s(\theta) - s(\theta) \right) dG(\theta) \right| \right).
\]
Since \( h'(x) \) is increasing in \( x \), for all \( \theta \in A(s) \), both \( |h'(s(\theta)) - h'(p_1(s))| \) and \( |h' \left( L_{\psi}^s(\theta) \right) - h' \left( p_1 \left( L_{\psi}^s \right) \right)| \) are bounded above by
\[
K = \max \left( |h'(1 - \varepsilon) - h'(\xi)|, |h'(1 - \varepsilon) - h'(\varepsilon)| \right).
\]
Therefore,
\[
|c \left( L_{\psi}^s \right) - c(\cdot)| \leq \int_{A(s)} K \cdot |L_{\psi}^s(\theta) - s(\theta)| dG(\theta)
= K \cdot \left\| L_{\psi}^s, s \right\|.
\]
This concludes the proof. ■

**Proof of Proposition 10.**

**Proof.** Without loss of generality, we only need to consider \( \varepsilon \) sufficiently small. Let
\[
N_{\theta^*} = \left\{ s \in S_M : \Pr \left( \left| s(\theta) - 1_{(\theta \geq \theta^*)} \right| \leq \varepsilon \right) = 1 \right\}.
\]
Suppose the aggregate stochastic choice rule \( \pi \in N_{\theta^*} \). Since \( \varepsilon \) is sufficiently small, the cutoff \( \theta_{\pi} \) induced by \( \pi \) is \( \theta^* \), as defined in Assumption A4. That is, \( \pi(\pi(\theta)) > 0 \) for \( \theta > \theta^* \) and \( \pi(\pi(\theta)) < 0 \) for \( \theta < \theta^* \). It suffices to show that when \( \lambda \) is sufficiently small, any best response \( s^*_{i, \lambda} \) to \( \pi \) also belongs to \( N_{\theta^*} \).

Since \( \theta^* \in (\theta_{\min}, \theta_{\max}) \) and \( \varepsilon \) is sufficiently small, we have
\[
\inf \left( \{ \pi(1 - \varepsilon, \theta) : \theta > \theta^* \} \right) > 0
\]
and
\[
\sup \left( \{ \pi(\varepsilon, \theta) : \theta < \theta^* \} \right) < 0.
\]
Let
\[
b = \min \left\{ \inf \left( \{ \pi(1 - \varepsilon, \theta) : \theta > \theta^* \} \right), -\sup \left( \{ \pi(\varepsilon, \theta) : \theta < \theta^* \} \right) \right\}.
\]
Since the cost functional satisfies cheap perfect discrimination, we can choose \( \rho > 0 \) and

---

\(^{19}\)The Fréchet derivative of a functional is analogous to the gradient of a function with vector argument.
$K > 0$ such that

$$| c(L^\lambda, s) - c(s)| \leq K \cdot \|L^\lambda, s, s\|$$

for all $s \in B_{\rho}(1_{\theta \geq \psi})$. By Lemma 8, there exists a $\lambda_1 > 0$ such that $S_\lambda(\pi) \in B_{\rho}(1_{\theta \geq \psi})$ for all $\lambda \in (0, \lambda_1)$. Let $\lambda = \min \{\lambda_1, \frac{b}{L}\}$. We show that any $s \in S_M \setminus N^\pi$ is strictly dominated by $L^\lambda, s$ when $\lambda \in (0, \lambda)$. The benefit from choosing $L^\lambda, s$ instead of $s$ is

