Coordination and the Relative Cost of Distinguishing Nearby States*

Stephen Morris and Ming Yang

Princeton University and Duke University

April 2016

Abstract

We study a coordination game where players simultaneously acquire information prior to the play of the game. We allow general information acquisition technologies, modeled by a cost functional mapping from information structures. Costly local distinguishability is a property requiring that the cost of distinguishing nearby states is hard relative to distinguishing distant states. This property is not important in decision problems, but is crucial in determining equilibrium outcomes in games. If it holds, there is a unique equilibrium; if it fails, there are multiple equilibria close to those that would exist if there was complete information.

We study these issues in the context of a regime change game with a continuum of players. We also provide a common belief foundation for equilibria of this game. This allows us to distinguish cases where the players could (physically) acquire information giving rise to multiple equilibria, but choose not to, and situations where players could not physically have acquired information in a way consistent with multiple equilibria. Our analysis corresponds to the former case, while the choosing precision of additive noise corresponds to the latter case.

JEL: C72 D82

KEYWORDS: coordination, endogenous information acquisition, costly local distinguishability, higher order beliefs

1 Introduction

Situations where players must coordinate their actions are ubiquitous. Under complete information, the resulting coordination game will have multiple equilibria. In this paper we ask: how does players' endogenous choice of information about payoffs impact multiplicity?

1

We study this in a setting where payoffs depend on a parameter drawn from the real line. In this setting, it is natural to assume that it is *harder* to distinguish nearby states *relative* to distant states. Our results will depend on *how much* harder it is. We will show that if it is relatively hard to distinguish nearby states ("costly local distinguishability"), then there is a unique equilibrium when there is endogenous information acquisition. If it is as easy to distinguish nearby states as distant states - or at least not too much harder - then multiple equilibria will persist.

The intuition for these results is as follows. Given strategic complementarities in action choices, players will always have an incentive to acquire the same information. If information is costly, they will not acquire complete information. Now suppose that in equilibrium, all players care about is whether the state is above or below some threshold value. Without costly local distinguishability, it will be an equilibrium for all players to focus on whether the true state is above or below that threshold. As a result, there will be many equilibria corresponding to different thresholds. But under costly local distinguishability, players will choose to have inaccurate signals in the neighborhood of the threshold state. Then perfect coordination in information acquisition is not an equilibrium. There will then be a contraction in best responses in information acquisition: for any given threshold which other players focus most attention on, a player will have an incentive to choose a threshold closer to a particular focal point, and so there is a contraction and a unique equilibrium.

Our results offer a novel perspective on recent work on endogenous information acquisition in coordination games. Szkup and Trevino (2015) and Yang (2015) have considered the case where players can (simultaneously) choose the precision of noisy signals about the state, with the cost increasing with precision. In this case, a low cost of information will imply that players will acquire signals with high precision. Carlsson and Damme (1993) have shown that in such "global game" environments, a unique equilibrium must then be played. However, these results rely heavily on the *inflexibility* of information acquisition: there is a one dimensional class of possible information structures, parameterized by the precision of private information. Players then have no ability to choose the qualitative properties of information they acquire and, in particular, where to focus their attention. Yang (2015) considered instead *flexible* information acquisition, where players can acquire any information. Yang (2015) used entropy reduction as a cost function for information. He showed that there are multiple equilibria. However, the entropy reduction cost function has the distinctive feature that it is equally easy to distinguish nearby and distant states (so costly local distinguishability fails). This paper incorporates cost functionals which are both flexible - allowing any information structure to be feasible - but allowing the cost of distinguishing states to be sensitive to the distance between states. Both uniqueness and multiplicity are consistent with flexible information acquisition in our setting. We characterize which case arises, depending on qualitative aspects of the cost of information.

Our results have implications for modelling information acquisition more broadly. 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. But 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 contexts, this paper contains a warning about use of entropy as a cost of information in strategic settings.

Our main model will be a *regime change* game, studied by Morris and Shin (1998) and extensively explored in the later literature. A continuum of players decide whether to invest or not invest in a project. There is a fixed cost of investing. There is a fixed benefit to investing realized only if the proportion of players investing is above a critical level, which is a decreasing function of the state. Before playing the game, each player can acquire any information about the underlying state. To model an environment with a low cost of information acquisition, we first fix a cost functional on information structures. We then ask what happens under endogenous information acquisition if we multiply the cost functional by a constant that we take to zero.

A player in this game chooses what information structure to acquire, and then what action to take as a function of his signal. These choices will imply an *effective strategy* for the player: for each state of the world, there will be an induced probability distribution over actions. In analyzing the game, one can restrict attention to effective strategies, abstract from the information structures that gave rise to the effective strategies, and restrict attention to cost functionals defined on effective strategies. In the binary action context studied here, an effective strategy then maps states of the world to a probability of investing.

Our main result links three properties of cost functionals. Consider a simple *threshold* decision problem, where a player picks an action (*invest* or *not invest*) and gets a payoff of 0 if she does not invest, a fixed positive payoff if she invests and the state is above a threshold and a fixed negative payoff if she invests and the state below that threshold. An information cost functional satisfies *continuous choice* if the optimal effective strategy in such environments is continuous. As we will discuss below, continuous choice captures the idea that it is relatively costly to distinguish nearby states. An information cost functional satisfies *translation insensitivity* if a small translation of an effective strategy results in a small change in cost. Translation insensitivity captures the idea that changing where attention is paid (holding the amount of attention fixed) does not change the cost too much. A cost functional satisfies

3

limit uniqueness if, as the cost goes to zero, there is a unique equilibrium played. If a cost functional satisfies continuous choice and translation insensitivity, then it satisfies limit uniqueness. In addition, every cost functional satisfying continuous choice and translation insensitivity has the same limit equilibrium which is the *Laplacian selection* - a many player version of risk dominance that is played in the corresponding exogenous information global game.

We also give a primitive property on cost functionals that is sufficient for continuous choice. A cost functional satisfies *costly local distinguishability* if it is sufficiently harder to distinguish nearby states than distant states: specifically, it requires that for any discontinuous effective strategy, the cost saving from acquiring a nearby continuous strategy is high compared to the distance between the two strategies. A high cost saving from choosing a continuous strategy corresponds to a high cost in distinguishing nearby states.

There is a partial converse to our main results. A cost functional is *Lipschitz* if the difference in costs between two effective strategies is of the order of the distance between the two effective strategies. This condition implies that the cost impact of changing the effective strategy on only a small set of states is small, even if the changes are large. This implies a failure of costly local distinguishability.

We also show that the limit uniqueness/multiplicity results extend beyond regime change games and describe local versions of results that hold under local versions of the sufficient conditions used in our main statements.

Our results can be understood using the properties of higher-order beliefs. For any given exogenous information structure, we characterize how equilibria depend on the implied higher-order beliefs.[1] For any given event (a set of states of the world), we can identify the event where a critical proportion of players assign at least a critical probability to the given event. If an event is a fixed point of this operator, we say that it is an equilibrium regime change event since there is an equilibrium in which the regime changes at states in that event. Under a global game information structure - where players observe the state plus some conditionally independent noise - all feasible choices imply a unique equilibrium regime change event. Thus there is a unique equilibrium independent of how information is chosen, endogenously or exogenously. We will study cost functionals where every information structure has finite cost, so it is always feasible to choose alternative information structure profiles giving rise to different equilibria. However, we show that under costly local distinguishability (and translation insensitivity), endogenous information choice implies that players will choose information structures giving rise to a unique equilibrium regime change event (even though it was feasible to choose information structures that would give rise to multiplicity).

We proceed as follows. Section 2 sets up the model. Section 3 presents a leading

---

[1]This exercise is analogous to that in Morris and Yildiz (2016); we will discuss the differences in the body of the paper.

example to illustrate the local distinguishability and its impact on equilibrium outcomes. Section 4 establishes the conditions for limit uniqueness and multiplicity under general information cost functionals. Section 5 provides an intuition building on understanding of higher order beliefs under endogenous information acquisition. Section 6 extends the results to coordination games with general payoffs under weaker conditions on the information cost functional. Section 7 discusses the relation between our general information cost and the entropy-based information cost that is widely used in similar settings, allowing players to observe others' actions and evidence on costly local distinguishability. Long proofs are relegated to the appendix.

2   The Model

We will define our game as follows. A continuum of players choose an action, "not invest" or "invest". We normalize the payoff from not investing to 0. A player's payoff if she invests is $\pi(l, \theta)$, where l is the proportion of players investing and θ is a payoff relevant state. We assume for now that we have "regime change" payoffs[2] with

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

where $t \in (0, 1)$ and $\beta : \mathbb{R} \to \mathbb{R}$ is a continuous and strictly decreasing function. Thus t can be interpreted as the cost of investment and $\beta(\theta)$ is the critical proportion of other players at which investing gives a return of 1. We assume that $\beta(\theta) > 1$ for small enough θ and $\beta(\theta) < 0$ for large enough θ. Without loss of generality, we let $\beta(0) = 1$ and $\beta(1) = 0$.[3] The regime changes whenever at least proportion $\beta(\theta)$ of players invest. Actions are strategic complements: players are (weakly) more willing to take an action if they expect others to take that action. Players do not know the payoff relevant state θ but do share a common prior on θ, denoted by density p. A maintained assumption is that p is continuous and strictly positive on $[0, 1]$.

Before selecting an action, players can simultaneously and privately acquire information about θ. Player i's information structure is a pair (X_i, q_i), where $X_i \subset \mathbb{R}$ is the set of realizations of player i's signal and $q_i(\cdot|\theta) \in \Delta(X_i)$ is the probability measure on X_i conditional on θ. Players' signals are conditionally independent. Information acquisition is costly. Let Q denote the space of all information structures and write $C : Q \to [0, \bar{c}]$ for the cost functional; thus we maintain for now the assumption that there is a uniform upper bound on the cost of information. This implies that no effective strategy is technologically

---

[2] In Section 6, we will extend our analysis to more general payoffs and relax a number of technical assumptions in the main statements of our results.

[3] In the literature of regime change games, it is often assumed that $\beta(\theta) = 1 - \theta$. It is convenient for us to work with more general β functions but the extra generality is not of intrinsic interest.

precluded.[4] A player incurs a cost $\lambda \cdot C\left((X, q)\right)$ if she chooses information structure $(X, q) \in Q$. We will hold the cost functional C fixed in our analysis and vary $\lambda \geq 0$, a parameter that represents the difficulty of information acquisition; we will refer to the resulting game as the λ-game. If $\lambda = 0$, the players can choose to observe θ at no cost and the model reduces to a complete information game. We will perturb this complete information game by letting λ be strictly positive but close to zero. Focussing on small but positive λ sharpens the statement and intuition of our results.

A player's strategy corresponds to an information structure (X, q) together with an action rule $\sigma : X \to [0, 1]$, with $\sigma(x)$ being the probability of investing upon signal realization x. The information structure and the action rule jointly determine the player's *effective strategy*, which is a function $s : \mathbb{R} \to [0, 1]$, with

$$s\left(\theta\right) = \int q\left(x|\theta\right) \cdot \sigma\left(x\right) dx \ .$$

That is, $s\left(\theta\right)$ is the player's probability of investing conditional on the state being θ.[5] We call s the *effective strategy* since it describes action choices integrating out signal realizations. Any effective strategy s can be viewed as arising from a binary information structure where a player's signal is an action recommendation. Formally, we can identify an effective strategy with the information structure given by $X = \{0, 1\}$ and

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

and the action rule given by

$$\sigma\left(x\right) = \begin{cases} 1, \text{ if } x = 1 \\ 0, \text{ if } x = 0 \end{cases} \ .$$

Players care only about the effective strategies of opponents, and not the information choice and strategy generating them. Moreover, if the cost functional is weakly increasing in the information content, as ordered by the Blackwell (1953), then each player will weakly prefer to acquire a two signal information structure corresponding to an effective strategy s. This observation corresponds to formal arguments in Woodford (2008) and Yang (2015). Thus we will identify information structures with effective strategies, unless otherwise stated. We will write S for the set of effective strategies and $c : S \to [0, \bar{c}]$ for the cost functional restricted to effective strategies.

We equip the space of effective strategies with the L^1-metric, so that the distance between

---

[4]This assumption is not necessary but helps simplify the derivation. We will relax it in Section 6.

[5]Here, the signal realization x could be continuous or discrete. To economize on notation, we use "$\int$" to refer to both the integration over a continuum of signal realizations and the summation over discrete realizations.

effective strategies s_1 and s_2 is given by

$$\|s_1, s_2\| = \int_{\mathbb{R}} |s_1(\theta) - s_2(\theta)|\, p(\theta)\, d\theta;$$

and write $B_\delta(s)$ for the set of effective strategies within δ of s under this metric.

