DISCUSSION PAPER SERIES

DP16495

Coordination and Continuous Stochastic Choice

Stephen Morris and Ming Yang

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

Coordination and Continuous Stochastic Choice

Stephen Morris and Ming Yang

Discussion Paper DP16495
Published 30 August 2021
Submitted 28 August 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Coordination and Continuous Stochastic Choice

Abstract

Players receive a return to investment that is increasing in the proportion of others who invest and the state, and incur a small cost for acquiring information about the state. Their information is reflected in a stochastic choice rule, specifying the probability of a signal leading to investment. If discontinuous stochastic choice rules are infinitely costly, there is a unique equilibrium as costs become small, in which actions are a best response to a uniform (Laplacian) belief over the proportion of others investing. Infeasibility of discontinuous stochastic choice rules captures the idea that it is impossible to perfectly distinguish states that are arbitrarily close together and is both empirically documented and satisfied by many natural micro-founded cost functionals on information. Our results generalize global game selection results (Carlsson and van Damme (1993) and Morris and Shin (2003)), and establish that they do not depend on the specific additive noise information structure.

JEL Classification: C72, D82

Keywords: coordination, endogenous information acquisition, continuous stochastic choice

Stephen Morris - stephen.edward.morris@gmail.com
M.I.T. and CEPR

Ming Yang - m-yang@ucl.ac.uk
University College London

Coordination and Continuous Stochastic Choice*

Stephen Morris† and Ming Yang‡
Massachusetts Institute of Technology and University College London

August 2021; click **here** for the latest version

Abstract

Players receive a return to investment that is increasing in the proportion of others who invest and the state, and incur a small cost for acquiring information about the state. Their information is reflected in a stochastic choice rule, specifying the probability of a signal leading to investment. If discontinuous stochastic choice rules are infinitely costly, there is a unique equilibrium as costs become small, in which actions are a best response to a uniform (*Laplacian*) belief over the proportion of others investing. Infeasibility of discontinuous stochastic choice rules captures the idea that it is impossible to perfectly distinguish states that are arbitrarily close together and is both empirically documented and satisfied by many natural micro-founded cost functionals on information. Our results generalize global game selection results (Carlsson and van Damme (1993) and Morris and Shin (2003)), and establish that they do not depend on the specific additive noise information structure.

JEL: C72 D82

KEYWORDS: coordination, endogenous information acquisition, continuous stochastic choice

*An earlier version of this paper was circulated under the title "Coordination and the Relative Cost of Distinguishing Nearby States" (Morris and Yang (2016b)) and presented at the INEX conference at NYU in February 2014 and the 2014 SAET annual meeting in Tokyo. We are grateful for the comments of conference discussants Muhamet Yildiz (who discussed the paper at both the "Global Games in Ames" conference in April 2016 and the Cowles Economic Theory conference in June 2016) and Jakub Steiner (who discussed the paper at the AEA winter meetings in January 2018). Morris discussed experimental ideas related to this project with Mark Dean and Isabel Trevino, discussed in Section 5.3. We are also grateful for feedback from Joel Flynn, Ben Hebert, Filip Matejka, Chris Sims, Philipp Strack, Tomasz Strzalecki, Mike Woodford and many seminar participants. We received valuable research assistence from Ole Agersnap, Aroon Narayanan and Victor Orestes. Morris gratefully acknowledges financial support from NSF grants SES-1824137 and SES-2049744.

†Department of Economics, MIT, The Morris and Sophie Chang Building, 77 Massachusetts Ave., Building E52-556, Cambridge, MA 02139 United States, semorris@mit.edu

‡Department of Economics and School of Management, UCL, Drayton House, 30 Gordon St, London WC1H 0AX, United Kingdom, m-yang@ucl.ac.uk

1 Introduction

Consider a canonical binary action coordination game, where each player receives a payoff to investing that is increasing in the proportion of other players investing and the payoff relevant state of the world. Suppose that before playing the coordination game, players acquire information about the state of the world. By a now standard observation (reviewed below), if the cost of information is increasing in its informativeness (in the sense of Blackwell),[1] it is without loss of generality to assume that players facing a binary choice will only acquire binary signals about the state of the world and choose to invest after observing one signal and to not invest after the other signal. Thus first choosing a costly experiment and then a strategy for the induced Bayesian game is equivalent to choosing a stochastic choice rule specifying a probability of investing in each state of the world.

We therefore study a *stochastic choice game* where each player chooses a stochastic choice rule and payoffs are given by the induced payoff in the base coordination game minus a small cost of information associated with the stochastic choice rule. We study what happens when we fix a cost functional on stochastic choice rules and multiply the cost functional by a constant that becomes small. Two properties of the cost functional are key. *Infeasible perfect discrimination* requires that discontinuous stochastic choice rules are infinitely costly (and thus infeasible). Infeasible perfect discrimination thus requires that it is impossible to perfectly discriminate between states that are arbitrarily close together. *Translation insensitivity* requires that translating a stochastic choice rule does not induce a discontinuous change in the cost. Translation insensitivity requires that the cost of information depends on how much discrimination between states occurs but does not change abruptly when the focus of that discrimination moves a small amount. Our main result is that - if the cost functional satisfies these two properties - there is a unique equilibrium of the stochastic choice game in the small cost limit. In particular, the unique equilibrium of the stochastic choice game selects a particular Nash equilibrium of each complete information game corresponding to a state of the world. In the limit, the *Laplacian* equilibrium of the complete information game is played: players invest when invest is a best response to a uniform belief over the proportion of other players choosing to invest in the complete information game.[2] Thus an arbitrarily small cost of information gives rise to a natural equilibrium selection with the selected outcome independent of the fine details of the cost functional.

Our infeasible perfect discrimination (IPD) property makes sense if the distance between states matters and nearby states are relatively hard to distinguish. Experimental evidence is consistent with IPD (and discussed in Section 5.3). We illustrate IPD (and other properties

[1] That is, if an experiment is a Blackwell garbling of another, then it is cheaper.