$$\begin{align*}
\int_{-\infty}^{\infty} \pi(\pi(\theta), \theta) \cdot [(L^\lambda, s)(\theta) - s(\theta)] g(\theta) d\theta - \lambda \cdot [c(L^\lambda, s) - c(s)] \\
\quad \geq \int_{-\infty}^{\theta^*} -b \cdot [(L^\lambda, s)(\theta) - s(\theta)] g(\theta) d\theta + \int_{\theta^*}^{\infty} b \cdot [(L^\lambda, s)(\theta) - s(\theta)] g(\theta) d\theta \\
\quad - \lambda \cdot [c(L^\lambda, s) - c(s)] \\
\quad = b \cdot \|L^\lambda, s, s\| - \lambda \cdot [c(L^\lambda, s) - c(s)] \\
\quad > (b - \lambda K) \cdot \|L^\lambda, s, s\| > 0,
\end{align*}$$

where the last inequality holds because $\lambda < \frac{b}{L}$ and $\|L^\lambda, s, s\| > 0$ by construction.

Hence, the best response to any $\pi \in N^\pi$ also belongs to $N^\pi$. By Helly’s selection theorem, $N^\pi$ is compact. Therefore, for any $\theta^* \in (\theta_{\min}, \theta_{\max})$ and $\varepsilon \in (0, 1/2)$, there is an equilibrium in $N^\pi$ when $\lambda$ is sufficiently small. This concludes the proof.

**Proof of Proposition 13.**

**Proof.** Let $\left\{ s^*_{i, \lambda} \right\}_{i \in \{0, 1\}}$ denote an equilibrium of the $\lambda$-game. By Lemma 12, any equilibrium strategy $s^*_{i, \lambda}$ is absolutely continuous. Then the desired result follows the same argument as in the proof of Proposition 7.

**Proof of Lemma 14.**

**Proof.** Let $s = \int s_i di$ denote the aggregate stochastic choice rule. It suffices to show that $s_i$ is player $i$’s best response to $s$ in the stochastic choice game. Suppose this is not true and there exists an $\bar{s}_i$ strictly better than $s_i$. Define $\varepsilon$ as

$$\varepsilon = \int_{-\infty}^{\infty} \pi(s(\theta), \theta) \cdot [\bar{s}_i(\theta) - s_i(\theta)] g(\theta) d\theta - \lambda \cdot [c(\bar{s}_i) - c(s_i)] .$$

Then $\varepsilon > 0$. By definition (9), there exists a $\tilde{q}_i \in Q(\bar{s}_i)$ and decision rule $\tilde{\sigma}_i$ such that

$$C(\tilde{q}_i) - c(\bar{s}_i) < \lambda^{-1} \cdot \varepsilon.$$
Then we have
\[
\int_{-\infty}^{\infty} \pi(s(\theta)\theta) \cdot [\tilde{s}_i(\theta) - s_i(\theta)] g(\theta) d\theta - \lambda \cdot [C(q_i) - c(s_i)] \\
\geq \int_{-\infty}^{\infty} \pi(s(\theta)\theta) \cdot [\tilde{s}_i(\theta) - s_i(\theta)] g(\theta) d\theta - \lambda \cdot [c(\tilde{s}_i) + \lambda^{-1} \cdot \varepsilon - c(s_i)] \\
= 0,
\]
so that \((q_i, \sigma_i)\) is strictly dominated by \((\tilde{q}_i, \tilde{\sigma}_i)\), contradicting the fact that \((q_i, \sigma_i)\) is an equilibrium strategy in the information acquisition game. ■

**Proof of Lemma 15.**

**Proof.** Let \(s = \int s_i d\theta\) denote the aggregate stochastic choice rule. It suffices to show that \((q_i, \sigma_i)\) is player \(i\)'s optimal strategy to \(s\) in the information acquisition game. Suppose this is not true and there exists a \((\tilde{q}_i, \tilde{\sigma}_i)\) strictly better than \((q_i, \sigma_i)\), i.e.,
\[
\int_{-\infty}^{\infty} \pi(s(\theta)\theta) \cdot [\tilde{s}_i(\theta) - s_i(\theta)] g(\theta) d\theta - \lambda \cdot [C(\tilde{q}_i) - C(q_i)] > 0,
\]
where \(\tilde{s}_i\) be the stochastic choice rule induced by \((\tilde{q}_i, \tilde{\sigma}_i)\). Since \((q_i, \sigma_i)\) is a minimum cost way of attaining \(s_i\) and by definition \(c(\tilde{s}_i) \leq C(\tilde{q}_i)\). The above inequality implies
\[
\int_{-\infty}^{\infty} \pi(s(\theta)\theta) \cdot [\tilde{s}_i(\theta) - s_i(\theta)] g(\theta) d\theta - \lambda \cdot [c(\tilde{s}_i) - c(s_i)] > 0,
\]
which contradicts the optimality of \(s_i\) as an equilibrium strategy in the stochastic choice game. ■