Now a player's ex ante payoff - if she chooses effective strategy s_i and the profile of others' effective strategies is $\{s_j\}_{j \in [0,1]}$ - is

$$u_i\left(s_i, \{s_j\}_{j \in [0,1]}\right) = \int_\theta s_i(\theta) \left(1_{\left\{\int s_j(\theta) dj \geq \beta(\theta)\right\}} - t\right) p(\theta)\, d\theta.$$

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

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

for each i.

We will later restrict attention to *monotonic* (non-decreasing) effective strategies $s :$ $\mathbb{R} \to [0,1]$. This is with loss of generality but is consistent with many applications (e.g., the effective strategy is always monotone in global game models) and allows us to highlight key insights. We write S_M for the set of monotonic effective strategies.

Definition 2 (Monotone Nash Equilibrium) $\{s_j\}_{j \in [0,1]}$ *is a monotone Nash equilibrium if it is a Nash equilibrium and each s_i is monotone.*

We will collect together in Section 6 a discussion of how maintained assumptions in the body of the text can be relaxed.

3   The Leading Example

For our leading example, we assume that $\beta(\theta) = 1 - \theta$ and that the players' common prior p is the uniform distribution over $[-A, 1 + A]$ for some $A > 0$; we focus on monotone Nash equilibria; and we consider the cost functional

$$c(s) = \max\left(0, 1 - \left(\sup_\theta\ s'(\theta)\right)^{-\gamma}\right) \tag{1}$$

with $\gamma > 0$. If s is discontinuous, then $s'(\theta)$ is understood to be infinity and so the cost of any discontinuous s is 1.[6]

---

[6]If $s(\theta)$ is not differentiable at θ, we can take it to equal the maximum of the left and right derivatives.

This cost functional highlights the role of distinguishing nearby states, in particular building in the property that it is more costly to distinguish the states that are closer to each other. To see this, fix the values of $s(\theta_1)$ and $s(\theta_2)$ for two states $\theta_1 \neq \theta_2$. The difference $|s(\theta_2) - s(\theta_1)|$ reflects to what extent these two states are distinguished from each other. The closer the two states, the steeper is the slope $\frac{|s(\theta_2) - s(\theta_1)|}{|\theta_2 - \theta_1|}$ and thus the higher is the cost.

Under this cost functional, the cost of a discontinuous effective strategy is set equal to 1. Choosing a continuous effective strategy lowers the cost and with a cost saving of $k^{-\gamma}$ if the maximum derivative of s is k (and not too small). Observe that higher γ results in lower marginal cost saving from reducing k, which makes the players more willing to choose steeper or even discontinuous effective strategies.

3.1 Characterizing the Equilibria

A monotone Nash equilibrium is characterized by a critical threshold θ^* above which the regime changes. In this case, each player i chooses an effective strategy s_i to maximize

$$U(s_i | \theta^*, \lambda) = \frac{1}{2 \cdot A + 1} \left[\int s_i(\theta) \cdot \left(1_{\{\theta \geq \theta^*\}} - t \right) \cdot d\theta \right] - \lambda \cdot c(s_i). \tag{2}$$

If there was no cost of information (i.e., $\lambda = 0$), a player's optimal effective strategy is the step function $1_{\{\theta \geq \theta^*\}}$, which perfectly distinguishes the threshold event $[\theta^*, 1 + A]$ from its complement, $[-A, \theta^*)$. For small but positive λ, since the information cost is determined by the maximal slope of s, any optimal effective strategy of player i will take the form

$$s_{\widehat{\theta}, k}(\theta) = \begin{cases} 0, \text{ if } \theta \leq \widehat{\theta} - \frac{1}{2k} \\ \frac{1}{2} + k\left(\theta - \widehat{\theta}\right), \text{ if } \widehat{\theta} - \frac{1}{2k} \leq \theta \leq \widehat{\theta} + \frac{1}{2k} \\ 1, \text{ if } \theta \geq \widehat{\theta} + \frac{1}{2k} \end{cases} \tag{3}$$

for some $\widehat{\theta}$ close to θ^* and $k \in \mathbb{R}_+ \cup \{\infty\}$.[7] Here k is the maximal slope and $\widehat{\theta}$ determines the position of the effective strategy. In this notation, $s_{\widehat{\theta}, \infty}$ is the step function $1_{\{\theta \geq \widehat{\theta}\}}$. Note that the information cost $c\left(s_{\widehat{\theta}, k}\right)$ is invariant to translation. That is, $c\left(s_{\widehat{\theta}, k}\right)$ does not depend on the position $\widehat{\theta}$. So we can first solve for $\widehat{\theta}$, taking k as given. Now

$$\frac{1}{2 \cdot A + 1} \int s_{\widehat{\theta}, k}(\theta) \cdot \left(1_{\{\theta \geq \theta^*\}} - t \right) \cdot d\theta$$

---

[7]The analysis that follows will be based on the assumption that $\left[\widehat{\theta} - \frac{1}{2k}, \widehat{\theta} + \frac{1}{2k}\right] \subset [-A, 1 + A]$. A sufficient condition for this will turn out to be that λ is sufficiently small, in particular with $\lambda < \frac{t(1-t)}{2(2A+1)}$.

does depend on $\widehat{\theta}$. Simple calculation shows that the above expression is maximized - for any given k- when the position $\widehat{\theta}$ is set equal to

$$\widehat{\theta}(\theta^*) = \theta^* + k^{-1}(t - 1/2). \tag{4}$$

By substituting (3) and (4) into (2), we can show that the optimal slope k^* is

$$k^* = \begin{cases} \left[\frac{t(1-t)}{2\lambda\gamma(2A+1)}\right]^{\frac{1}{1-\gamma}} & \text{if } \gamma < 1 \\ \infty & \text{if } \gamma \geq 1 \end{cases}. \tag{5}$$

Note that when $\gamma \geq 1$, the optimal slope is $k^* = \infty$ and the optimal position $\widehat{\theta}(\theta^*) = \theta^*$. Hence the optimal effective strategy $s_{\widehat{\theta}(\theta^*),k^*}$ becomes the step function $1_{\{\theta \geq \theta^*\}}$, which is discontinuous at the threshold θ^*. This is because the cost saving $k^{-\gamma}$ from replacing the step function by a continuous effective strategy with slope $k < \infty$ is too small to compensate the sacrificed benefit. As a result, the player chooses the step function that sharply distinguishes states below θ^* from those above it. When $\gamma < 1$, the cost saving is large enough and the optimal effective strategy becomes continuous. It is worth noting that whether the optimal effective strategy is continuous or discontinuous does not depend on λ, which controls the overall difficulty of information acquisition. The (dis)continuity is purely determined by γ and hence is a property of the information cost functional.

Since the optimal effective strategy is unique and all players are facing the same decision problem in equilibrium, this game has only symmetric equilibria. Assuming the continuum law of large numbers, $s_{\widehat{\theta}(\theta^*),k^*}$ will also describe the proportion of players investing as a function of the state. In order to be an equilibrium, the threshold of regime change must coincide with θ^*. That is,

$$s_{\widehat{\theta}(\theta^*),k^*}(\theta) \begin{cases} > 1 - \theta & \text{if } \theta > \theta^* \\ < 1 - \theta & \text{if } \theta < \theta^* \end{cases}.$$

To search for the equilibria, we first consider the case of $\gamma \geq 1$. For any threshold $\theta^* \in [0,1]$, the optimal effective strategy (by (4) and (5)) is

$$s_{\theta^*,\infty} = 1_{\{\theta \geq \theta^*\}}, \tag{6}$$

which is discontinuous at $\theta = \theta^*$. The induced threshold does coincide with θ^*. As a result, any threshold $\theta^* \in [0,1]$ is an equilibrium threshold and the corresponding strategy in equilibrium is $s_i = 1_{\{\theta \geq \theta^*\}}$ for all players $i \in [0,1]$. If $\gamma < 1$, for any threshold $\theta^* \in [0,1]$, the optimal effective strategy (by (4) and (5)) is continuous and the induced threshold $\widetilde{\theta}(\theta^*)$ satisfies

$$\frac{1}{2} + k^*\left[\widetilde{\theta}(\theta^*) - \widehat{\theta}(\theta^*)\right] = 1 - \widetilde{\theta}(\theta^*) \ .$$

Together with (4), the above equation leads to

$$\widetilde{\theta}\left(\theta^{*}\right)=\frac{k^{*}\cdot\theta^{*}+t}{1+k^{*}}. \tag{7}$$

Since $k^{*}<\infty$, $\widetilde{\theta}:[0,1]\rightarrow[0,1]$ defined by equation (7) is a contraction and thus has a unique fixed point,

$$\theta^{*}=t, \tag{8}$$

which uniquely characterizes the threshold of regime change in equilibrium. The corresponding effective strategy in equilibrium is

$$s_{\widehat{\theta},k^{*}}\left(\theta\right)=\left\{\begin{array}{l} 0,\text{ if }\theta\leq\widehat{\theta}-\frac{1}{2k^{*}}\\ \frac{1}{2}+k^{*}\left(\theta-\widehat{\theta}\right),\text{ if }\widehat{\theta}-\frac{1}{2k^{*}}\leq\theta\leq\widehat{\theta}+\frac{1}{2k^{*}}\\ 1,\text{ if }\theta\geq\widehat{\theta}+\frac{1}{2k^{*}}\end{array}\right. \tag{9}$$

with

$$\widehat{\theta}=t+\frac{t-1/2}{k^{*}} \tag{10}$$

and

$$k^{*}=\left[\frac{t\left(1-t\right)}{2\lambda\gamma\left(2A+1\right)}\right]^{\frac{1}{1-\gamma}}. \tag{11}$$

We summarize the above results in the following proposition.

Proposition 3 *Let the information cost parameter $\lambda>0$ be small enough. When $\gamma<1$, there exists a unique equilibrium in which each player takes the effective strategy characterized by (9), (10) and (11) and there is regime change if and only if $\theta\geq t$. When $\gamma\geq 1$, for any $\theta^{*}\in[0,1]$, there exists an equilibrium where each player takes the effective strategy given by (6) and there is regime change if $\theta\geq\theta^{*}$.*

This proposition clearly relates equilibrium uniqueness and multiplicity to the local distinguishability of the information cost. The intuition is as follows. The players have a motive to coordinate their actions, which also induces a motive to coordinate their information acquisition. In the case of cheap local distinguishability (i.e., $\gamma\geq 1$), it is easy to distinguish any event $[\theta^{*},1+A]$ from its complement. This facilitates the players' coordination in information acquisition in the sense that the players can coordinate to identify any event $[\theta^{*},1+A]$ with threshold $\theta^{*}\in[0,1]$, resulting in multiple equilibria. In contrast, when the information cost exhibits costly local distinguishability (i.e., $\gamma<1$), the players are not willing to sharply distinguish any event $[\theta^{*},1+A]$ from its complement. This weakens the players' ability and incentive to coordinate their information acquisition and leads to the unique equilibrium. It is also worth noting that this unique equilibrium is exactly the same as the Laplacian selection in the global games literature. In order to fully appreciate the

10

economic insight behind the result, Section 5 further elaborates on the relation between the uniqueness and local distinguishability of the information cost in terms of players' higher order beliefs.

3.2 Two Interpretations of the Leading Example

This subsection provides two economic interpretations of the effective strategy given by (3), (4) and (5), in order to connect our model to the global game literature. The first interpretation is straightforward. Fixing the slope $k > 0$, the effective strategy $s_{\widehat{\theta}(\theta^*),k}$ can be simply viewed as a binary information structure with $X = \{0,1\}$ and

$$q(x|\theta) = \begin{cases} s_{\widehat{\theta}(\theta^*),k}(\theta) & \text{if } x = 1 \\ 1 - s_{\widehat{\theta}(\theta^*),k}(\theta) & \text{if } x = 0 \end{cases},$$

as well as an action rule $\sigma(x) = x$.