[2] This is also the equilibrium selected by the global games approach, which is a special case of our theory as discussed below.

of the cost of information in the paper) by considering *pairwise-separable cost functionals*, a variant of the log-likelihood ratio cost functions recently introduced and axiomatized by Pomatto, Strack, and Tamuz (2019) (their axiomatization is discussed in Section 5.2). This cost functional sums up, over pairs of states, the product of the indistinguishability of the two states and divergence between the distributions of signals at those two states. We focus on a parameterization where the indistinguishability of states θ and θ' is $\left|\theta - \theta'\right|^{-\alpha}$, so a higher α corresponds to a greater relative cost of distinguishing nearby versus distant states; IPD is satisfied if $\alpha \geq 2$.

Infeasible perfect discrimination (together with translation insensitivity) is sufficient for Laplacian selection and is intuitive, easy to check and easy to interpret. To better understand what is driving our result, we also provide a partial converse and a weaker sufficient condition. If nearby states are relatively easy to distinguish - we say that there is *cheap perfect discrimination* (CPD) - then multiplicity is restored. It has become common in the recent literature to measure the cost of information by the reduction in entropy. Entropy is an information theoretic notion under which the distance between states has no meaning or significance. This implies that the entropy reduction cost functional satisfies our cheap perfect discrimination property and thus gives rise to multiple equilibria. For the pairwise-separable cost functional, CPD is satisfied when α is close to zero.

The key to our main result is that players choose continuous stochastic choice rules in equilibrium, and continuous stochastic choice is the observable implication of IPD. Infeasible perfect discrimination directly imposes continuous stochastic choice. We also report a weaker condition - *expensive perfect discrimination* (EPD) - which implies that players will choose continuous stochastic choice rules in equilibrium, even when it is feasible to choose discontinuous stochastic choice rules that perfectly discriminate between neighboring states at finite cost. EPD is satisfied if α is close to but smaller than 2.

Our results cleanly embed and generalize a leading result on *global games* as a special case. Suppose that each player could observe a signal equal to the true state plus some additive noise. A higher precision of the signal (i.e., lower variance of the noise) is more expensive. But it is infeasible to acquire a perfect signal. The induced cost functional on stochastic choice rules satisfies infeasible perfect discrimination (because perfect signals are infeasible) and translation insensitivity (because noise is additive) and our main result thus implies leading results about equilibrium selection in global games (we review the relevant literature below).[3] Our results thus establish that the conclusions of the global games literature do not rely on the particular (additive noise) information structure but follow much more generally from infeasible perfect discrimination.

[3] The global game result is not quite a special case of our main result, because we maintain a submodularity restriction on cost functionals in order to maintain supermodularity of the underlying game that fails for the global game cost functional. But in Section 5.1, we discuss how the global game result is a special case of a variation of our main result.

An alternative approach to modelling our problem would be to consider a two stage game where players first decide what information to acquire about the state of the world, and then decide a strategy in the induced game. But any experiment acquired and strategy chosen will induce a stochastic choice rule, which can be understood as a two-signal experiment. Now suppose that the cost of information satisfies Blackwell-consistency, i.e., an experiment that is weakly less informative in the sense of Blackwell (1953)) than another, is always weakly less expensive. Thus it is without loss of generality to consider the one stage game, where players choose a stochastic choice rule that can be understood as both a two-signal experiment and the effective strategy of the player.[4] Our focus on stochastic choice rules is thus motivated by appeal to Blackwell-consistency on cost functionals on general experiments. Almost all the cost functionals we study in this paper (pairwise-separable, entropy reduction, Fisher information, etc.) are defined on general experiments and satisfy Blackwell-consistency in that general space. The additive noise cost in global games does not satisfy it, but where we discuss this case (in Section 5.2.3), we work with a Blackwell-consistent version where players have free disposal of information.

Our main treatment relies on a maintained restriction of submodularity on cost functionals. This ensures that the game is supermodular and thus simplifies our analysis. Submodularity requires that the sum of the cost of two stochastic choice rules is more than the cost of the maximum of the stochastic choice rules plus the cost of the minimum. However, we also discuss the case without submodular costs, noting that submodularity is not necessary and also noting that our results continue to hold if we restricted players to choose monotonic stochastic choice rules (in Section 5.1).

While we have an information acquisition interpretation of the cost of stochastic choice rules, our results are independent of the interpretation of the cost functional. Alternative interpretations are possible, including "control costs" of taking an action, and all our results are of interest also with this alternative interpretations. We focus on the informational interpretation in the body of the paper, but briefly discuss alternative interpretations in Section 5.4.

The properties of cost functional we appeal to (e.g., IPD, EPD and CPD) are rather abstract. We discuss the relation of these properties to various cost functionals studied in the literature (in addition to the pairwise-separable cost functionals) and discuss their foundations (based on axioms, sequential choice foundations, additive noise and coding) in Section 5.2.

The observable implication of IPD is continuous stochastic choice and we discuss variety of empirical (field and experimental) evidence in favor of IPD in Section 5.3. We also discuss the experimental work of Goryunov and Rigos (2020), confirming our theory.

We proceed as follows. Section 2 sets up the model of a stochastic choice game. Section

[4] See Woodford (2009), Yang (2015) and the working paper version of this paper, Morris and Yang (2016a) for formal versions of this argument.

3 contains our main result about stochastic choice games. Section 4 reports a converse and weaker sufficient conditions for Laplacian selection, in order to deepen our understanding of the main result. Section 5 collects together discussion of assumptions, extensions, and alternative foundations and interpretations of the cost functionals.

All proofs are in the Appendix unless otherwise stated.