**Lemma 19** Let \(P\) be any probability measure over \(\mathbb{R}\). A set of functions \(S \subset L^1(\mathbb{R}, P)\) is relatively compact if \(S\) is uniformly bounded and equicontinuous.

**Proof.** Let \(B > 0\) be the uniform bound and \(\{s_n\}_{n=1}^{\infty} \subset S\) be a sequence of functions. For \(T \in \mathbb{N}\), let
\[
A_T = \left\{-T, -T + \frac{1}{T}, -T + \frac{2}{T}, \ldots, 0, \ldots, T - \frac{2}{T}, T - \frac{1}{T}, T\right\}.
\]
Then \(\bigcup_{T=1}^{\infty} A_T\) is dense in \(\mathbb{R}\). Since \(\bigcup_{T=1}^{\infty} A_T\) is countable, we can list its elements as \(\{\theta_1, \theta_2, \theta_3, \ldots\}\). Since the sequence \(\{s_n(\theta_1)\}_{n=1}^{\infty}\) is bounded, the Bolzano-Weierstrass theorem implies that it has a convergent subsequence, denoted by \(\{s_{1,n}(\theta_1)\}_{n=1}^{\infty}\). Since the sequence \(\{s_{1,n}(\theta_2)\}_{n=1}^{\infty}\) is also bounded, again it has a convergent subsequence \(\{s_{2,n}(\theta_2)\}_{n=1}^{\infty}\). Note that the sequence of functions \(\{s_{2,n}\}_{n=1}^{\infty}\) converges at both \(\theta_1\) and \(\theta_2\) as it is a subsequence of \(\{s_{1,n}\}_{n=1}^{\infty}\). Proceeding in this way allows us to obtain a countable collection of
subsequences of original sequence \( \{s_n\}_{n=1}^{\infty} \):

\[
\begin{array}{cccc}
s_{1,1} & s_{1,2} & s_{1,3} & \cdots \\
s_{2,1} & s_{2,2} & s_{2,3} & \cdots \\
s_{3,1} & s_{3,2} & s_{3,3} & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{array}
\]

where the sequence in the \( n \)-th row converges at the points \( \theta_1, \theta_2, \ldots, \theta_n \) and each row is a subsequence of the one above it. Hence the diagonal sequence \( \{s_{n,n}\}_{n=1}^{\infty} \) is a subsequence of the original sequence \( \{s_n\}_{n=1}^{\infty} \) that converges at each point of \( \bigcup_{T=1}^{\infty} A_T \). We next show that \( \{s_{n,n}\}_{n=1}^{\infty} \) is a Cauchy sequence in \( L^1(\mathbb{R}, P) \).