Alternatively, suppose each player receives a signal $z = \theta + k^{-1} \cdot \varepsilon$, where $\varepsilon \sim Uniform_{[-1/2,1/2]}$ is independent from θ, and k measures the precision of the signal. The signals are conditionally independent across different players. This information structure is that assumed in the global games literature (in the special case of uniform noise). Each player purchases her own precision k at the price $\lambda \cdot \max(0, 1 - k^{-\gamma})$. For any given precision $k > 0$, facing the decision problem (2), simple calculations show that each player follows a switching strategy

$$\sigma(z) = \begin{cases} 1 & \text{if } z \geq \widehat{z}(\theta^*) \\ 0 & \text{if } z < \widehat{z}(\theta^*) \end{cases},$$

where $\widehat{z}(\theta^*) = \theta^* + k^{-1}(t - 1/2)$. This leads to

$$\Pr(\sigma(z) = 1| \text{ state} = \theta) = \begin{cases} 0, & \text{if } \theta \leq \widehat{\theta}(\theta^*) - \frac{1}{2k} \\ \frac{1}{2} + k\left(\theta - \widehat{\theta}(\theta^*)\right), & \text{if } \widehat{\theta}(\theta^*) - \frac{1}{2k} \leq \theta \leq \widehat{\theta}(\theta^*) + \frac{1}{2k} \\ 1, & \text{if } \theta \geq \widehat{\theta}(\theta^*) + \frac{1}{2k} \end{cases},$$

which is exactly the effective strategy $s_{\widehat{\theta}(\theta^*),k}$ characterized by (3), (4) and (5). It is worth noting that the information cost is $\lambda \cdot \max(0, 1 - k^{-\gamma})$ under both interpretations. This explains why the players will choose the same level of k, no matter whether it stands for the slope in the first interpretation or the precision in the second one. The resulting effective strategies and equilibrium outcomes are the same. In Subsection 5.2, however, we show that the underlying mechanisms are very different.

4 Main Results

We will be focusing on the properties of monotone Nash equilibria. In particular, we identify
the properties of the information cost functionals that lead to a unique or multiple equilibria
when λ is close to zero.

4.1 Threshold Decision Problems

A key ingredient of the analysis will be the choice in a simple class of "threshold decision
problems". Suppose a player must choose an effective strategy s when the cost to investing
is t, there is a payoff 1 if she invests and the state is at least θ^* and the information cost of
effective strategy is $\lambda \cdot c(s)$. Thus the payoff of effective strategy s is

$$U(s|\theta^*, \lambda) = \left[\int s(\theta) \left(1_{\{\theta \geq \theta^*\}} - t \right) p(\theta) d\theta \right] - \lambda \cdot c(s) .$$

This decision problem is parameterized by θ^* and λ, and we will refer to it as the (θ^*, λ)-
decision problem. We write $S_M(\theta^*, \lambda)$ for the set of optimal monotonic effective strategies
in the (θ^*, λ)-decision problem, i.e.,

$$S_M(\theta^*, \lambda) = \arg\max_{s \in S_M} U(s|\theta^*, \lambda) .$$

We first show that it is optimal for players to choose strategies that are close to a step
function at θ^* when the cost of information is small.

Lemma 4 (Optimal Effective Strategies in the Threshold Decision Problems) *The
essentially unique monotonic optimal effective strategy if $\lambda = 0$ is a step function at θ^*, i.e.,*

$$S_M(\theta^*, 0) = \left\{ 1_{\{\theta \geq \theta^*\}} \right\} .$$

*For any $\delta > 0$, there exists $\overline{\lambda} > 0$ such that $S_M(\theta^*, \lambda) \subset B_\delta \left(1_{\{\theta \geq \theta^*\}} \right)$ for all $\theta^* \in [0, 1]$ and
$\lambda \leq \overline{\lambda}$.*

Proof. When $\lambda = 0$, it is straightforward to see that the player chooses $s(\theta) = 1$ if
$1_{\{\theta \geq \theta^*\}} - t > 0$ and $s(\theta) = 0$ if $1_{\{\theta \geq \theta^*\}} - t < 0$. Hence, $S(\theta^*, 0) = \left\{ 1_{\{\theta \geq \theta^*\}} \right\}$.

Now consider the case of $\lambda > 0$. For any $s_{\theta^*, \lambda} \in S_M(\theta^*, \lambda)$ and $s_{\theta^*, \lambda} \neq 1_{\{\theta \geq \theta^*\}}$, the
optimality of $s_{\theta^*, \lambda}$ implies

$$\int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*\}} - t \right] \cdot \left[1_{\{\theta \geq \theta^*\}} - s_{\theta^*, \lambda}(\theta) \right] p(\theta) d\theta \quad < \quad \lambda \cdot \left[c \left(1_{\{\theta \geq \theta^*\}} \right) - c(s_{\theta^*, \lambda}) \right]$$

$$\leq \quad \lambda \cdot \overline{c}.$$

Note that

$$\int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*\}} - t \right] \cdot \left[1_{\{\theta \geq \theta^*\}} - s_{\theta^*, \lambda}(\theta) \right] p(\theta) \, d\theta$$

$$= \quad t \cdot \int_{-\infty}^{\theta^*} s_{\theta^*, \lambda}(\theta) \, p(\theta) \, d\theta + (1-t) \cdot \int_{\theta^*}^{\infty} \left[1 - s_{\theta^*, \lambda}(\theta) \right] p(\theta) \, d\theta$$

$$\geq \quad \min(t, 1-t) \cdot \left[\int_{-\infty}^{\theta^*} s_{\theta^*, \lambda}(\theta) \, p(\theta) \, d\theta + \int_{\theta^*}^{\infty} \left[1 - s_{\theta^*, \lambda}(\theta) \right] p(\theta) \, d\theta \right]$$

$$= \quad \min(t, 1-t) \cdot \left\| 1_{\{\theta \geq \theta^*\}}, s_{\theta^*, \lambda} \right\|.$$

The above two inequalities imply

$$\left\| 1_{\{\theta \geq \theta^*\}}, s_{\theta^*, \lambda} \right\| < \frac{\lambda \cdot \overline{c}}{\min(t, 1-t)} \ . \tag{12}$$

Hence for any $\delta > 0$, $\left\| 1_{\{\theta \geq \theta^*\}}, s_{\theta^*, \lambda} \right\| < \delta$ if $\lambda < \overline{\lambda} = \frac{\delta \cdot \min(t, 1-t)}{\overline{c}}$. ∎

The fact that the decision maker's optimal effective strategies approximate $1_{\{\theta \geq \theta^*\}}$ as $\lambda \to 0$ reflects her motive to sharply identify event $\{\theta \geq \theta^*\}$ from its complement. In a decision problem, whether this is achieved by a continuous or discontinuous $s_{\theta^*, \lambda} \in S_M(\theta^*, \lambda)$ is not important, since the loss caused by deviating from $1_{\{\theta \geq \theta^*\}}$ is of the order of magnitude of $\left\| 1_{\{\theta \geq \theta^*\}}, s_{\theta^*, \lambda} \right\|$. In contrast, in the game considered here, as we will see, the continuity of $s_{\theta^*, \lambda}$ is critical in determining the equilibrium outcomes. In particular, we will be interested in when the optimal effective strategies are always absolutely continuous whenever $\lambda > 0$.[8]

Definition 5 (Continuous Choice) *Cost functional $c(\cdot)$ satisfies continuous choice if all optimal strategies are absolutely continuous, i.e., $S_M(\theta^*, \lambda)$ consists only of absolutely continuous functions, for all $\theta^* \in [0, 1]$ and $\lambda \in \mathbb{R}_{++}$.*

Before proceeding to its equilibrium implications, we introduce several concepts to reveal the economic meaning of the continuous choice property.

A *continuous approximation* of an effective strategy s that is not absolutely continuous is a sequence of absolutely continuous effective strategies $\{s^n\}_{n=1}^{\infty}$ with

$$\lim_{n \to \infty} \| s, s^n \| = 0.$$

---

[8] We conjecture that continuity should be sufficient for the result, but currently use absolute continuity in proving the result.

The approximation is *cheap* if, first, $c(s^n) < c(s)$ for all n; and, second,

$$\lim_{n \to \infty} \frac{c(s) - c(s^n)}{\|s, s^n\|} = \infty .$$

That is, choosing s^n instead of s, the cost saving $c(s) - c(s^n)$ relative to the degree of approximation can be arbitrarily large.

Definition 6 (costly local distinguishability) *Cost functional $c(\cdot)$ satisfies costly local distinguishability if every effective strategy that is not absolutely continuous has a cheap continuous approximation.*

The *costly local distinguishability* captures the idea that it is sufficiently harder to distinguish nearby states than distant states. This is because sharply distinguishing nearby states, e.g., the states just above a threshold and those just below it, requires the effective strategy to jump at the threshold. Hence, a high cost saving from choosing a continuous strategy corresponds to a high cost in distinguishing nearby states.

Lemma 7 (continuous choice) *If $c(\cdot)$ satisfies costly local distinguishability, then $c(\cdot)$ satisfies continuous choice.*

Proof. Suppose $s_{\theta^*, \lambda} \in S_M(\theta^*, \lambda)$ is not absolutely continuous. Since $c(\cdot)$ satisfies costly local distinguishability, we can find an absolutely continuous s such that

$$\|s_{\theta^*, \lambda}, s\| < \lambda \cdot [c(s_{\theta^*, \lambda}) - c(s)].$$

Then, the gain from replacing $s_{\theta^*, \lambda}$ by s is

$$
\begin{aligned}
& [U(s, \theta^*) - \lambda \cdot c(s)] - [U(s_{\theta^*, \lambda}, \theta^*) - \lambda \cdot c(s_{\theta^*, \lambda})] \\
= & \int [s(\theta) - s_{\theta^*, \lambda}] \cdot [1_{\{\theta \geq \theta^*\}} - t] p(\theta) d\theta + \lambda \cdot [c(s_{\theta^*, \lambda}) - c(s)] \\
> & -\int |s(\theta) - s_{\theta^*, \lambda}| \cdot 1 \cdot p(\theta) d\theta + \|s_{\theta^*, \lambda}, s\| \\
= & 0,
\end{aligned}
$$

which contradicts the optimality of $s_{\theta^*, \lambda}$. ∎

This lemma explains the logic of continuous choice based on a primitive property on cost functionals. We next turn to the equilibrium implications of continuous choice.

14

4.2 Characterizing the Equilibria

A profile of effective strategies $\{s_i\}_{i\in[0,1]}$ will induce an *aggregate effective strategy*

$$\widehat{s}(\theta) = \int_{i\in[0,1]} s_i(\theta)\, di \tag{13}$$

which can be interpreted, assuming a continuum law of large numbers, as the proportion of players that invest conditional on the state being θ. If all individual effective strategies are monotone, then so is the aggregate effective strategy.

Now a profile of monotone effective strategies $\{s_i\}_{i\in I}$ will induce a unique threshold $\theta^* \in [0,1]$ such that $\widehat{s}(\theta) > \beta(\theta)$ for $\theta > \theta^*$ and $\widehat{s}(\theta) < \beta(\theta)$ for $\theta < \theta^*$. Thus an aggregate effective strategy gives rise to an event in the payoff state space

$$F_{\theta^*} = \{\theta \in \mathbb{R} : \theta \geq \theta^*\} \ .$$

We will call this a regime change event since it characterizes the set of states where there is regime change (i.e., the proportion investing exceeds $\beta(\theta)$). Now any player's opponents' strategies are summarized by a threshold θ^*. Hence, her optimal best response is equivalent to maximizing $U(s|\theta^*, \lambda)$.

Lemma 8 *A strategy profile $\{s_i\}_{i\in[0,1]}$ of monotone strategies is an equilibrium of the λ-game if (1) they induce a threshold θ^* and (2) all strategies are optimal in the (θ^*, λ)-decision problem, i.e., each $s_i \in S(\theta^*, \lambda)$.*

The proof is straightforward and hence omitted.

4.3 Limit Uniqueness

We say that there is limit uniqueness if, as $\lambda \to 0$, all monotone equilibria of the game converge to a unique equilibrium.

Definition 9 (limit uniqueness) *Cost functional $c(\cdot)$ satisfies limit uniqueness if there exists s^* such that, for any $\delta > 0$, there exists $\overline{\lambda} > 0$ such that $\|s, s^*\| \leq \delta$ whenever s is a monotone equilibrium strategy in the λ-game.*

We now study how costs vary as we translate the effective strategy. Let $T_\Delta : S \to S$ be a translation operator: that is, for any $\Delta \in \mathbb{R}$ and $s \in S$,

$$(T_\Delta s)(\theta) = s(\theta + \Delta) \ .$$

Definition 10 (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 information cost responds at most linearly to translations of the effective strategies. 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 does not depend on where attention is paid. This is not a stringent property as it is satisfied in most models of information acquisition in the literature.[9]

We will have a particular equilibrium played in the limit when there is limit uniqueness. Since the threshold function β is continuous and strictly increasing, it has a well-defined inverse. Setting $\theta^{**} = \beta^{-1}(1-t)$, the limit equilibrium will be the step function at θ^{**}. Morris and Shin (2001) defined the Laplacian selection to be the behavior that is a best response to a uniform belief over the proportion of other players choosing one action. In the regime change game, this is invest as long as $\theta \geq \theta^{**}$. This corresponds to the behavior that gets selected in global games.