1.1 Related Literature and Broader Context

Carlsson and Damme (1993) introduced global games, where players exogenously observe the true payoffs with a small amount of additive noise, and showed that there is a unique equilibrium played in the limit as noise goes to zero. These results have since been significantly generalized and widely applied. Morris and Shin (2003) provide an early survey of theory and applications; the class of continuum player, binary action, symmetric payoff games studied in this paper is essentially that of Morris and Shin (2003), which embeds most applications of global games. Szkup and Trevino (2015) and Yang (2015)[5] showed that global game uniqueness and selection results will go through essentially unchanged if players endogenously choose the precision of their private signals, in the small cost limit.

Yang (2015) emphasized that the global game information structure was *inflexible* - players were constrained to a very restricted parameterized class of information structures, with all other information structures being infeasible. Sims (2003) suggested that the ability to process information is a binding constraint, which implies - via results in information theory - that there is a bound on feasible entropy reduction. If information capacity can be bought, this suggests a cost functional that is an increasing function of entropy reduction. An attractive feature of entropy reduction treated as a cost of information is that it is flexible. But Yang (2015) showed that global game uniqueness and selection results are reversed if entropy reduction is used as a cost functional. One contribution of this paper is to reconcile these results. We show how flexible information acquisition is consistent with global game uniqueness results: the key property of the global game information structure is *not* its inflexibility, but rather the natural implicit assumption of infeasible perfect discrimination.

Our paper has implications for the widespread use of entropy reduction in economic applications. Because of its purely information theoretic foundations, this cost function is not sensitive to the labelling of states, and thus it is built in that it is as easy to distinguish nearby states as distant states. Because entropy reduction has a tractable functional form for the cost of information, it has been widely used in economic settings where it does not reflect information processing costs and where the insensitivity to the distance between states does not make sense. While this may not be important in single person decision making, this paper contains a warning about use of entropy reduction as a cost of information in strategic settings. Important papers of Hebert and Woodford (2021a) and Pomatto, Strack,

[5] The formal statement of this result appears in the working paper version, Yang (2013).

and Tamuz (2019) have recently highlighted these themes and proposed alternative micro-founded cost functionals that satisfy IPD. We will discuss these papers in Section 5.

Our results address a debate about equilibrium uniqueness and selection without common knowledge. Weinstein and Yildiz (2007) have emphasized that equilibrium selection arguments in the global games literature rely on a particular relaxation of common knowledge (noisy signals of payoffs). They show that while other natural exogenous local perturbations from common knowledge imply uniqueness, any rationalizable play may be selected depending on the perturbation. We show that an alternative perturbation - based on costly information acquisition - selects the same outcome as the global game literature but does not rely on the specific additive noisy structure of information in the global games literature, but only on infeasible perfect discrimination.

While our limit uniqueness result generalizes the global games literature initiated by Carlsson and Damme (1993), we cannot appeal to the arguments in Carlsson and Damme (1993) and later papers on binary action games because the relevant space of stochastic choice rules cannot be characterized by a threshold. Our results are closer to the argument for uniqueness in general (many action) supermodular games in Frankel, Morris, and Pauzner (2003). Here too, translation insensitivity has a crucial role, with contraction like properties giving rise to uniqueness.[6] Mathevet and Steiner (2013) highlighted the role of translation insensitivity in obtaining uniqueness results. All these papers assume noisy information structures, depend on translation insensitivity and implicitly use a more restrictive form of continuous stochastic choice. In contrast, we show that translation insensitivity leads to limit uniqueness (multiplicity) if continuous stochastic choice holds (fails) in equilibrium, and thus highlight continuous stochastic choice as the essential property that leads to the equilibrium uniqueness.

2 Setting

We first describe a canonical class of games with symmetric payoffs and strategic complementarities, parameterized by a payoff relevant state of the world. This class of games is widely used in applications and corresponds to the class studied in the survey of global games of Morris and Shin (2003). We then describe the *stochastic choice game* in which each player chooses a stochastic choice rule mapping the state space to the probability simplex of the actions, and pays a cost for the stochastic choice rule.

[6] Mathevet (2008) provided a proof of the results of Frankel, Morris, and Pauzner (2003) using the contraction mapping theorem under slightly stronger assumptions

2.1 The Base Game

A continuum of players simultaneously choose an action, "not invest" or "invest".[7] The mass of players is normalized to 1 and a generic player is indexed by $i \in [0,1]$. A player's return if she invests is $\pi(l,\theta)$, where $l \in [0,1]$ is the proportion of players investing and $\theta \in \mathbb{R}$ is a payoff relevant state. The return to not investing is normalized to 0. Note that while we find it convenient to label actions "invest" and "not invest" as an aid to comprehension, the return of 0 to "not invest" is just a normalization and, modulo this normalization, we are allowing for arbitrary continuum player, binary action, symmetric payoff games.

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

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

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

Assumption A3 (Limit Dominance): *There exist* $\theta_{\min} \in \mathbb{R}$ *and* $\theta_{\max} \in \mathbb{R}$ *such that (i)* $\pi(l,\theta) < 0$ *for all* $l \in [0,1]$ *and* $\theta < \theta_{\min}$*; and (ii)* $\pi(l,\theta) > 0$ *for all* $l \in [0,1]$ *and* $\theta > \theta_{\max}$.

Assumption A1 and A2 are basic monotonicity assumptions. A1 states that the incentive to invest is increasing in the proportion of other players who are also investing. Assumption A2 states that the incentive to invest is increasing in the state.

Assumption A3 requires that players have a dominant strategy to not invest or invest when the state is, respectively sufficiently low or sufficiently high. As in the global games literature, small information frictions will not be able to select among equilibria unless it is sometimes a dominant strategy to choose a particular action.

We need some additional assumptions of strict monotonicity and continuity. It would be enough for us to assume that $\pi(l,\theta)$ is continuous and strictly increasing in l and θ, but this would rule out important applications. Assumptions A4-A6 are weaker strictness and continuity requirements.