For any \( \varepsilon > 0 \), there exists a \( T_0 \) such that

\[
\Pr \left( \left[ -T_0, T_0 \right] \right) \geq 1 - \frac{\varepsilon}{5 \cdot B},
\]

where \( B \) is the uniform bound such that \(|s(\theta)| < B\) for all \( \theta \in \mathbb{R} \) and \( s \in S \). Since \( S \) is equicontinuous, there exists \( T_1 > T_0 \) such that \( \forall s \in S, \forall \theta_1, \theta_2 \in \Theta, \)

\[
| \theta_1 - \theta_2 | < \frac{1}{T_1}
\]

implies

\[
|s(\theta_1) - s(\theta_2)| < \frac{\varepsilon}{5}.
\]

Since \( A_{T_1} \) is a finite set and \( \{s_{n,n}\}_{n=1}^{\infty} \) converges at every point of \( A_{T_1} \), there exists \( n_0 \in \mathbb{N} \) such that

\[
|s_{n,n}(\theta) - s_{n',n'}(\theta)| < \frac{\varepsilon}{5}
\]

for all \( n,n' > n_0 \) and all \( \theta \in A_{T_1} \). For any \( y \in [-T_1, T_1] \), there exists \( \theta \in A_{T_1} \) such that

\[
|y - \theta| < \frac{1}{T_1},
\]

thus we have

\[
|s_{n,n}(y) - s_{n,n}(\theta)| < \frac{\varepsilon}{5}
\]

and

\[
|s_{n',n'}(y) - s_{n',n'}(\theta)| < \frac{\varepsilon}{5}.
\]
Hence for any \( y \in [-T_1, T_1] \) and \( n, n' > n_0 \),
\[
|s_{n', n'} (y) - s_{n, n} (y)| \\
\leq |s_{n', n'} (y) - s_{n', n'} (\theta) + |s_{n', n'} (\theta) - s_{n, n} (\theta)| + |s_{n, n} (\theta) - s_{n, n} (y)| \\
< \frac{3 \cdot \varepsilon}{5}.
\]
(24)

Then,
\[
\|s_{n', n'} - s_{n, n}\|_{L^1(\mathbb{R}, P)} \\
= \int_{\mathbb{R}} |s_{n', n'} (y) - s_{n, n} (y)| dP (y) \\
= \int_{[-T_1, T_1]} |s_{n', n'} (y) - s_{n, n} (y)| dP (y) + \int_{\mathbb{R} \setminus [-T_1, T_1]} |s_{n', n'} (y) - s_{n, n} (y)| dP (y) \\
< \frac{3 \cdot \varepsilon}{5} \cdot \Pr([-T_1, T_1]) + \int_{\mathbb{R} \setminus [-T_1, T_1]} 2 \cdot B \cdot dP (y) \\
\leq \frac{3 \cdot \varepsilon}{5} \cdot 1 + \frac{\varepsilon}{5} \cdot B \cdot 2 \cdot B \\
= \varepsilon,
\]
where the first inequality follows (24) and the second inequality follows (23).

Therefore, \( \{s_{n, n}\}_{n=1}^{\infty} \) is a Cauchy subsequence of \( \{s_{n}\}_{n=1}^{\infty} \) in \( L^1 (\mathbb{R}, P) \) and \( S \) is relatively compact in \( L^1 (\mathbb{R}, P) \). This concludes the proof. \( \blacksquare \)

**Proof of Proposition 16.**

**Proof.** We first show that \( k^\lambda \) goes to \( \infty \) as \( \lambda \) vanishes. We prove this claim by contradiction. If this is not true, there exists a \( K > 0 \) and a sequence \( \{\lambda_n\}_{n=1}^{\infty} \) that converges to zero, such that for each \( \lambda_n \), the information acquisition game has an equilibrium in which \( k^{\lambda_n} < K \).

Let \( h (x) = 1_{[-1/2 \leq x \leq 1/2]} \) denote the probability density function of the uniform distribution on \([-1/2, 1/2]\). Write \( g_k \) for the density function over signals induced by precision \( k \), i.e.,
\[
g_k (x) = \int_{\Theta} k \cdot h (k (x - \theta)) \cdot g (\theta) \cdot d\theta,
\]
and write \( l_k (\cdot|x) \) for the induced posterior density over \( \theta \):
\[
l_k (\theta|x) = \frac{k \cdot h (k (x - \theta)) \cdot g (\theta)}{g_k (x)}.
\]