Lemma 11 *If $c(\cdot)$ satisfies continuous choice and translation insensitivity, then $c(\cdot)$ satisfies limit uniqueness. In particular, in the limit all players follow strategy $1_{\{\theta \geq \theta^{**}\}}$, i.e., letting $\left\{ s_{i,\ \lambda}^{*} \right\}_{i \in [0,1]}$ denote an an equilibrium of the λ-game,*

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

for all $j \in [0,1]$.

According to the lemma, when $c(\cdot)$ satisfies continuous choice and translation insensitivity, and information costs are low, all equilibria are close to the switching strategy with threshold θ^{**}.[10] The property of continuous choice is essential to the limit uniqueness result. Recall that a player's ideal strategy is to sharply identify the event of regime change whenever it occurs. This requires perfectly distinguishing the states above the threshold of regime change from those below it, calling for an effective strategy discontinuous at the threshold. The property of continuous choice means that the players do not choose such sharp strategies for any relevant decision problems (i.e., (θ^{*}, λ)-decision problem with $\theta^{*} \in [0,1]$). This property of the information cost restricts the players' ability, and in

---

[9]For example, when information acquisition is modeled by the costly state verification (CSV) approach, as introduced by Townsend (1979) and employed in many later applications, all the non-trivial effective strategies incur the same cost and hence the translation insensitivity is naturally satisfied. Another popular example is to model information acquisition as paying to reduce the magnitude of the additive noise. In this case, translating an effective strategy amounts to translating the mapping from signal realizations to actions, which does not change the underlying information structures nor the information cost. Again, the translation insensitivity is naturally satisfied. It is worth noting that this second example nests the (exogenous) additive information structures in the global game models as a special case. Moreover, in these examples, translation insensitivity takes an even stronger form, translation invariance, i.e., the information cost does not change with respect to the translation of the effective strategies.

[10]The continuous choice property and the translation insensitivity can be greatly relaxed to their local versions. In particular, they only need to hold in a small neighborbood of the step functions. See Section 7.1 for the formal definitions and results.

turn, incentive to coordinate in acquiring information. As a result, the players can only coordinate in identifying a unique threshold event $F_{\theta^{**}}$ in equilibrium.

Now Lemmas 7 and 11 imply:

Proposition 12 *If $c(\cdot)$ satisfies costly local distinguishability and translation insensitivity, then $c(\cdot)$ satisfies limit uniqueness. In particular, in the limit all players follow strategy $1_{\{\theta \geq \theta^{**}\}}$.*

It is worth relating this result to our leading example. Consider approximating the step function $1_{\{\theta \geq \theta^*\}}$ by an effective strategy s with maximal slope $k > 1$. In our leading example

$$c\left(1_{\{\theta \geq \theta^*\}}\right) - c(s) = k^{-\gamma}$$

and

$$\left\|1_{\{\theta \geq \theta^*\}}, s\right\| \leq \left\|1_{\{\theta \geq \theta^*\}}, s_{\theta^*, k}\right\| = \frac{1}{4 \cdot (2A + 1) \cdot k},$$

where $s_{\theta^*, k}$ is given by (3). Hence, in the scenario of $\gamma < 1$, we obtain

$$\frac{c\left(1_{\{\theta \geq \theta^*\}}\right) - c(s)}{\left\|1_{\{\theta \geq \theta^*\}}, s\right\|} \geq 4 \cdot (2A + 1) \cdot k^{1-\gamma} \to \infty \text{ as } k \to \infty.$$

This inequality suggests that the step functions can be cheaply approximated by absolutely continuous functions when $\gamma < 1$. One can generalize this example, letting

$$c(s) = \max\left(0, 1 - \left(\sup_\theta \; \|k\|_p\right)^{-\gamma}\right)$$

where $k : \mathbb{R} \to \mathbb{R}_+$ is the derivative of the effective strategy s and $\|k\|_p$ is the p-norm of k

$$\|k\|_p = \left[\int [k(\theta)]^p \, g(\theta) \, d\theta\right]^{\frac{1}{p}}.$$

In this case, ==one show== that costly local distinguishability is satisfied if

$$\frac{1}{\gamma} + \frac{1}{p} > 1.$$

This reduces to our leading example as $p \to \infty$ and shows the robustness of the example to focussing on the maximum derivative.

4.4 Limit Multiplicity

In order to appreciate the importance of the conditions for limit uniqueness, this subsection contrasts these conditions to a sufficient condition for limit multiplicity.

Definition 13 (Lipschitz) *Cost functional $c\left(\cdot\right)$ is Lipschitz, if there exists a $K > 0$ such that $\left|c\left(s_1\right) - c\left(s_2\right)\right| < K \cdot \left\|s_1, s_2\right\|$ for all $s_1, s_2 \in S$.*

The Lipschitz property requires that the information cost responds at most linearly to any change of the effective strategies. This property is sufficient for the result of limit multiplicity.

Proposition 14 *If cost functional $c\left(\cdot\right)$ is Lipschitz, then there exists $\overline{\lambda} > 0$ such that the game has multiple equilibria for all $\lambda \in \left(0, \overline{\lambda}\right)$. In particular, for every $\theta^* \in [0, 1]$, $s_{i,\,\lambda}^* = 1_{\{\theta \geq \theta^*\}}$ for each $i \in [0, 1]$ is an equilibrium.*

Yang (2015) also showed this sufficient condition for multiplicity (in a closely related setting), as well as showing that there was multiplicity when the cost functional corresponded to entropy reduction. Entropy reduction is not a special case of Lipschitz, but the argument takes a very similar form (see section **??** for more discussion of this point).

Proof. Let $\overline{\lambda} = \min\left(\frac{t}{K}, \frac{1-t}{K}\right)$. It suffices to show that $1_{\{\theta \geq \theta^*\}} \in S\left(\theta^*, \lambda\right)$ for any $\theta^* \in [0, 1]$ and $\lambda \in \left(0, \overline{\lambda}\right)$. This is true because for any $s \neq 1_{\{\theta \geq \theta^*\}}$,

$$
\left[U\left(1_{\{\theta \geq \theta^*\}}, \theta^*\right) - \lambda \cdot c\left(1_{\{\theta \geq \theta^*\}}\right)\right] - \left[U\left(s, \theta^*\right) - \lambda \cdot c\left(s\right)\right]
$$

$$
= \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*\}} - t\right] \cdot \left[1_{\{\theta \geq \theta^*\}} - s\left(\theta\right)\right] \, p\left(\theta\right) d\theta - \lambda \cdot \left[c\left(s\right) - c\left(1_{\{\theta \geq \theta^*\}}\right)\right]
$$

$$
> t \cdot \int_{-\infty}^{\theta^*} s\left(\theta\right) \, p\left(\theta\right) d\theta + \left(1 - t\right) \cdot \int_{\theta^*}^{\infty} \left[1 - s\left(\theta\right)\right] \, p\left(\theta\right) d\theta - \lambda K \cdot \left\|1_{\{\theta \geq \theta^*\}}, s\right\|
$$

$$
\geq \min\left(t, 1 - t\right) \cdot \left[\int_{-\infty}^{\theta^*} s\left(\theta\right) \, p\left(\theta\right) d\theta + \int_{\theta^*}^{\infty} \left[1 - s\left(\theta\right)\right] \, p\left(\theta\right) d\theta\right] - \lambda K \cdot \left\|1_{\{\theta \geq \theta^*\}}, s\right\|
$$

$$
= \left[\min\left(t, 1 - t\right) - \lambda K\right] \cdot \left\|1_{\{\theta \geq \theta^*\}}, s\right\| > 0,
$$

where the first inequality follows the Lipschitz property. ∎

It is easy to see that the Lipschitz property implies the translation insensitivity.[11] It is also worth noting that by definition the Lipschitz property implies the failure of costly local distinguishability. Thus the Lipschitz property preserves translation insensitivity but fails costly local distinguishability, highlighting the importance of the latter condition.

---

[11] This is because

$$
\left\|T_\Delta s, s\right\| = \int \left|s\left(\theta + \Delta\right) - s\left(\theta\right)\right| \cdot p\left(\theta\right) d\theta \leq \overline{p} \cdot \int \left[s\left(\theta + \Delta\right) - s\left(\theta\right)\right] d\theta
$$

$$
= \overline{p} \cdot \int s'\left(\theta\right) d\theta \cdot \Delta + o\left(\Delta\right) \leq \overline{p} \cdot 1 \cdot \Delta + o\left(\Delta\right) < K \cdot \Delta,
$$

where $K > \overline{p} = \sup_{\theta \in \mathbb{R}} p\left(\theta\right)$.

In our leading example, the Lipschitz property corresponds to the case of $\gamma \geq 1$. To see this, choose s_1 and s_2 and let k_1 and k_2 be their maximal slopes, respectively. Since the optimal effective strategies converge to step function $1_{\{\theta \geq \theta^*\}}$ as $\lambda \to 0$, let $k_2 > k_1 \geq 1$ and s_1 and s_2 belong to $B_\delta \left(1_{\{\theta \geq \theta^*\}}\right)$ for some small $\delta > 0$. Then

$$\|s_1, s_2\| \geq [4 \cdot (2A + 1)]^{-1} \cdot \left[k_1^{-1} - k_2^{-1}\right] + O(\delta).$$

The first term of the right hand side is what the distance would be if $s_1 = s_{\widehat{\theta}, k_1}$ and $s_2 = s_{\widehat{\theta}, k_2}$, which provides a lower bound. So

$$\frac{c(s_2) - c(s_1)}{\|s_1, s_2\|} \leq 4 \cdot (2A + 1) \cdot \frac{k_1^{-\gamma} - k_2^{-\gamma}}{k_1^{-1} - k_2^{-1}} + O(\delta).$$

Since k_1^{-1} and k_2^{-1} belong to $[0, 1]$ and the derivative of $f(y) = y^\gamma$ is bounded if and only if $\gamma \geq 1$, the information cost in our leading example has the Lipschitz property if and only if $\gamma \geq 1$. It is then not surprising that any step function constitutes an equilibrium.

5 Information Acquisition and Higher Order Beliefs about Regime Change

The regime changes only if a sufficiently large proportion of players invest. A player is willing to invest only if she assigns high probability to regime change. Hence, the regime changes only if enough players assign high probability to regime change, enough players assign high probability to enough players assigning high probability to regime change, and so on. Therefore, an equilibrium regime change event is characterized by players' higher order beliefs about that event. In this section, we first provide a tight characterization of the equilibrium regime change event in terms of higher order beliefs given any profile of players' information choices. This exercise is an analogue to Morris and Yildiz (2016).[12]

Using the common belief characterization, we then analyze whether it is feasible and incentive compatible for the players to acquire information to form specific higher order beliefs in our leading example. This analysis reveals that although different information acquisition technologies, namely, the binary and additive information structures, result in observationally the same equilibrium outcomes, the underlying mechanisms could be very different.

---

[12] This exercise is less general as it relies on the one dimensional state space and continuum player assumption to get a common belief characterization, although it gives a common belief characterization for a different (regime change) game.

5.1 Common Belief Foundations of Regime Change

We introduce a notion of common belief that is relevant to the regime change game studied in this paper. Throughout this subsection, we fix an arbitrary profile of players' information choices $\{(X_i, q_i)\}_{i \in [0,1]}$. Let $\widetilde{q}_i(\cdot|x_i) \in \Delta(\mathbb{R})$ denote player i's posterior belief upon observing $x_i \in X_i$. For any event $E \subset \mathbb{R}$ and $\alpha \in \mathbb{R}$, define

$$B^{\alpha,\beta}(E) = \left\{ \theta \in \mathbb{R} : \int_{i \in [0,1]} q_i\left(\{x_i \in X_i : \widetilde{q}_i(E|x_i) \geq \alpha\}|\theta\right) di \geq \beta(\theta) \right\}. \tag{14}$$

This is the set of θ's that at least proportion $\beta(\theta)$ of players α-believe event E. Here a player α-believes event E if she assigns probability at least α to the true state being in E. More precisely, say that event E is (α, β)-believed in state θ if $\theta \in B^{\alpha,\beta}(E)$. Say event E is (α, β)-evident if $E \subset B^{\alpha,\beta}(E)$. That is, a proportion of at least β players α-believe event E whenever E occurs.