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

Given assumption A2, assumption A4 simply rules out the possibility that there is an open interval of θ for which $\pi(l,\theta) = 0$. Notice also that A2 and A4 imply A3 *limit dominance*. Specifically, we can define $\theta_{\min}$ and $\theta_{\max}$ by setting $\theta_{\min} = \theta_1$ and $\theta_{\max} = \theta_0$ as defined in assumption A4. Then for any $l \in [0,1]$, $\pi(l,\theta) \leq \pi(1,\theta) < 0$ for all $\theta < \theta_{\min}$, and $\pi(l,\theta) \geq \pi(0,\theta) > 0$ for all $\theta > \theta_{\max}$, i.e., limit dominance holds.

We will be especially concerned about a player's "Laplacian payoff" when he has a

[7] The continuum player assumption is for convenience. The same results go through for a finite number of players.

uniform, or "Laplacian", belief about the proportion of opponents who invest in state θ, or

$$\Pi(\theta) = \int_0^1 \pi(l, \theta)\, dl.$$

We impose two assumptions on the Laplacian payoff. First, we slightly strengthen A4 by requiring that the Laplacian payoff also satisfies state single crossing:

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

We will refer to θ^{**} as the Laplacian threshold, which will play a key role in our analysis. A player with the Laplacian belief who knows the state will invest if the state exceeds θ^{**} and not invest if the state is less than θ^{**}.

Second, we impose a continuity assumption on the Laplacian payoff (note that we impose no continuity assumption on individual payoffs).

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

Finally, we require:

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

Assumption A7 simplifies the proof but could be relaxed.

We will refer to a game satisfying the above assumptions as the base game. To illustrate the assumptions, we will describe two games widely studied in the applied literature.

The following *regime change game*[8] will be analyzed in Section 5.1.

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

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

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

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

[8] Morris and Shin (1998) introduced a more complicated version. Morris and Shin (2004) studied this stripped down version and Angeletos, Hellwig, and Pavan (2007) popularized this name.

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

The following *investment* game satisfies stronger continuity and monotonicity properties.

Example 2 ***[Investment Game]*** *Each player enjoys a return $f(\theta) - r \cdot (1 - l)$ from investing and zero otherwise, i.e.,*

$$\pi(l, \theta) = f(\theta) - r \cdot (1 - l) \ ,$$

where $f : \mathbb{R} \to \mathbb{R}$ is a continuous function that is strictly increasing, bounded, and there exist θ and θ' such that $f(\theta) \leq 0$ and $f(\theta') \geq r$.

This example satisfies assumptions A1 through A7. In particular, $\pi(l, \theta)$ is strictly increasing in θ and l and so A4 (state single crossing) holds with θ_l the unique solution to $f(\theta_l) = r \cdot (1 - l)$. The Laplacian payoff is

$$\Pi(\theta) = \int_0^1 \pi(l, \theta)\, dl = f(\theta) - \frac{r}{2} \ ,$$

so A5 (Laplacian single crossing) and A6 (Laplacian continuity) hold, with the Laplacian threshold the unique solution to $f(\theta^{**}) = \frac{r}{2}$, so that $\theta^{**} = \theta_{1/2}$.

2.2 The Stochastic Choice Game

We now define a stochastic choice game to model a situation where there is an additional component of payoffs reflecting the informational cost of the stochastic choice rule. Players share a *common prior* on θ, denoted by probability density g, which is continuous and strictly positive on $[\theta_{\min}, \theta_{\max}]$. We also assume that the common prior assigns positive probability to both dominance regions.

A generic player i chooses a *stochastic choice rule* (henceforth SCR) $s_i : \mathbb{R} \to [0, 1]$, with $s_i(\theta)$ being the player's probability of investing conditional on the state being θ. The SCR can be viewed as a binary signal experiment, where $s_i(\theta)$ is the probability of observing signal 1 and $1 - s_i(\theta)$ is the probability of observing signal 0. We write S for the set of all SCRs, which consists of all Lebesgue measurable functions that map from the real line to $[0, 1]$. Players privately and simultaneously choose their SCRs so that their actions are independent conditional on the state. As is usual, we adopt the law of large numbers convention that given a strategy profile $\{s_j\}_{j \in [0,1]}$, the proportion of players that invest when the state is θ is $\int s_j(\theta)\, dj$.[9]

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

The first component of a player's payoff in the stochastic choice game is his base game payoff

$$u\left(s_i, \{s_j\}_{j\in[0,1]}\right) = \int s_i(\theta)\,\pi\left(\int s_j(\theta)\,dj, \theta\right) g(\theta)\,d\theta\ . \tag{1}$$

The second component of a player's payoff depends only on his own SCR. This component reflects the informational cost of the SCR. A *cost functional* $c : S \to \mathbb{R}_+ \cup \{\infty\}$ maps SCRs to the extended positive real line. Here $c(s) = \infty$ will be interpreted to mean that s is not feasible. To guarantee the existence of best responses of players' decision problems, we assume that c is lower semi-continuous in SCRs.

A player incurs cost $\lambda \cdot c(s)$ if she chooses $s \in S$. We will hold the cost functional c fixed in our analysis and vary $\lambda \geq 0$, a parameter that represents the cost of information; we will refer to the resulting stochastic choice game as the λ-game. The payoff of player i in the λ-game is thus given by

$$v_\lambda\left(s_i, \{s_j\}_{j\in[0,1]}\right) = u\left(s_i, \{s_j\}_{j\in[0,1]}\right) - \lambda \cdot c(s_i)\ . \tag{2}$$

Since $u\left(s_i, \{s_j\}_{j\in[0,1]}\right)$ is linear in s_i and $c(s_i)$ lower semi-continuous, $v_\lambda\left(s_i, \{s_j\}_{j\in[0,1]}\right)$ is upper semi-continuous in s_i. Hence, each player i's best responses to her opponents strategy profile exist. Then, the Nash equilibrium of the stochastic game is defined as follows:

Definition 3 (Nash Equilibrium) *A strategy profile $\{s_j\}_{j\in[0,1]}$ is a Nash equilibrium of the λ-game if*

$$s_i \in \arg\max_{s_i'} v_\lambda\left(s_i', \{s_j\}_{j\in[0,1]}\right)$$

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

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

We equip the SCR space S with the L^1-metric, so that the distance between SCRs s_1 and s_2 is given by

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

and write $B_\delta(s)$ for the open set of SCRs within distance δ of s under this metric. This metric captures one notion of closeness that will be relevant for evaluating base game payoffs, but we emphasize that our focus will be on cost functionals that are *not* continuous in metric.