Let
\[
\Sigma \triangleq \{\sigma \text{ Lebesgue measurable : } \forall x \in \mathbb{R}, \sigma (x) \in [0, 1]\}
\]
and

\[ S_K := \left\{ s \in S : \exists k \in [0, K] \text{ and } \sigma \in \Sigma, \text{ s.t. } s(\theta) = \int_x k \cdot h(k(x - \theta)) \cdot \sigma(x) \cdot dx \text{ for all } \theta \in \mathbb{R} \right\}. \]

\( S_K \) contains a player’s conjectures on the aggregate stochastic choice rule that can be generated from other players’ play when \( k \in [0, K] \).

Note that for any \( s \in S_K \)

\[ \left| \frac{ds(\theta)}{d\theta} \right| = k \cdot \left| \sigma\left(\theta + \frac{1}{2k}\right) - \sigma\left(\theta - \frac{1}{2k}\right) \right| \leq k < K, \]

which implies that \( S_K \) is equicontinuous. Since \( S_K \) is uniformly bounded by definition, it is relatively compact as an implication of Lemma 19.

If a player \( j \) chooses \((k_j, \sigma_j)\) against conjecture \( s \in S_K \), her expected utility is

\[ V_j(k_j, \sigma_j, s) = \int_{x_j} \sigma_j(x_j) \cdot \left[ \int_{\theta} \pi(s(\theta), \theta) \cdot l_{k_j}(\theta|x_j) \cdot d\theta \right] \cdot g_{k_j}(x_j) \cdot dx_j. \]

With an optimal choice of \( \sigma_j \) her expected utility becomes

\[ V_j^*(k_j, s) = \int_{x_j} \max \left\{ 0, \int_{\theta} \pi(s(\theta), \theta) \cdot l_{k_j}(\theta|x_j) \cdot d\theta \right\} \cdot g_{k_j}(x_j) \cdot dx_j. \]

(25)

Define

\[ V_j^{**}(s) := \lim_{k_j \to \infty} V_j^*(k_j, s) \]

\[ = \int_{x_j} \max \left\{ 0, \pi(s(x_j), x_j) \cdot g(x_j) \right\} \cdot dx_j \]

\[ = \int_{\theta} \max \left\{ 0, \pi(s(\theta), \theta) \right\} \cdot g(\theta) \cdot d\theta. \]

(26)

\( V_j^{**}(s) \) is player \( j \)'s ex ante expected utility against conjecture \( s \) if she can always observe the exact realization of the fundamental (i.e., if \( k_j \to \infty \)).

We next show that \( \forall s \in S, k_j > 0, V_j^{**}(s) > V_j^*(k_j, s) \). Since both dominance regions occur with positive probability under the common prior, we obtain

\[ \Pr(\pi(s(\theta), \theta) < 0) > 0 \]

48
and
\[ \Pr((s(\theta), \theta) > 0) > 0. \]

Then the convexity of function \( \max \{0, \cdot\} \) implies
\[
\max \left\{ 0, \int_\theta \pi(s(\theta), \theta) \cdot k_j \cdot h(k_j(x_j - \theta)) \cdot g(\theta) \, d\theta \right\}
\leq \int_\theta \max \{0, \pi(s(\theta), \theta) \cdot k_j \cdot h(k_j(x_j - \theta))\} \cdot g(\theta) \, d\theta
\]
\[
= \int_\theta \max \{0, \pi(s(\theta), \theta)\} \cdot k_j \cdot h(k_j(x_j - \theta)) \cdot g(\theta) \, d\theta.
\]
(27)