The belief operator $B^{\alpha,\beta}(\cdot)$ can be iterated. The set of θ's that the proportion of players who α-believe that E is (α, β)-believed is at least $\beta(\theta)$ corresponds to $B^{\alpha,\beta}(B^{\alpha,\beta}(E))$, written as $\left(B^{\alpha,\beta}\right)^2(E)$. Then the sequence of operators $\left(B^{\alpha,\beta}\right)^n(\cdot), n \in \mathbb{N}$ can be defined accordingly. Define

$$C^{\alpha,\beta}(E) = \cap_{n=1}^{\infty} \left(B^{\alpha,\beta}\right)^n(E) \ .$$

At any state $\theta \in C^{\alpha,\beta}(E)$, at least proportion $\beta(\theta)$ of players α-believe event E, at least proportion $\beta(\theta)$ of players α-believe that at least proportion $\beta(\theta)$ of players α-believe event E, and so on. Say that event E is *commonly* (α, β)-believed at state θ if $\theta \in C^{\alpha,\beta}(E)$.

In the remainder of this subsection, we fix the profile of information choices $\{(X_i, q_i)\}_{i \in [0,1]}$ and characterize the set of equilibria using the language for higher-order beliefs described above. A strategy profile can be summarized by its *regime change event*, i.e., the set of states where the regime changes under those strategies. We characterize which regime change events are consistent with equilibrium.

Suppose that F was the regime change event. Player i will invest only if she observes a signal at which she assigns probability at least t to event F, i.e., if

$$\widetilde{q}_i(F|x_i) \geq t.$$

Thus the probability that i will invest conditional on θ is

$$q_i\left(\{x_i \in X_i : \widetilde{q}_i(F|x_i) \geq t\}|\theta\right) \ .$$

By a continuum law of large numbers assumption, the proportion of players investing at θ

would be

$$\int_{i \in [0,1]} q_i \left(\{ x_i \in X_i : \widetilde{q}_i \left(F | x_i \right) \geq t \} | \theta \right) di \ .$$

So the regime will change if this expression is greater than $\beta(\theta)$. Thus it is an equilibrium condition that

$$F = \left\{ \theta \in \mathbb{R} : \int_{i \in [0,1]} q_i \left(\{ x_i \in X_i : \widetilde{q}_i \left(F | x_i \right) \geq t \} | \theta \right) di \geq \beta(\theta) \right\} .$$

This condition is equivalent to the requirement that

$$F = B^{t,\beta}(F) . \tag{15}$$

Hence the event of regime change must be a fixed point of the belief operator $B^{t,\beta}(\cdot)$. That is, the event of regime change is (t, β)-evident; and once the regime change is (t, β)-believed, the regime changes. It is clear that (15) is necessary for F to be an event of regime change. Since any event F that satisfies (15) can be an event of regime change, it is also sufficient. We summarize these results in the following Proposition.

Proposition 15 *Given any profile of the players' information choices, a subset $F \subset \mathbb{R}$ can be an event of regime change if and only if $F = B^{t,\beta}(F)$.*

To fully characterize the outcomes of the game, consider the complement event $\mathbb{R} \backslash F$, which is the event of no regime change. A symmetric argument shows that

$$\mathbb{R} \backslash F = B^{1-t,1-\beta}(\mathbb{R} \backslash F) \tag{16}$$

and

$$\mathbb{R} \backslash F = C^{1-t,1-\beta}(\mathbb{R} \backslash F) .$$

According to Lemma 22 in the appendix, (16) is equivalent to (15). Hence, once the profile of players' information choices is fixed, (15) fully characterizes the outcome of the game. Moreover, since

$$C^{t,\beta}(F) = \cap_{n=1}^{\infty} \left(B^{t,\beta} \right)^n (F) = F,$$

the proposition states that regime changes if and only if it is *commonly* (t, β)-believed. In games with endogenous information acquisition, whether an event is *commonly* (t, β)-believed depends on the players' information choices.[13] Hence, an event characterizes the regime change in equilibrium if and only if the players can form *common* (t, β)-belief on this

---

[13]Note that the belief operator $B^{t,\beta}(\cdot)$ is a function of the profile of the players' information choices.

event through their information acquisition. Propositions **??** and 14 relate the properties of the equilibria to the nature of the information acquisition technology: the players can only form *common* (t, β)-belief on a unique threshold event $F_{\theta^{**}}$ if the information technology features costly local distinguishability; otherwise, they can form *common* (t, β)-belief on almost all threshold events F_{θ^*} with $\theta^* \in [0, 1]$.

5.2 Information Acquisition and Higher Order Beliefs about Regime Change

The threshold events with threshold $\theta^* \in [0, 1]$ are potential candidates that could be the events of regime change. This subsection uses our leading example to illustrate how the players coordinate to form *common* (t, β)-beliefs on any of such threshold events. In particular, we will study both whether it is *feasible* to do so, in the sense that it is technologically possible under the players' information acquisition technology; and whether it is *incentive compatible*, in the sense that there is a Nash equilibrium of the information acquisition game giving rise to that outcome. The information cost and the threshold function in this subsection follow that of Section 3.

Proposition 16 *Any threshold event F_{θ^*} with threshold $\theta^* \in [0, 1]$ can be an event of regime change, if each player chooses the binary information structure*

$$q(x|\theta) = \begin{cases} s_{\widehat{\theta},k}(\theta) & \text{if } x = 1 \\ 1 - s_{\widehat{\theta},k}(\theta) & \text{if } x = 0 \end{cases},$$

where $s_{\widehat{\theta},k}$ is defined by (3) with

$$\widehat{\theta} = \left(1 + \frac{1}{k}\right)\theta^* - \frac{1}{2k}$$

and

$$k \in \left[\frac{t \cdot (1 - \theta^*)^2 + (1 - t) \cdot (\theta^*)^2}{2 \cdot \min\left[t \cdot (\theta^* + A), (1 - t) \cdot (1 + A - \theta^*)\right]}, \infty\right].$$

Proposition 16 states that by acquiring information properly, it is feasible for the players to achieve *common* (t, β)-beliefs on any threshold event F_{θ^*}, and thus make it the event of regime change. The players can choose a binary information structure that focuses around the threshold θ^* to distinguish F_{θ^*} from its complement sharply (i.e., k large enough). This seems to be a very natural result for binary information structures, but does not hold for the additive information structures that are typically employed in the global game models, as shown later in this subsection.

It is worth comparing the results of Propositions 3 and 16. When $\gamma \geq 1$, any threshold event F_{θ^*} with $\theta^* \in [0, 1]$ can be an event of regime change in equilibrium. That is, it

is feasible and also incentive compatible for the players to coordinate to form common (t, β)-beliefs on any of such threshold events. In contrast, when $\gamma < 1$, although by Proposition 16 it is feasible for the players to coordinate to form common (t, β)-beliefs on all the aforementioned threshold events, Proposition 3 implies that it is not incentive compatible for them to do so except for F_t. The key to understand this is the costly local distinguishability of the information cost. Simple calculation shows that, given any payoff gain function $1_{\{\theta \geq \theta^*\}} - t$ with $\theta^* \in [0, 1]$, the players' optimal binary information structure $s_{\widehat{\theta}(\theta^*), k^*}$ leads them to commonly (t, β)-believe event $F_{\widetilde{\theta}(\theta^*)}$, where $\widetilde{\theta}(\theta^*)$ is defined by equation (7). When $\gamma \geq 1$, as shown in Subsection 4.4 the information cost is Lipschitz and thus fails the costly local distinguishability. The players are willing to sharply identify any threshold event using a discontinuous binary information structure. This results in a maximal slope $k^* = \infty$ and thus $\widetilde{\theta}(\theta^*) = \theta^*$, making any $\theta^* \in [0, 1]$ the threshold of a regime change event in equilibrium. When $\gamma < 1$, as shown in Subsection 4.1 the information cost exhibits costly local distinguishability. The players find it suboptimal to sharply distinguish any threshold event and their optimal effective strategy is continuous ($k^* < \infty$), which makes $\widetilde{\theta}(\theta^*)$ strictly closer to t than θ^* for all $\theta^* \neq t$. This suggests that the players are not willing to coordinate to form common (t, β)-beliefs on any threshold event other than F_t.

It is also worth examining the higher order beliefs in a global game model. The setting follows Subsection 3.2. Each player i receives a signal $z_i = \theta + k^{-1} \cdot \varepsilon_i$, where $\varepsilon_i \sim U_{[-1/2, 1/2]}$ is independent from θ and across the players, and k^{-1} measures the magnitude of noise.

We first fix an $k < \infty$ and examine the corresponding belief operator $B^{t, \beta}(\cdot)$. Player i t-believes a threshold event F_{θ^*} if and only if

$$\frac{z_i + k^{-1}/2 - \theta^*}{k^{-1}} \geq t,$$

i.e.,

$$z_i \geq \theta^* + (t - 1/2) \cdot k^{-1}.$$

Hence, F_{θ^*} is (t, β)-believed if and only if

$$\frac{\theta + k^{-1}/2 - \theta^* - (t - 1/2) \cdot k^{-1}}{k^{-1}} \geq 1 - \theta,$$

i.e.,

$$\theta \geq \frac{k \cdot \theta^* + t}{1 + k} = \widetilde{\theta}(\theta^*).$$

Therefore,

$$B^{t, \beta}(F_{\theta^*}) = F_{\widetilde{\theta}(\theta^*)}.$$

Since $k < \infty$, it is clear that F_t is the only fixed point of $B^{t, \beta}(\cdot)$. That is, no matter how precise are the signals, it is only feasible for the players to form common (t, β)-beliefs on

23

F_t. The common (t, β)-beliefs on other threshold events are technologically precluded by the additive information structure.

In order to compare to the case of information acquisition with binary information structures, we next let the players acquire information about θ by increasing k as in subsection 3.2. The players choose a precision $k^* < \infty$ if and only if $\gamma < 1$, which leads to exactly the same effective strategies as with binary information structures. Hence if $\gamma < 1$, it is only feasible for the players to form common (t, β)-beliefs on F_t through acquiring additive signals. Since the players have no other choice, it is also incentive compatible to form common (t, β)-belief on F_t, which results in the same event of regime change in equilibrium as with binary information structures. It is worth highlighting the very different mechanisms behind the same equilibrium outcome of the two technologies. The additive information structure setting directly shrinks the players' "choice" of common (t, β)-beliefs to a singleton, which can only be formed on F_t. In contrast, the binary information structure setting makes common (t, β)-beliefs on a whole spectrum of threshold events feasible but the players are willing to form common (t, β)-belief on only one of these events.

6 Extensions

This section extends our main results in two directions: generalizing payoffs of the game and relaxing key assumptions and, in particular, replacing global properties of information cost functionals with local properties.

Recall that player i's payoff from investing is given by $\pi(l, \theta)$ where $l \in [0, 1]$ is the proportion of players that invest and θ is the state of the world. Instead of specifying a functional form of $\pi(l, \theta)$, we will impose the following properties on this general payoffs.

Assumption A2 (Monotonicity and Boundedness): *a)* $\pi(l, \theta)$ *is non-decreasing in l and θ; b)* $|\pi(l, \theta)|$ *is uniformly bounded.*

Assumption A3 (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 $\theta < \theta_l$.*

Assumption A4 (Strict Laplacian State Monotonicity and Continuity): *Let $\psi(\theta) = \int_0^1 \pi(l, \theta) \, dl$. Then, a) there exists a unique $\theta^{**} \in \mathbb{R}$, such that $\psi(\theta^{**}) = 0$; b) ψ is continuous, and ψ^{-1} exists on an open neighborhood of $\psi(\theta^{**})$.*

These assumptions are standard in the global game literature.[14] In particular, Assumptions A2 and A3 imply that any $s \in S_M$ induces 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 $\theta < \theta_s$. Assumption A2 further implies that $\theta_s \in [\theta_1, \theta_0]$ for all $s \in S_M$, where θ_1 and θ_0 are defined by choosing $l = 1$ and $l = 0$ in Assumption A3. Consequently, we obtain the limit dominance condition often assumed in the global game literature. That is, $\pi(l, \theta) > 0$ for all $l \in [0, 1]$ and $\theta > \theta_0$, and $\pi(l, \theta) < 0$ for all $l \in [0, 1]$

---

[14]See the general assumptions surveyed in Subsection 2.2 of Morris and Shin (2001).

and $\theta < \theta_1.$[15] Let

$$U\left(\widetilde{s}, s\right) = \int \pi\left(s\left(\theta\right), \theta\right) \cdot \widetilde{s}\left(\theta\right) p\left(\theta\right) d\theta$$

denote a player's expected payoff from playing effective strategy $\widetilde{s} \in S$ if all other players choose strategy $s \in S.$[16] Then the player's decision problem is

$$\max_{\widetilde{s} \in S} U\left(\widetilde{s}, s\right) - \lambda \cdot c\left(\widetilde{s}\right) . \tag{17}$$

Slightly abusing earlier notations, call this problem the (s, λ)-decision problem and let $S\left(s, \lambda\right)$ denote the set of solutions. We will again focus on the monotone equilibria. Note that the state contingent payoff $\pi\left(s\left(\theta\right), \theta\right)$ is non-decreasing in θ if s is non-decreasing. Hence, it is reasonable to make the following assumption that parallels Assumption A1 in Subsection 4.1.