We also equip the SCR space S with a partial order $\succeq$. In particular, for any s_1 and s_2 in

S, $s_2 \succeq s_1$ if and only if $s_2(\theta) \geq s_1(\theta)$ almost surely under the common prior. Accordingly, $\vee$ and $\wedge$, the join and meet operators, take the form $[s_2 \vee s_1](\theta) = \max\{s_2(\theta), s_1(\theta)\}$ and $[s_2 \wedge s_1](\theta) = \min\{s_2(\theta), s_1(\theta)\}$, respectively. It is straightforward to see that for any s_1 and s_2 in S, both their join and meet belong to S, so that $(S, \succeq)$ forms a complete lattice.

We now describe two maintained assumptions on the cost functional $c(\cdot)$.

Intuitively, "flatter" SCRs are less informative than steeper ones and so should be cheaper. The following restriction reflects this intuition.

Assumption A8 (Submodularity): *The cost functional $c(\cdot)$ is submodular on $(S, \succeq)$; i.e., $c(s_2 \vee s_1) + c(s_2 \wedge s_1) \leq c(s_1) + c(s_2)$* for all s_1 and s_2 in S.

Here, the meet and the join are "flatter" than the original SCRs and thus jointly cheaper.

An important SCR will be the (discontinuous) step function $1_{\{\theta \geq \psi\}}$, where a player invests if and only if the state exceeds a threshold ψ. While allowing such discontinuous SCRs to be infeasible, we want to require that they can at least be approximated by feasible (i.e., finite cost) SCRs.

Assumption A9 (Feasible Almost Perfect Discrimination): *For any $\psi \in \mathbb{R}$ and $\delta > 0$, there exists an SCR $s \in S$ such that $s \in B_\delta\left(1_{\{\theta \geq \psi\}}\right)$ and $c(s) < \infty$.*

We will see that if information is costless ($\lambda = 0$), the best response for a player who expects other players to follow a step function SCR would be to choose a step function. Assumption A9 ensures that it is at least feasible to approximate this best response arbitrarily closely (even if with a continuous SCR). Without this assumption, there would be step function SCRs that cannot be approximated by finite-cost SCRs, and complete-information equilibria consisting of such SCRs are thus *exogenously* precluded in the analysis of equilibrium selection. This assumption holds for almost all cost functionals used in the literature, such as entropy reduction and those used in Hebert and Woodford (2021a) and Pomatto, Strack, and Tamuz (2019).

2.3 Pairwise-Separable Cost Functionals

We will illustrate maintained assumptions A8 (submodularity) and A9 (feasible almost perfect discrimination), and other key properties of cost functionals introduced throughout the paper with *pairwise-separable* (henceforth PS) cost functionals of the form:

$$c_{PS}(s) = \int_\theta \int_{\theta'} \left|\theta' - \theta\right|^{-\alpha} D\left(s(\theta), s(\theta')\right) h\left(\theta, \theta'\right) d\theta' d\theta \ , \tag{3}$$

where $0 \leq \alpha < \beta + 1$, $h\left(\theta, \theta'\right)$ is any density on $\mathbb{R}^2$ such that both $\frac{h(\theta,\theta')}{g(\theta)g(\theta')}$ and $\frac{g(\theta)g(\theta')}{h(\theta,\theta')}$ are bounded above, and D is a divergence function satisfying (i) $D(x_1, x_2) > 0$ if $x_1 \neq x_2$; (ii) $D(x_1, x_2) = O\left(|x_1 - x_2|^\beta\right)$ as $x_1 - x_2 \to 0$, for some $\beta \geq 1$;[10] (iii) differentiability; and

[10]Note that (ii) implies $D(x_1, x_2) = 0$ if $x_1 = x_2$. We require $\beta \geq 1$ so that popular functional forms like $|x_1 - x_2|^\beta$ satisfy decreasing differences as required by Condition (iv).

(iv) decreasing differences, i.e.,

$$\left[D\left(x_1', x_2'\right) - D\left(x_1', x_2\right)\right] - \left[D\left(x_1, x_2'\right) - D\left(x_1, x_2\right)\right] \le 0$$

if $x_1 \le x_1'$ and $x_2 \le x_2'$.

Thus PS cost functionals are the sum of terms that represent the difficulty of distinguishing pairs of states θ and θ', where those terms are the product of (i) $\left|\theta' - \theta\right|^{-\alpha}$ measuring the the intrinsic difficulty of distinguishing the two states; and (ii) the divergence $D\left(s\left(\theta\right), s\left(\theta'\right)\right)$ measuring the difference between the distributions of signals at those states. We require $\alpha < \beta + 1$ to ensure absolutely continuous SCRs have finite cost. The density h is a weight function required only to ensure that the double integral is well-defined, and our results do not depend on the specific form of this density.[11] The intuition for decreasing differences is that if $s\left(\theta'\right) \ge s\left(\theta\right)$, increasing $s\left(\theta'\right)$ better differentiates state θ' from θ, but to a lesser extent than if $s\left(\theta\right)$ also increases. Note that decreasing differences is equivalent to the assumption that $\frac{\partial^2 D(x_1, x_2)}{\partial x_1 \partial x_2} \le 0$ if D is second-order differentiable.