Since
\[
\Pr((s(\theta), \theta) \cdot k_j \cdot h(k_j(x_j - \theta)) > 0)
= \Pr((s(\theta), \theta) > 0)
> 0
\]
and
\[
\Pr((s(\theta), \theta) \cdot k_j \cdot h(k_j(x_j - \theta)) < 0)
= \Pr((s(\theta), \theta) < 0)
> 0
\]
for all \( x_j \in \mathbb{R} \), (27) holds strictly. Then, (25) implies
\[
V_j^*(k_j, s) = \int_{x_j} \max \left\{ 0, \int_\theta \pi(s(\theta), \theta) \cdot k_j \cdot h(k_j(x_j - \theta)) \cdot g(\theta) \, d\theta \right\} \cdot dx_j
\]
\[
< \int_{x_j} \int_\theta \max \{0, \pi(s(\theta), \theta)\} \cdot k_j \cdot h(k_j(x_j - \theta)) \cdot g(\theta) \, d\theta \cdot dx_j
\]
\[
= \int_\theta \max \{0, \pi(s(\theta), \theta)\} \cdot \int_{x_j} k_j \cdot h(k_j(x_j - \theta)) \cdot dx_j \cdot d\theta
\]
\[
= \int_\theta \max \{0, \pi(s(\theta), \theta) \cdot g(\theta)\} \cdot \int_{x_j} k_j \cdot h(k_j(x_j - \theta)) \cdot dx_j \cdot d\theta
\]
\[
= \int_\theta \max \{0, \pi(s(\theta), \theta) \cdot g(\theta)\} \cdot 1 \cdot d\theta
\]
\[
= V_j^{**}(s),
\]
where the last equality follows (26). Therefore,
\[
\forall s \in S, \forall k_j > 0, V_j^{**}(s) > V_j^*(k_j, s).
\]
(28)
We then show that for any $k_j > 0$, there exists a $k'_j > 0$ and $\delta > 0$, s.t. $\forall s \in S_K$,

$$V_j^* (k'_j, s) - V_j^* (k_j, s) > \delta .$$

If this is not true, there exists a $k_j > 0$ s.t. for any $k'_j > 0$ and $l \in \mathbb{N}$, there is an $s_{k_j, k'_j}^l \in S_K$ satisfying

$$V_j^* (k'_j, s_{k_j, k'_j}^l) - V_j^* (k_j, s_{k_j, k'_j}^l) \leq 1/l .$$

Since $S_K$ is relatively compact, $\forall k'_j > 0$, there exists a subsequence of $\{ s_{k_j, k'_j}^l \}_{l=1}^{\infty}$ converging to some $s_{k_j, k'_j} \in S$. To economize on notation, we will use $\{ s_{k_j, k'_j}^l \}_{l=1}^{\infty}$ to denote the subsequence. Since $V_j^* (k, s)$ is a continuous functional of $s$ for all $k > 0$, we obtain

$$V_j^* (k'_j, s_{k_j, k'_j}) - V_j^* (k_j, s_{k_j, k'_j}) = \lim_{l \to \infty} V_j^* (k'_j, s_{k_j, k'_j}^l) - V_j^* (k_j, s_{k_j, k'_j}^l) \leq 0 .$$

However, (28) implies that

$$V_j^* (k'_j, s_{k_j, k'_j}) - V_j^* (k_j, s_{k_j, k'_j}) > 0$$

for $k'_j$ large enough, which is a contradiction.

Now we are ready to prove $\lim_{n \to \infty} k_j (\lambda_n) = \infty$, where $k_j (\lambda_n)$ is player $j$’s optimal choice of precision at information cost $\lambda_n$. If this is not true, there exists a $\overline{k}_j > 0$ and a subsequence $\{ \lambda_m \}_{m=1}^{\infty} \subset \{ \lambda_n \}_{n=1}^{\infty}$ s.t. $\lim_{m \to \infty} k_j (\lambda_m) = \overline{k}_j$. Note that $V_j^* (k, s)$ is continuous in $k$, hence there exists a $k'_j > \overline{k}_j$ and $\delta > 0$ such that when $m$ is sufficiently large