Assumption A1': *The optimal effective strategies in the (s, λ)-decision problem are non-decreasing for all $s \in S_M$, i.e., $S\left(s, \lambda\right) \subset S_M$ for all $s \in S_M$ and $\lambda \geq 0$.*

We next relax the assumptions on the information cost functionals.

First, we define the local translation insensitivity instead of the global translation insensitivity.

Definition 17 (local translation insensitivity) *Cost functional $c\left(\cdot\right)$ is said to be locally translation insensitive at $s \in S$, if there exists a $\delta > 0$ and $K > 0$ such that $\left|c\left(T_\Delta \widetilde{s}\right) - c\left(\widetilde{s}\right)\right| < K \cdot \left|\Delta\right|$ holds for all $\widetilde{s} \in B_\delta\left(s\right)$ and $\Delta \in \mathbb{R}$ providing that $T_\Delta \widetilde{s} \in B_\delta\left(s\right)$.*

Second, we define a local version of the continuous choice property.

Definition 18 (locally continuous choice) *Cost functional $c\left(\cdot\right)$ satisfies locally continuous choice at $s \in S$, if there exists a $\delta > 0$ such that $S\left(\widetilde{s}, \lambda\right)$ consists only of absolutely continuous functions for all $\widetilde{s} \in B_\delta\left(s\right)$ and $\lambda \in \mathbb{R}_{++}$.*

These local properties are weaker than their counterparts in Section 4. Together with the assumptions on the general payoffs, they generalize our main results as follows.

Proposition 19 *If $c\left(\cdot\right)$ satisfies locally continuous choice and is locally translation insensitive at $1_{\{\theta \geq \theta_s\}}$ for all $\theta_s \in [\theta_1, \theta_0]$, then $c\left(\cdot\right)$ satisfies limit uniqueness. In particular, in the limit all players follow strategy $1_{\{\theta \geq \theta^{**}\}}$, i.e., letting $\left\{s_{i, \lambda}^*\right\}_{i \in [0,1]}$ denote an equilibrium of the λ-game,*

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

for all $j \in [0, 1]$.

---

[15] Note that we have $\theta_1 = 0$ and $\theta_0 = 1$ in the regime change game.

[16] Equivalently, $s\left(\theta\right)$ can be interpreted as the aggregate effective strategy, which is the proportion of the players that invest when the state is θ.

This proposition generalizes the results of Lemma 14 and shares the same intuition. In particular, as shown in the proof, the maintained assumption that the information cost is uniformly bounded for all effective strategies can be further relaxed. Indeed, the proof goes through when the information cost is bounded on a subset $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$ instead of all effective strategies. As shown by Lemma 23 in the appendix, this condition guarantees that the optimal strategies in $S(s, \lambda)$ uniformly converge to $1_{\{\theta \geq \theta_s\}}$ for all $s \in S_M$. Hence, even in the applications where the information cost is unbounded on $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$, our results are still valid providing the uniform convergence of the optimal effective strategies, a property satisfied in most models of information acquisition.

Analogously, the locally continuous choice property reflects a relatively high cost of distinguishing nearby states and can be microfounded by a local version of costly local distinguishability. The presentation is similar to its counterpart in Section 4 and is omitted here.

Finally, we relax the Lipschitz property and show that it implies limit multiplicity.

Definition 20 (locally Lipschitz) *Cost functional $c(\cdot)$ is locally Lipschitz at $s \in S$, if there exists a $\delta > 0$ and $K > 0$ such that $|c(s_2) - c(s_1)| < K \cdot \|s_1, s_2\|$ for all $s_1, s_2 \in B_\delta(s)$.*

Again, it is easy to see that the local Lipschitz property implies the local translation insensitivity. It is also worth noting that by definition the Lipschitz property implies the failure of local continuous choice. Thus the local Lipschitz property preserves local translation insensitivity but fails local continuous choice, highlighting the importance of the latter condition, as shown by the following proposition.

Proposition 21 *If the information cost $c(\cdot)$ is locally Lipschitz at $1_{\{\theta \geq \theta_s\}}$ for some $\theta_s \in [\theta_1, \theta_0]$, then there exists a $\overline{\lambda} > 0$ such that the game has multiple equilibria for all $\lambda \in (0, \overline{\lambda})$. In particular, for every $\theta'_s \in (\theta_1, \theta_0)$ in a neighborhood of θ_s, $s^*_{i, \lambda} = 1_{\{\theta \geq \theta'_s\}}$ for each $i \in [0, 1]$ is an equilibrium.*

This proposition is a generalization of Proposition 14 and the two propositions share the same intuition.

7 Discussion

7.1 Entropy Reduction and Learning about Others' Actions

We did not discuss entropy reduction cost in the body of the paper although, as discussed in the introduction, it has been shown to give rise to limit multiplicity by Yang (2015). Essentially the arguments in subsection 4.4 could be used to establish limit multiplicity. In this sense, entropy reduction is a particular cost function that delivers the economic results in our setting.

However, the entropy reduction cost function does not in fact satisfy the Lipschitz condition. To see why, note that a well known property of entropy reduction is that the marginal cost of pushing $s(\theta)$ to 0 or 1 tends to infinity, so that in any threshold decision problem, it would be optimal to choose a discontinuous effective strategy but not a step function. But the Lipschitz condition rules out the possibility that the marginal cost going to infinity. While this distinction is not important for our analysis, it is important in other contexts.

A maintained assumption in our analysis is that players acquire information about the state only. Denti (2016) has recently considered the problem when players can acquire information about others' information. Because players do not acquire perfect information under entropy reduction, there is some residual uncertainty about others' signals under endogenous information acquisition, and this gives rise to smoother best responses and a different answer for us - limit uniqueness - in this case. Note that under a Lipschitz condition, he too would get limit multiplicity.

7.2   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 scarce resources to discriminate. However, the ability to discriminate between states on either side of the threshold disappears as we approach the threshold, giving continuous choice in our sense in this setting.[17]   The allocation of resources in this case is at the unconscious neuro level. Subjects in Caplin and Dean (2014) are asked to discriminate between the number of balls on the screen, where allocation of resources is presumably a conscious choice (e.g., how much time to devote to the task). Ongoing work in Dean, Morris, and Trevino (2016) confirms that, given a threshold decision problem, an inability to distinguish nearby states arises as expected.

References

Blackwell, D. (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics 24*(2), 265–272.

Caplin, A. and M. Dean (2014). Revealed preference, rational inattention, and costly information acquisition. Technical report, National Bureau of Economic Research.

---

[17]We are grateful to Michael Woodford for providing this reference.

Carlsson, H. and E. v. Damme (1993). Global games and equilibrium selection. *Econometrica 61*(5), 989–1018.

Dean, M., S. Morris, and I. Trevino (2016). Endogenous information structures and play in global games. Technical report.

Denti, T. (2016). Games with unrestricted information acquisition. Technical report, MIT.

Jazayeri, M. and J. A. Movshon (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature 446*(7138), 912–915.

Morris, S., H. S. and M. Yildiz (2016). Common belief foundations of global games. *Journal of Economic Theory (forthcoming)*.

Morris, S. and H. S. Shin (1998). Unique equilibrium in a model of self-fulfilling currency attacks. *The American Economic Review 88*(3), pp. 587–597.

Morris, S. and H. S. Shin (2001). Global games: Theory and applications. Cowles Foundation Discussion Papers 1275R, Cowles Foundation for Research in Economics, Yale University.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Szkup, M. and I. Trevino (2015). Information acquisition in global games of regime change. *Journal of Economic Theory 160*, 387 – 428.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory 21*(2), 265–293.

Woodford, M. (2008). Inattention as a source of randomized discrete adjustment. Working paper, Columbia University.

Yang, M. (2015). Coordination with flexible information acquisition. *Journal of Economic Theory 158, Part B*, 721 – 738.

8   Appendix

Proof of Proposition 11.

Proof. Lemma 4 implies

$$\lim_{\lambda \to 0} \sup_{s_{\theta^*,\lambda} \in \ S(\theta^*,\lambda) \ \text{and} \ \theta^* \in [0,1]} \left\| s_{\theta^*,\lambda}, 1_{\{\theta \geq \theta^*\}} \right\| = 0. \tag{18}$$

In equilibrium, the aggregate effective strategy is given by

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

Assuming the continuum law of large numbers, the proportion of players that take action 1 conditional on θ is $\widehat{s}^*_{\lambda}(\theta)$. Since all $\left\{ s^*_{i,\,\lambda} \right\}_{i \in [0,1]}$ are absolutely continuous, $\widehat{s}^*_{\lambda}$ is also absolutely continuous. Hence, there exists a unique θ^*_{λ} such that

$$\widehat{s}^*_{\lambda}(\theta^*_{\lambda}) = \beta(\theta^*_{\lambda}). \tag{19}$$

Note that θ^*_{λ} is the threshold of regime change in this equilibrium. Since

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

and (18) implies

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

it suffices to show that $\theta^*_{\lambda} \to \theta^{**}$ as $\lambda \to 0$.

We first show that $\int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_{\lambda}\}} - t \right] \cdot p(\theta)\, d\widehat{s}^*_{\lambda}(\theta)$ is arbitrarily close to zero when λ 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

$$W(\Delta) = \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_{\lambda}\}} - t \right] \cdot s^*_{i,\,\lambda}(\theta + \Delta) \cdot p(\theta)\, d\theta - \lambda \cdot c\left(T_\Delta s^*_{i,\,\lambda} \right).$$

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

$$\int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t\right] \cdot \frac{ds^*_{i,\,\lambda}(\theta)}{d\theta} \cdot p(\theta)\, d\theta - \lambda \cdot \left. \frac{dc\left(T_\Delta s^*_{i,\,\lambda}\right)}{d\Delta}\right|_{\Delta=0}$$

$$= \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t\right] \cdot p(\theta)\, ds^*_{i,\,\lambda}(\theta) - \lambda \cdot \left. \frac{dc\left(T_\Delta s^*_{i,\,\lambda}\right)}{d\Delta}\right|_{\Delta=0} = 0.$$

Here $W'(0)$ exists because $s^*_{i,\,\lambda}$ is absolutely continuous. In addition, the translation insensitivity implies $-K < \left.\frac{dc\left(T_\Delta s^*_{i,\,\lambda}\right)}{d\Delta}\right|_{\Delta=0} < K$. Hence, for any small $\varepsilon > 0$, by choosing $\lambda \in (0, \varepsilon)$ we obtain

$$-K\varepsilon < \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t\right] \cdot p(\theta)\, ds^*_{i,\,\lambda}(\theta) < K\varepsilon \ .$$

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

$$-K\varepsilon < \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t\right] \cdot p(\theta)\, d\widehat{s}^*_\lambda(\theta) < K\varepsilon \ ,$$

i.e.,

$$\left| \int_{-\infty}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t\right] \cdot p(\theta)\, d\widehat{s}^*_\lambda(\theta) \right| < K\varepsilon \ . \tag{20}$$

Since the density function $p(\theta)$ is continuous on $[0, 1]$, it is also uniformly continuous on $[0, 1]$. Hence, for any $\varepsilon > 0$, we can find an $\eta > 0$ such that $\left|p(\theta) - p(\theta')\right| < \varepsilon$ for all $\theta, \theta' \in [0, 1]$ and $\left|\theta - \theta'\right| < \eta$. By (18), for all i, the effective strategy $s^*_{i,\,\lambda}$ converges to $1_{\{\theta \geq \theta^*_\lambda\}}$ in $L^1 - norm$, so does the aggregate effective strategy $\widehat{s}^*_\lambda$. Together with the monotonicity of $\widehat{s}^*_\lambda$, this implies the existence of a $\lambda_1 > 0$ such that for all $\lambda \in (0, \lambda_1)$, $\left|\widehat{s}^*_\lambda(\theta) - 1_{\{\theta \geq \theta^*_\lambda\}}\right| < \varepsilon$ for all $\theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty)$. Choosing $\lambda \in$