The pairwise-separable cost function is an adaptation to our context of the log-likelihood ratio cost function recently introduced and elegantly axiomatized by Pomatto, Strack, and Tamuz (2019). In particular, Pomatto, Strack, and Tamuz (2019) considered a finite state space, a uniform prior on the state space, a more general indistinguishability function[12] and the particular Kullback-Leibler divergence.[13] Note that we cannot have a uniform prior on $\left(\theta, \theta'\right)$ on $\mathbb{R}^2$, but our results hold independent on the density h.[14] We will discuss the axiomatic foundations for the log-likelihood ratio cost functional provided by Pomatto, Strack, and Tamuz (2019) in Section 5.2. The PS cost functional is a particularly useful example to illustrate our results since the one parameter α captures the intrinsic difficulty to distinguishing pairs of states; if $\alpha = 0$, all pairs of states are treated equally, and as α increases, it becomes increasingly harder to distinguish nearby states than distant states.

We now verify that the PS cost functional satisfies our two maintained assumptions on cost functions. Consider first A8 (submodularity). For any two SCRs s_1 and s_2, we have

$$\begin{aligned} & c_{PS}\left(s_2 \vee s_1\right) + c_{PS}\left(s_2 \wedge s_1\right) - c_{PS}\left(s_1\right) - c_{PS}\left(s_2\right) \\ = & \int_\theta \int_{\theta'} \left|\theta' - \theta\right|^{-\alpha} \left[\begin{array}{c} D\left(\max\left(s_2\left(\theta\right), s_1\left(\theta\right)\right), \max\left(s_2\left(\theta'\right), s_1\left(\theta'\right)\right)\right) \\ + D\left(\min\left(s_2\left(\theta\right), s_1\left(\theta\right)\right), \min\left(s_2\left(\theta'\right), s_1\left(\theta'\right)\right)\right) \\ - D\left(s_2\left(\theta\right), s_2\left(\theta'\right)\right) - D\left(s_1\left(\theta\right), s_1\left(\theta'\right)\right) \end{array}\right] h\left(\theta, \theta'\right) d\theta' d\theta\ , \end{aligned}$$

[11] Note that $\left|\theta' - \theta\right|^{-\alpha} \to \infty$ as $\theta' - \theta$ vanishes, while $h\left(\theta, \theta'\right)$ is of the same order of $g\left(\theta\right) g\left(\theta'\right)$. Thus h will not completely undo the functional form restrictions of (3).

[12] They also gave conditions under which the indistinguishability function was $\left|\theta - \theta'\right|^{-2}$ as we will discuss in Section 5.2.

[13] The Kullback-Leibler divergence in our binary signal case is $D_{KL}\left(x_1, x_2\right) = x_1 \ln \frac{x_1}{x_2} + \left(1 - x_1\right) \ln \frac{1 - x_1}{1 - x_2}$, and satisfies the decreasing-difference property since $\frac{\partial^2 D_{KL}(x_1, x_2)}{\partial x_1 \partial x_2} = -\frac{1}{x_2(1 - x_2)} < 0$.

[14] We also could have restricted our analysis to a bounded interval of real line, maintained uniformity of the prior on $\left(\theta, \theta'\right)$, at the cost of extra cases to deal with in our analysis.

and c_{PS} is submodular if the term of the square brackets (for each pair of θ and θ') in the integral is always non-positive. This is indeed the case because of the decreasing-difference property of D.

To establish A9 (feasible almost perfect discrimination), it is enough to identify a sequence of finite cost SCRs that approach the step function $1_{\{\theta \geq \psi\}}$. We will consider the piecewise linear approximation:

$$\widehat{s}_{k,\psi}(\theta) = \begin{cases} 0, \text{ if } \theta \leq \psi - \frac{1}{2k} \\ \frac{1}{2} + k(\theta - \psi), \text{ if } \psi - \frac{1}{2k} \leq \theta \leq \psi + \frac{1}{2k} \\ 1, \text{ if } \theta \geq \psi + \frac{1}{2k} \end{cases}, \tag{4}$$

which is illustrated in Figure 2.3. This approximation will be used throughout the paper and we will refer to $\widehat{s}_{k,\psi}$ as the "slope k threshold ψ approximation" of $1_{\{\theta \geq \psi\}}$.

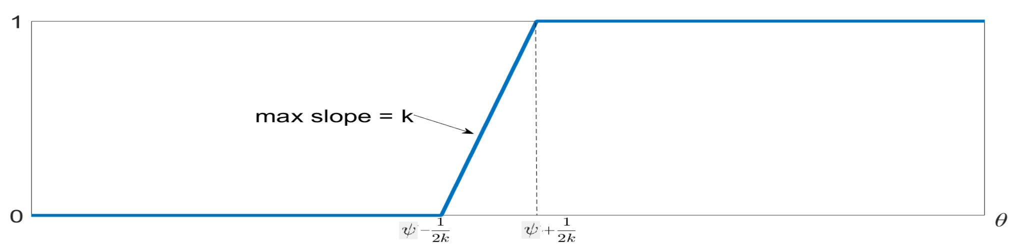

Figure 2.3: the slope k jump ψ approximation to the step function at ψ.

Note that $\left\|\widehat{s}_{k,\psi}, 1_{\{\theta \geq \psi\}}\right\|$ is of order $\frac{1}{k}$, so as $k \to \infty$, $\left\|\widehat{s}_{k,\psi}, 1_{\{\theta \geq \psi\}}\right\| \to 0$. Lemma 19 in the Online Appendix shows that $\widehat{s}_{k,\psi}$ has finite cost under the PS cost functional for all finite k and ψ.