$$V_j^* (k'_j, s) - V_j^* (k_j (\lambda_m), s) > \delta / 2$$

for all $s \in S_K$. Since $\lim_{m \to \infty} \lambda_m = 0$, we can choose $m$ large enough such that

$$\lambda_m < \frac{\delta}{2 \cdot [f (k'_j) - f (k_j (\lambda_m))]} .$$

Hence we have $\forall s \in S_K$,

$$V_j^* (k'_j, s) - \lambda_m \cdot f (k'_j) > V_j^* (k_j (\lambda_m), s) - \lambda_m \cdot f (k_j (\lambda_m)) ,$$

which contradicts the assumption that $k_j (\lambda_m)$ is player $j$’s equilibrium response in the information acquisition game with information cost $\lambda_m f (\cdot)$. Therefore we prove $\lim_{n \to \infty} k_j (\lambda_n) = \infty$. Since all players are infinitesimal, having the same preference and facing the same decision problem in any equilibrium, we obtain that $k^A$ goes to $\infty$ as $\lambda$ vanishes.

Finally, since $\lim_{k \to \infty} [f (\infty) - f (k)] \cdot k = \infty$, the information acquisition technology satisfies EPD so that in any decision problem it is never optimal to choose $k = \infty$. Therefore,
we conclude that $\lim_{\lambda \to 0} k^\lambda = \infty$ and $k^\lambda < \infty$ for all $\lambda > 0$. Then the desired limit uniqueness result follow the standard global game arguments. This concludes the proof. ■

**Proof of Proposition 17**

**Proof.** Since the information cost is Blackwell-consistent, the equilibrium information structure $q^\lambda$ is a binary information structure characterized by a stochastic choice rule $s^\lambda$, i.e.,

$$q^\lambda (x|\theta) = \begin{cases} s^\lambda (\theta), & \text{if } x = 1 \\ 1 - s^\lambda (\theta), & \text{if } x = 0 \end{cases}$$

where $x \in \{0, 1\}$ is the recommendation of the information structure. Since the information cost satisfies Laplacian selection, for any $\varepsilon > 0$, there exists a $\overline{\lambda} > 0$ such that for any $\lambda \in (0, \overline{\lambda}), \|s^\lambda, 1_{\{\theta \geq \theta^*\}}\| < \varepsilon$. Note that the four equilibria of the coordination game exist if $E [\pi (1, \theta) | x = 1] > 0, E [\pi (0, \theta) | x = 0] < 0, E [\pi (1, \theta) | x = 0] > 0$, and $E [\pi (0, \theta) | x = 1] < 0$ hold for $\lambda$ sufficiently small. We next verify these four inequalities.

Note that $E [\pi (1, \theta) | x = 1] > 0$ iff $\int \pi (1, \theta) s^\lambda (\theta) g (\theta) \, d\theta > 0$.

$$\left| \int \pi (1, \theta) s^\lambda (\theta) g (\theta) \, d\theta - \int \pi (1, \theta) 1_{\{\theta \geq \theta^*\}} g (\theta) \, d\theta \right|$$

$$= \left| \int \pi (1, \theta) \left[ s^\lambda (\theta) - 1_{\{\theta \geq \theta^*\}} \right] g (\theta) \, d\theta \right|$$

$$\leq \pi \cdot \|s^\lambda, 1_{\{\theta \geq \theta^*\}}\| < \pi \cdot \varepsilon ,$$

where $\pi > 0$ is the uniform upper bound of $|\pi (l, \theta)|$. For all $\theta > \theta^*$,

$$\pi (1, \theta) \geq \int_0^1 \pi (l, \theta) \, dl > 0 ,$$

where the first inequality follows Assumption A1 and the second inequality follows Assumption A5. Hence $\int \pi (1, \theta) 1_{\{\theta \geq \theta^*\}} g (\theta) \, d\theta > 0$, which implies $\int \pi (1, \theta) s^\lambda (\theta) g (\theta) \, d\theta > 0$ whenever $\varepsilon > 0$ is small enough. This proves $E [\pi (1, \theta) | x = 1] > 0$.