$(0, \min(\lambda_1, \varepsilon))$, by (20), we obtain

$$
\left| \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) \right|
$$

$$
< \left| \int_{-\infty}^{\theta^*_\lambda - \eta} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) + \int_{\theta^*_\lambda + \eta}^{\infty} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) \right| + K\varepsilon
$$

$$
\leq \int_{-\infty}^{\theta^*_\lambda - \eta} \left| 1_{\{\theta \geq \theta^*_\lambda\}} - t \right| \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) + \int_{\theta^*_\lambda + \eta}^{\infty} \left| 1_{\{\theta \geq \theta^*_\lambda\}} - t \right| \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) + K\varepsilon
$$

$$
\leq 2\overline{p}\varepsilon + K\varepsilon , \tag{21}
$$

where $\overline{p} = \sup_{\theta \in \mathbb{R}} p(\theta) < \infty$. By the definition of η, $|p(\theta) - p(\theta^*_\lambda)| < \varepsilon$ for all $\theta \in [\theta^*_\lambda - \eta, \theta^*_\lambda + \eta]$. Hence,

$$
\left| p(\theta^*_\lambda) \cdot \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] d\widehat{s}^*_\lambda(\theta) - \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] \cdot p(\theta) \, d\widehat{s}^*_\lambda(\theta) \right| < \varepsilon. \tag{22}
$$

Further note that

$$
\left| 1 - \beta(\theta^*_\lambda) - t - \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \left[1_{\{\theta \geq \theta^*_\lambda\}} - t \right] d\widehat{s}^*_\lambda(\theta) \right|
$$

$$
= \left| 1 - \beta(\theta^*_\lambda) - t - \widehat{s}^*_\lambda(\theta^*_\lambda + \eta) + \widehat{s}^*_\lambda(\theta^*_\lambda) + t \cdot [\widehat{s}^*_\lambda(\theta^*_\lambda + \eta) - \widehat{s}^*_\lambda(\theta^*_\lambda - \eta)] \right|
$$

$$
= \left| (1 - t) \cdot [1 - \widehat{s}^*_\lambda(\theta^*_\lambda + \eta)] - t \cdot \widehat{s}^*_\lambda(\theta^*_\lambda - \eta) \right|
$$

$$
\leq (1 - t) \cdot |1 - \widehat{s}^*_\lambda(\theta^*_\lambda + \eta)| + t \cdot |\widehat{s}^*_\lambda(\theta^*_\lambda - \eta)|
$$

$$
\leq \varepsilon , \tag{23}
$$

where the second equality follows (19), the last inequality follows the facts that $\widehat{s}^*_\lambda(\theta^*_\lambda - \eta) \leq \varepsilon$ and $1 - \widehat{s}^*_\lambda(\theta^*_\lambda + \eta) \leq \varepsilon$ when $\lambda \in (0, \lambda_1)$.

Inequalities (21), (22) and (23) together imply that

$$
|1 - \beta(\theta^*_\lambda) - t| < \varepsilon + \frac{2\overline{p} + K + 1}{p(\theta^*_\lambda)} \varepsilon
$$

$$
\leq \varepsilon + \frac{2\overline{p} + K + 1}{\underline{p}} \varepsilon ,
$$

where $\underline{p} = \inf_{\theta \in [0,1]} p(\theta) > 0$ since p is assumed to be continuous and strictly positive on $[0, 1]$. Hence, $\beta(\theta^*_\lambda)$ is arbitrarily close to $1 - t$ as $\lambda \to 0$. Therefore, the continuity of

$\beta^{-1}(\cdot)$ implies $\lim_{\lambda \to 0} \theta^*_\lambda = \beta^{-1}(1-t) = \theta^{**}$. ∎

Lemma 22 *For any event $E \subset \mathbb{R}$, $\alpha \in \mathbb{R}$ and $\beta : \mathbb{R} \to \mathbb{R}$, $E = B^{\alpha,\beta}(E)$ if and only if $\mathbb{R}\backslash E = B^{1-\alpha,1-\beta}(\mathbb{R}\backslash E)$.*

Proof. By the symmetry of the statement, it suffices to show that $E = B^{\alpha,\beta}(E)$ implies $\mathbb{R}\backslash E = B^{1-\alpha,1-\beta}(\mathbb{R}\backslash E)$. Since $E = B^{\alpha,\beta}(E)$, any state θ belongs to $\mathbb{R}\backslash E$ if and only if θ belongs to $\mathbb{R}\backslash B^{\alpha,\beta}(E)$, i.e., the proportion of players that do not α-believe E is at least $1-\beta(\theta)$. If a player does not α-believe E, she must $(1-\alpha)$-believe $\mathbb{R}\backslash E$. Hence, θ belongs to $\mathbb{R}\backslash E$ if and only if the proportion of players that $(1-\alpha)$-believe $\mathbb{R}\backslash E$ is at least $1-\beta(\theta)$, i.e., $\mathbb{R}\backslash E = B^{1-\alpha,1-\beta}(\mathbb{R}\backslash E)$. ∎

Proof of Proposition 16.

Proof. Let $s(\theta)$ be the proportion of players whose signal realization is 1, conditional on the true state being θ. Then

$$\widehat{\theta} = \left(1 + \frac{1}{k}\right)\theta^* - \frac{1}{2k} \tag{24}$$

implies that $s(\theta) \geq 1 - \theta$ if and only if F_{θ^*} is true. Together with (24),

$$k \geq \frac{t \cdot (1-\theta^*)^2 + (1-t) \cdot (\theta^*)^2}{2 \cdot \min\left[t \cdot (\theta^* + A), (1-t) \cdot (1+A-\theta^*)\right]}$$

implies that each player t-believes F_{θ^*} if and only if her signal realization is 1. Hence, $F_{\theta^*} = B^{t,\beta}(F_{\theta^*})$. Proposition 15 then leads to the desired result. ∎

Lemma 23 *If $c\left(1_{\{\theta \geq \theta_s\}}\right)$ as a function of θ_s is bounded for $\theta_s \in [\theta_1, \theta_0]$, then for any $\rho > 0$, there exists a $\lambda_1 > 0$ such that $S(s, \lambda) \subset B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ for all $s \in S_M$ and $\lambda \in (0, \lambda_1)$.*

Proof. For any $\delta > 0$, define

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

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 A3, this function is always strictly positive. Hence, its infimum on $[0, 1]$ exists and is strictly positive. That is, $z(\delta) > 0$ for all $\delta > 0$. In addition, for any $s \in S_M$ and $\theta \notin [\theta_s - \delta, \theta_s + \delta]$, we have

$$|\pi(s(\theta), \theta)| \geq |\pi(s(\theta_s), \theta)| \geq z(\delta), \tag{25}$$

where the first inequality follows Assumptions A2 and A3, and the second inequality follows the definition of $z(\delta)$.

If $S(s,\lambda) = \{1_{\{\theta \geq \theta_s\}}\}$, we are done. Now for any $\widetilde{s} \in S(s,\lambda)$ such that $\widetilde{s} \neq 1_{\{\theta \geq \theta_s\}}$, the optimality of $\widetilde{s}$ implies

$$\int_{-\infty}^{\infty} \pi(s(\theta),\theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right] p(\theta)\, d\theta \quad < \quad \lambda \cdot \left[c\left(1_{\{\theta \geq \theta_s\}}\right) - c(\widetilde{s})\right]$$

$$\leq \quad \lambda \cdot c\left(1_{\{\theta \geq \theta_s\}}\right). \tag{26}$$

Note that

$$\int_{-\infty}^{\infty} \pi(s(\theta),\theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right] p(\theta)\, d\theta$$

$$\geq \int_{\theta \notin [\theta_s - \delta, \theta_s + \delta]} \pi(s(\theta),\theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right] p(\theta)\, d\theta$$

$$\geq \int_{-\infty}^{\infty} z(\delta) \cdot \left|1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right| p(\theta)\, d\theta - \int_{\theta_s - \delta}^{\theta_s + \delta} z(\delta) \cdot \left|1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right| p(\theta)\, d\theta$$

$$\geq \quad z(\delta) \cdot \left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}\right\| - 2 \cdot z(\delta) \cdot \overline{p} \cdot \delta, \tag{27}$$

where $\overline{p} = \sup_{\theta \in \mathbb{R}} p(\theta) < \infty$, the first inequality holds since $\pi(s(\theta),\theta)$ and $\left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right]$ always have the same sign and thus

$$\int_{\theta_s - \delta}^{\theta_s + \delta} \pi(s(\theta),\theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right] p(\theta)\, d\theta > 0,$$

and the second inequality follows (25). Inequalities (26) and (27) imply

$$\left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}\right\| < \frac{\lambda \cdot c\left(1_{\{\theta \geq \theta_s\}}\right)}{z(\delta)} + 2 \cdot \overline{p} \cdot \delta. \tag{28}$$

Hence, for any $\rho > 0$, choose $\delta < \frac{\rho}{4 \cdot \overline{p}}$ and $\lambda_1 < \frac{z(\delta) \cdot \rho}{2 \cdot c\left(1_{\{\theta \geq \theta_s\}}\right)}$, we obtain $\left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}\right\| < \rho$ for all $\lambda \in (0, \lambda_1)$. (Note that $c\left(1_{\{\theta \geq \theta_s\}}\right) > 0$, otherwise we return to the case $S(s,\lambda) = \{1_{\{\theta \geq \theta_s\}}\}$.)

Let $c_1 = \sup_{\theta_s \in [\theta_1, \theta_0]} c\left(1_{\{\theta \geq \theta_s\}}\right)$. For any $\rho > 0$, choose $\delta < \frac{\rho}{4 \cdot \overline{p}}$ and $\lambda_1 < \frac{z(\delta) \cdot \rho}{2 \cdot c_1}$. Then inequality (28) implies $\left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}\right\| < \rho$ for all $s \in S_M$ and $\lambda \in (0, \lambda_1)$. ∎

Proof of Proposition 19.

Proof. The idea of the proof is similar to that of Proposition 11.

We have assumed that the information cost functional is uniformly bounded for all effective strategies. Here we will prove our results under a weaker condition that $c\left(1_{\{\theta \geq \theta_s\}}\right)$

as a function of θ_s is bounded for $\theta_s \in [\theta_1, \theta_0]$.

The local translation insensitivity implies that $c\left(1_{\{\theta \geq \theta_s\}}\right)$ is a continuous function of θ_s for $\theta_s \in [\theta_1, \theta_0]$, which further implies $\sup_{\theta_s \in [\theta_1, \theta_0]} c\left(1_{\{\theta \geq \theta_s\}}\right) < \infty$. Then by Lemma 23, we have

$$\lim_{\lambda \to 0} \sup_{\widetilde{s}_{s,\lambda} \in S(s,\lambda) \text{ and } s \in S_M} \left\| \widetilde{s}_{s,\lambda}, 1_{\{\theta \geq \theta_s\}} \right\| = 0. \tag{29}$$

Let $\left\{ s_{i,\lambda}^* \right\}_{i \in [0,1]}$ denote a monotone equilibrium of the λ-game. Then the aggregate effective strategy is given by

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

which by Assumptions A2 and A3 induces a threshold θ_λ^* such that $\pi\left(\widehat{s}_\lambda^*(\theta), \theta\right) > 0$ if $\theta > \theta_\lambda^*$ and $\pi\left(\widehat{s}_\lambda^*(\theta), \theta\right) < 0$ if $\theta < \theta_\lambda^*$. By (29),

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

Since

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

it suffices to show that θ_λ^* becomes arbitrarily close to θ^{**} as $\lambda \to 0$.