2.4 Supermodularity and Symmetric Monotonic Equilibria

Because the first component of player i's payoff is linear in s_i and the cost functional is submodular, each player's payoff in the λ-game, $v_\lambda\left(s_i, \{s_j\}_{j\in[0,1]}\right)$, is supermodular in his own SCR s_i. Assumption A1 (strategic complementarities) implies that $v_\lambda\left(s_i, \{s_j\}_{j\in[0,1]}\right)$ has increasing differences over s_i and other players' SCRs $\{s_j\}_{j\in[0,1]}$. Hence, the stochastic choice game is a symmetric payoff supermodular game. Thus, its equilibria exist and form a sub-lattice of $(S, \succeq)$ with both the largest and the smallest equilibria being monotonic and symmetric SCRs (Van Zandt and Vives (2007)). If these two monotonic equilibria converge to a unique one as λ vanishes, so do all other equilibria (if they exist) between them. We thus establish the uniqueness among all (possibly non-monotonic and asymmetric) equilibria. Therefore, it suffices to focus on *symmetric* equilibria in *monotonic* (non-decreasing) SCRs in the analysis.[15]

Thus we will write S_M for the set of monotonic SCRs and restrict attention to monotonic SCRs in the body of the paper. In the light of this, we will adapt our notation by now writing

$$u(\widetilde{s}, s) = \int \widetilde{s}(\theta)\, \pi(s(\theta), \theta)\, g(\theta)\, d\theta$$

and

$$v_\lambda(\widetilde{s}, s) = u(\widetilde{s}, s) - \lambda \cdot c(\widetilde{s})$$

for a player's payoff from the base game and the λ-game respectively if he chooses SCR $\widetilde{s}$ and others all choose SCR s.

Definition 4 (Symmetric Monotonic Nash Equilibrium) *An SCR s is a symmetric monotonic Nash equilibrium of the λ-game if*

$$s \in \arg\max_{\widetilde{s} \in S} v_\lambda(\widetilde{s}, s)$$

and $s \in S_M$, where S_M is the set of monotonic SCRs.

In what follows, we will refer to symmetric monotonic Nash equilibria as simply "equilibria".

3 Main Result

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

Definition 5 (Laplacian Selection) *Laplacian selection occurs if, for any $\delta > 0$, there exists $\overline{\lambda} > 0$ such that $\left\| s, 1_{\{\theta \geq \theta^{**}\}} \right\| \leq \delta$ whenever s is an equilibrium of the λ-game and $\lambda \in \left(0, \overline{\lambda}\right)$.*

We have two key sufficient conditions for Laplacian selection. First:

Definition 6 (IPD) *Cost functional $c(\cdot)$ satisfies infeasible perfect discrimination (IPD) if any SCR that is not absolutely continuous has infinite cost.*

[15] We review what happens if submodularity of the cost functional fails in Section 5.1.

IPD implies that the probability of observing a given signal cannot jump discontinuously (e.g., from 0 to 1) at some θ. To do so, it would have to be feasible to perfectly discriminate between states below θ and states above θ. We consider this property to be intuitive and consistent with empirical evidence reviewed in Section 5.3.

The PS cost functional (3) provides an illustration of IPD. In the cost functional, the divergence $D\left(s\left(\theta\right), s\left(\theta'\right)\right)$, the difference between signals at each pair of states θ and θ', is weighted by $\left|\theta' - \theta\right|^{-\alpha}$ with $\alpha \geq 0$, so that it is more costly to distinguish θ and θ' if the distance $\left|\theta' - \theta\right|$ is smaller. Moreover, increasing the parameter α makes it increasingly harder to distinguish nearby states than to distinguish distant states. The pairwise-separable cost functional satisfies IPD if and only if $\alpha \geq 2$. We refer interested readers to Lemma 20 in the online Appendix for the proof of this result. We will discuss further examples of cost functionals satisfying IPD in Section 5.

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

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

Definition 7 (Translation Insensitivity) *Cost functional $c\left(\cdot\right)$ satisfies translation insensitivity if for any $\psi \in \left[\theta_{\min}, \theta_{\max}\right]$, there exist $\delta > 0$ and $K > 0$ such that, for any feasible $\widetilde{s}$ in $B_\delta\left(1_{\{\theta \geq \psi\}}\right)$, $\left|c\left(T_\Delta \widetilde{s}\right) - c\left(\widetilde{s}\right)\right| < K \cdot \left|\Delta\right|$ for all $T_\Delta \widetilde{s}$ in $B_\delta\left(1_{\{\theta \geq \psi\}}\right)$.*

This property requires that the cost responds at most linearly to translations of finite cost SCRs in a small neighborhood of any step function of interest. The definition only involes finite cost SCRs since infinitely costly (i.e., infeasible) SCRs will not be chosen in equilibrium. Note that this is a local property, which does not require a uniform bound on the responses of the cost to translations throughout the strategy space. Translation insensitivity captures the idea that the cost of information acquisition reflects the cost of paying attention to some neighborhood of the state space, but is not too sensitive to where attention is paid. We regard this as a mild restriction held by all the cost functionals discussed in this paper. For example, it is straightforward to see that the PS cost functional satisfies translation insensitivity, as the distance between any pair of states and the associated weight are invariant to translations. Now we have our main result:

Proposition 8 (IPD and Laplacian Selection) *If the cost functional satisfies infeasible perfect discrimination and translation insensitivity, then there is Laplacian selection.*

Thus when $c\left(\cdot\right)$ satisfies infeasible perfect discrimination and translation insensitivity, and the cost multiplier λ is small, all equilibria are close to the Laplacian switching strategy.