Note that $E [\pi (0, \theta) | x = 0] < 0$ iff $\int \pi (0, \theta) \left[ 1 - s^\lambda (\theta) \right] g (\theta) \, d\theta < 0$.

$$\left| \int \pi (0, \theta) \left[ 1 - s^\lambda (\theta) \right] g (\theta) \, d\theta - \int \pi (0, \theta) \left[ 1 - 1_{\{\theta \geq \theta^*\}} \right] g (\theta) \, d\theta \right|$$

$$= \left| \int \pi (0, \theta) \left[ 1_{\{\theta \geq \theta^*\}} - s^\lambda (\theta) \right] g (\theta) \, d\theta \right|$$

$$\leq \pi \cdot \|s^\lambda, 1_{\{\theta \geq \theta^*\}}\| < \pi \cdot \varepsilon .$$

For all $\theta < \theta^*$,

$$\pi (0, \theta) \leq \int_0^1 \pi (l, \theta) \, dl < 0 ,$$

where the first inequality follows Assumption A1 and the second inequality follows Assump-
tion A5. Hence \( \int \pi (0, \theta) \left[ 1 - 1_{\{\theta \geq \theta^*\}} \right] g(\theta) \, d\theta < 0 \), which implies \( \int \pi (0, \theta) \left[ 1 - s^\lambda (\theta) \right] g(\theta) \, d\theta < 0 \) whenever \( \varepsilon > 0 \) is small enough. This proves \( E [\pi (0, \theta) | x = 0] < 0 \).

Note that \( E [\pi (1, \theta) | x = 0] > 0 \) iff \( \int \pi (1, \theta) \left[ 1 - s^\lambda (\theta) \right] g(\theta) \, d\theta > 0 \).

\[
\begin{align*}
&= \left| \int \pi (1, \theta) \left[ 1 - 1_{\{\theta \geq \theta^*\}} \right] g(\theta) \, d\theta - \int \pi (1, \theta) \left[ 1 - 1_{\{\theta \geq \theta^*\}} \right] g(\theta) \, d\theta \right| \\
&\leq \pi \cdot \|s^\lambda, 1_{\{\theta \geq \theta^*\}}\| < \pi \cdot \varepsilon .
\end{align*}
\]

Since \( \int \pi (1, \theta) \left[ 1 - 1_{\{\theta \geq \theta^*\}} \right] g(\theta) \, d\theta > 0 \) by the assumption of the proposition, we obtain \( \int \pi (1, \theta) \left[ 1 - s^\lambda (\theta) \right] g(\theta) \, d\theta > 0 \) whenever \( \varepsilon > 0 \) is small enough. This proves \( E [\pi (1, \theta) | x = 0] > 0 \).

Note that \( E [\pi (0, \theta) | x = 1] < 0 \) iff \( \int \pi (0, \theta) s^\lambda (\theta) g(\theta) \, d\theta < 0 \).

\[
\begin{align*}
&= \left| \int \pi (0, \theta) s^\lambda (\theta) g(\theta) \, d\theta - \int \pi (0, \theta) 1_{\{\theta \geq \theta^*\}} g(\theta) \, d\theta \right| \\
&\leq \pi \cdot \|s^\lambda, 1_{\{\theta \geq \theta^*\}}\| < \pi \cdot \varepsilon .
\end{align*}
\]

Since \( \int \pi (0, \theta) 1_{\{\theta \geq \theta^*\}} g(\theta) \, d\theta < 0 \) by the assumption of the proposition, we obtain \( \int \pi (0, \theta) s^\lambda (\theta) g(\theta) \, d\theta < 0 \) whenever \( \varepsilon > 0 \) is small enough. This proves \( E [\pi (0, \theta) | x = 1] < 0 \).