We first show that the local translation insensitivity and local continuous choice property can be extended to a neighborhood of $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$. For any $\theta_s \in [\theta_1, \theta_0]$, since $c(\cdot)$ is locally translation insensitive at $1_{\{\theta \geq \theta_s\}}$, there exists $\delta(\theta_s) > 0$ and $K(\theta_s) > 0$ such that $|c(T_\Delta \widetilde{s}) - c(\widetilde{s})| < K(\theta_s) \cdot |\Delta|$ holds for all $\widetilde{s} \in B_{\delta(\theta_s)}\left(1_{\{\theta \geq \theta_s\}}\right)$ and $\Delta \in \mathbb{R}$, providing that $c(\widetilde{s}) < \infty$ and $T_\Delta \widetilde{s} \in B_{\delta(\theta_s)}\left(1_{\{\theta \geq \theta_s\}}\right)$. It is straightforward to see that $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$ is a sequentially compact subset of the metric space S and thus it is also compact. Since $\left\{ B_{\delta(\theta_s)}\left(1_{\{\theta \geq \theta_s\}}\right) : \theta_s \in [\theta_1, \theta_0] \right\}$ is an open cover of $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$, it has a finite subcover, denoted by

$$\left\{ B_{\delta(\theta_s^1)}\left(1_{\{\theta \geq \theta_s^1\}}\right), B_{\delta(\theta_s^2)}\left(1_{\{\theta \geq \theta_s^2\}}\right), \cdots, B_{\delta(\theta_s^n)}\left(1_{\{\theta \geq \theta_s^n\}}\right) \right\} \ .$$

Then

$$B_{\delta(\theta_s^1)}\left(1_{\{\theta \geq \theta_s^1\}}\right) \cup B_{\delta(\theta_s^2)}\left(1_{\{\theta \geq \theta_s^2\}}\right) \cup \cdots \cup B_{\delta(\theta_s^n)}\left(1_{\{\theta \geq \theta_s^n\}}\right)$$

is a finite open cover of $\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\}$. Hence, there exists a $\rho > 0$ such that

$$\begin{aligned}
\left\{ 1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0] \right\} \quad &\subset \quad \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right) \\
&\subset \quad B_{\delta(\theta_s^1)}\left(1_{\{\theta \geq \theta_s^1\}}\right) \cup B_{\delta(\theta_s^2)}\left(1_{\{\theta \geq \theta_s^2\}}\right) \cup \cdots \cup B_{\delta(\theta_s^n)}\left(1_{\{\theta \geq \theta_s^n\}}\right) \ .
\end{aligned}$$

Let $K = \max\left\{K\left(\theta_s^1\right), K\left(\theta_s^2\right), \cdots, K\left(\theta_s^n\right)\right\}$. Therefore, $\left|c\left(T_\Delta \widetilde{s}\right) - c\left(\widetilde{s}\right)\right| < K \cdot |\Delta|$ holds for all $\widetilde{s} \in \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ and $\Delta \in \mathbb{R}$, providing that $c\left(\widetilde{s}\right) < \infty$ and $T_\Delta \widetilde{s} \in \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$. The same argument shows that $c\left(\cdot\right)$ satisfies the locally continuous choice in an open neighborhood of $\left\{1_{\{\theta \geq \theta_s\}} : \theta_s \in [\theta_1, \theta_0]\right\}$. That is, for all s in such neighborhood, $S\left(s, \lambda\right)$ consists of only absolutely continuous effective strategies. Slightly abusing the notation but without loss of generality, we denote the neighborhood in which both locally continuous choice property and local translation insensitivity hold by $\cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$.

We next show that $\displaystyle\int_{-\infty}^{\infty} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right)$ is arbitrarily close to zero when λ is small enough. By (29), there exists a $\lambda_1 > 0$ such that $S\left(s, \lambda\right) \subset \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ for all $s \in S_M$ and $\lambda \in (0, \lambda_1)$. Hence, by choosing $\lambda < \lambda_1$, we have $s_{i,\lambda}^* \in S\left(\widehat{s}_\lambda^*, \lambda\right) \subset \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ for all $i \in [0, 1]$. This also implies that the aggregate effective strategy $\widehat{s}_\lambda^* \in \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ and thus $s_{i,\lambda}^*$ is absolutely continuous for all $i \in [0, 1]$. Now consider player i's expected payoff from slightly shifting her equilibrium strategy $s_{i,\lambda}^*$ to $T_\Delta s_{i,\lambda}^* \in \cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$, which is given by

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

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

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

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

Here $W'\left(0\right)$ exists because $s_{i,\lambda}^*$ is absolutely continuous. Since the local translation insensitivity has been extended to $\cup_{\theta_s \in [\theta_1, \theta_0]} B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$, we have $-K < \left.\frac{dc\left(T_\Delta s_{i,\lambda}^*\right)}{d\Delta}\right|_{\Delta = 0} < K$. Hence, for any small $\varepsilon > 0$, by choosing $\lambda \in (0, \min\left(\lambda_1, \varepsilon\right))$ we obtain

$$-K\varepsilon < \int_{-\infty}^{\infty} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) ds_{i,\lambda}^*\left(\theta\right) < K\varepsilon .$$

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

$$-K\varepsilon < \int_{-\infty}^{\infty} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) < K\varepsilon ,$$

i.e.,

$$\left| \int_{-\infty}^{\infty} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) \right| < K\varepsilon .$$ (30)

Since the density function $p\left(\theta\right)$ is continuous on $[\theta_1, \theta_0]$, it is also uniformly continuous on $[\theta_1, \theta_0]$. For the same reason, $\psi\left(\theta\right)$ is also uniformly continuous on $[\theta_1, \theta_0]$. Hence, for any $\varepsilon > 0$, we can find an $\eta > 0$ such that $\left| p\left(\theta\right) - p\left(\theta'\right) \right| < \varepsilon$ and $\left| \psi\left(\theta\right) - \psi\left(\theta'\right) \right| < \varepsilon$ for all $\theta, \theta' \in [\theta_1, \theta_0]$ and $\left| \theta - \theta' \right| < 2\eta$. Without loss of generality, we can choose $\eta < \varepsilon$. By (29), for all i, the effective strategy $s_{i,\lambda}^*$ converges to $1_{\{\theta \geq \theta_\lambda^*\}}$ in $L^1 - norm$, so does the aggregate effective strategy $\widehat{s}_\lambda^*$. Together with the monotonicity of $\widehat{s}_\lambda^*$, this implies the existence of a $\lambda_2 > 0$ such that for all $\lambda \in (0, \lambda_2)$, $\left| \widehat{s}_\lambda^*\left(\theta\right) - 1_{\{\theta \geq \theta_\lambda^*\}} \right| < \varepsilon$ for all $\theta \in (-\infty, \theta_\lambda^* - \eta) \cup (\theta_\lambda^* + \eta, \infty)$. Choosing $\lambda \in (0, \min\left(\lambda_1, \lambda_2, \varepsilon\right))$, by (30), we obtain

$$\left| \int_{\theta_\lambda^* - \eta}^{\theta_\lambda^* + \eta} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) \right|$$

$$< \int_{-\infty}^{\theta_\lambda^* - \eta} \left| \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \right| \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) + \int_{\theta_\lambda^* + \eta}^{\infty} \left| \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \right| \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) + K\varepsilon$$

$$\leq 2L\overline{p}\varepsilon + K\varepsilon ,$$ (31)

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

$$\left| p\left(\theta_\lambda^*\right) \cdot \int_{\theta_\lambda^* - \eta}^{\theta_\lambda^* + \eta} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) d\widehat{s}_\lambda^*\left(\theta\right) - \int_{\theta_\lambda^* - \eta}^{\theta_\lambda^* + \eta} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) \cdot p\left(\theta\right) d\widehat{s}_\lambda^*\left(\theta\right) \right| < L\varepsilon.$$ (32)

Inequalities (31) and (32) imply

$$\left| \int_{\theta_\lambda^* - \eta}^{\theta_\lambda^* + \eta} \pi\left(\widehat{s}_\lambda^*\left(\theta\right), \theta\right) d\widehat{s}_\lambda^*\left(\theta\right) \right| < \frac{2L\overline{p} + K + L}{\underline{p}}\varepsilon ,$$ (33)

where $\underline{p} = \inf_{\theta \in [\theta_1, \theta_0]} p\left(\theta\right) > 0$ since p is assumed to be continuous and strictly positive on $[\theta_1, \theta_0]$.

Next note that

$$\left| \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda + \eta)\, ds - \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda - \eta)\, ds \right| \leq |\psi(\theta^*_\lambda + \eta) - \psi(\theta^*_\lambda - \eta)| + 4L\varepsilon$$

$$< \varepsilon + 4L\varepsilon, \tag{34}$$

where the first inequality follows the fact that $\left| \widehat{s}^*_\lambda(\theta) - 1_{\{\theta \geq \theta^*_\lambda\}} \right| < \varepsilon$ for all $\theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty)$, and the second inequality follows the uniform continuity of $\psi(\theta)$ on $[\theta_1, \theta_0]$.

Further note that Assumption A2 implies

$$\int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda - \eta)\, ds \leq \int_{\theta^*_\lambda-\eta}^{\theta^*_\lambda+\eta} \pi(\widehat{s}^*_\lambda(\theta), \theta)\, d\widehat{s}^*_\lambda(\theta) \leq \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda + \eta)\, ds,$$

which together with (33) and (34) implies

$$-\left(\frac{2L\overline{p} + K + L}{\underline{p}} + 4L + 1 \right)\varepsilon \quad < \quad \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda - \eta)\, ds$$

$$\leq \quad \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda + \eta)\, ds < \left(\frac{2L\overline{p} + K + L}{\underline{p}} + 4L + 1 \right)\varepsilon \tag{35}$$

By Assumption A2, the monotonicity of $\pi(s, \theta)$ in θ implies

$$\left| \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi(s, \theta^*_\lambda)\, ds \right| < \left(\frac{2L\overline{p} + K + L}{\underline{p}} + 4L + 1 \right)\varepsilon .$$

Again, using the fact that $\left| \widehat{s}^*_\lambda(\theta) - 1_{\{\theta \geq \theta^*_\lambda\}} \right| < \varepsilon$ for all $\theta \in (-\infty, \theta^*_\lambda - \eta) \cup (\theta^*_\lambda + \eta, \infty)$, the above inequality implies

$$\left| \int_0^1 \pi(s, \theta^*_\lambda)\, ds \right| < \left(\frac{2L\overline{p} + K + L}{\underline{p}} + 6L + 1 \right)\varepsilon .$$

Therefore, we have

$$\lim_{\lambda \to 0} \psi(\theta^*_\lambda) = 0,$$

which implies

$$\lim_{\lambda \to 0} \theta_\lambda^* = \theta^{**}$$

according to Assumption A4. ∎

The proof of Proposition 21.

Proof. Choose $\rho > 0$ and $K > 0$ such that $|c(s_2) - c(s_1)| < K \cdot \|s_1, s_2\|$ for all $s_1, s_2 \in B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$. Note that we can always let $\theta_s \in (\theta_1, \theta_0)$. This is without loss of generality because by definition, for any $1_{\{\theta \geq \theta_s'\}} \in B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ with $\theta_s' \in (\theta_1, \theta_0)$, the information cost is also locally Lipschitz at $1_{\{\theta \geq \theta_s'\}}$.

Let $s \in S_M$ denote the effective strategy that induces the cutoff θ_s. Assumption A2 then implies that θ_s is also the cutoff for $\pi\left(1_{\{\theta \geq \theta_s\}}, \theta\right)$. This is because $\pi\left(1_{\{\theta \geq \theta_s\}}, \theta\right) \geq \pi\left(s(\theta), \theta\right) > 0$ for $\theta > \theta_s$ and $\pi\left(1_{\{\theta \geq \theta_s\}}, \theta\right) \leq \pi\left(s(\theta), \theta\right) < 0$ for $\theta < \theta_s$. In addition, $\theta_s \in (\theta_1, \theta_0)$ implies

$$\inf\left(\{\pi(1, \theta) : \theta > \theta_s\}\right) > 0$$

and

$$\sup\left(\{\pi(0, \theta) : \theta < \theta_s\}\right) < 0 .$$

Let

$$b = \min\left\{\inf\left(\{\pi(1, \theta) : \theta > \theta_s\}\right), -\sup\left(\{\pi(0, \theta) : \theta < \theta_s\}\right)\right\} .$$

We next show that $s_{i,\lambda}^* = 1_{\{\theta \geq \theta_s\}}$ for all $i \in [0, 1]$ is an equilibrium. Since θ_s is the cutoff for $\pi\left(1_{\{\theta \geq \theta_s\}}, \theta\right)$, Lemma 23 implies the existence of a $\lambda_1 > 0$ such that $S\left(1_{\{\theta \geq \theta_s\}}, \lambda\right) \in B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ for all $\lambda \in (0, \lambda_1)$. Let $\overline{\lambda} = \min\left(\lambda_1, \frac{b}{K}\right)$. It thus suffices to show that $1_{\{\theta \geq \theta_s\}}$ dominates all $\widetilde{s} \in B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ when $\lambda \in \left(0, \overline{\lambda}\right)$. This is true because

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \pi\left(1_{\{\theta \geq \theta_s\}}, \theta\right) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}(\theta)\right] p(\theta)\, d\theta - \lambda \cdot \left[c\left(1_{\{\theta \geq \theta_s\}}\right) - c(\widetilde{s})\right] \\
&\geq \quad b \cdot \left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}\right\| - \lambda \cdot \left[c\left(1_{\{\theta \geq \theta_s\}}\right) - c(\widetilde{s})\right] \\
&> \quad (b - \lambda K) \cdot \left\|1_{\{\theta \geq \theta_s\}}, s\right\| > 0,
\end{aligned}
$$

where the first inequality follows the definition of b and the second inequality follows the local Lipschitz property.

Finally, note that by definition, for any $1_{\{\theta \geq \theta_s'\}} \in B_\rho\left(1_{\{\theta \geq \theta_s\}}\right)$ with $\theta_s' \in (\theta_1, \theta_0)$, the information cost is also locally Lipschitz at $1_{\{\theta \geq \theta_s'\}}$. Hence $s_{i,\lambda}^* = 1_{\{\theta \geq \theta_s'\}}$ for all $i \in [0, 1]$ is another equilibrium. ∎