We will give some intuition for the result and the proof by sketching three key steps in the argument. Holding fixed the strategy of others $s \in S_M$, write

$$S_\lambda(s) = \arg\max_{\widetilde{s} \in S_M} v_\lambda(\widetilde{s}, s)$$

for a player's best responses in the λ-game. We will first show that, for small λ, a player's best response will approximate a step function. To see why, observe that Assumptions A1 (strategic complementarities) and A4 (state single crossing) and the monotonicity of s imply that there exists a threshold ψ such that $\pi(s(\theta), \theta) > 0$ if $\theta > \psi$ and $\pi(s(\theta), \theta) < 0$ if $\theta < \psi$. So if $\lambda = 0$, it is optimal to choose the step function with cutoff ψ, so $S_\lambda(s) = \left\{1_{\{\theta \geq \psi\}}\right\}$. Now the base game payoff loss from choosing strategy $\widetilde{s}$ instead of $1_{\{\theta \geq \psi\}}$ is given by

$$\begin{aligned} u\left(1_{\{\theta \geq \psi\}}, s\right) - u(\widetilde{s}, s) &= \int_\theta \left(1_{\{\theta \geq \psi\}}(\theta) - \widetilde{s}(\theta)\right) \pi(s(\theta), \theta)\, g(\theta)\, d\theta \\ &\leq \left\|1_{\{\theta \geq \psi\}} - \widetilde{s}\right\| \cdot \sup_\theta |\pi(s(\theta), \theta)| \ . \end{aligned}$$

By A7, the latter term is finite and so the payoff loss is of order $\left\|\widetilde{s}, 1_{\{\theta \geq \psi\}}\right\|$. But Assumption A9 (feasible almost perfect discrimination) implies that we can approximate the step function to an arbitrary degree of accuracy with a nearby SCR with finite cost. Thus as λ approaches 0, the player's payoff from his best response to s will approach his payoff when $\lambda = 0$ and his best response must approximate the step function with cutoff ψ arbitrarily closely. Hence, to show there is Laplacian selection, it suffices to show that the critical threshold ψ is close to the Laplacian threshold θ^{**} when λ is small.

Second, if we fix a strategy that is continuous but close to a step function with cutoff ψ, the base game payoff gain to a player from deviating from s to $T_\Delta s$ (when others keep playing according to s) is approximately proportional to $\Pi(\psi)$, his Laplacian payoff at ψ. We first illustrate this argument graphically and then sketch its algebraic counterpart.

Observe that s and $T_\Delta s$ result in different outcomes only when the player does not invest under s but invests under $T_\Delta s$. This event is indicated as the yellow-shaded region in Figure 3. When Δ is sufficiently close to zero, s and $T_\Delta s$ vary continuously from almost 0 to almost 1 within a sufficiently small interval $[\psi - \eta, \psi + \eta]$ in which the density $g(\theta)$ is approximately equal to $g(\psi)$. Then the probability that the mass of players investing falls into $[l, l + dl]$ when this event occurs is approximately equal to $g(\psi) \cdot \Delta \cdot dl$. Note that this probability does not depend on l (up to the approximation). Hence the distribution of l conditional on the event that s and $T_\Delta s$ induce different outcomes is a uniform one, resulting in a marginal impact on the expected return approximately equal to $g(\psi) \int_0^1 \pi(l, \psi)\, dl$.

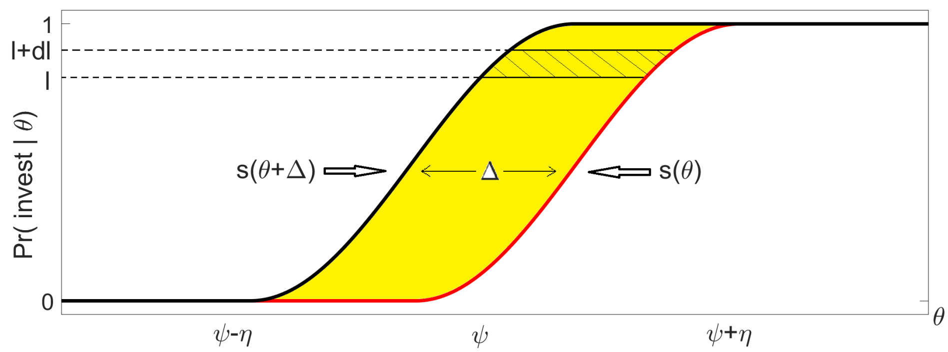

Figure 3: translation leads to Laplacian belief

To see the algebraic counterpart to this argument, note that the expected return from deviating from s to $T_\Delta s$ is

$$u\left(T_\Delta s, s\right) - u\left(s, s\right) = \int\limits_\theta \left(s\left(\theta+\Delta\right) - s\left(\theta\right)\right)\pi\left(s\left(\theta\right), \theta\right) g\left(\theta\right) d\theta \ .$$

Thus, the marginal impact on the expected return is

$$\begin{aligned}
\left.\frac{d}{d\Delta}u\left(T_\Delta s, s\right)\right|_{\Delta=0} &= \int\limits_\theta s'\left(\theta\right)\pi\left(s\left(\theta\right), \theta\right) g\left(\theta\right) d\theta \\
&= \int\limits_{l=0}^{1} \pi\left(l, s^{-1}\left(l\right)\right) g\left(s^{-1}\left(l\right)\right) dl \ ,
\end{aligned}$$

where the second equality follows the change of variables $l = s\left(\theta\right)$. Now for any $\eta > 0$, if s is close enough to a step function with cutoff ψ, we will have $\left|s^{-1}\left(l\right) - \psi\right| \leq \eta$ for all $l \in (\eta, 1-\eta)$. So as s approaches $1_{\{\theta \geq \psi\}}$,

$$\left.\frac{d}{d\Delta}u\left(T_\Delta s, s\right)\right|_{\Delta=0} \to g\left(\psi\right) \int\limits_{l=0}^{1} \pi\left(l, \psi\right) dl \ .$$

The Laplacian belief, which is uniform over all $l \in [0, 1]$, reflects a player's uncertainty about other players' action, conditional on the event that the translation has changed his action. His uncertainty takes this form because of the the continuity of the equilibrium strategy s, in the sense that all $l \in [0, 1]$ are possible. In particular, all $l \in [0, 1]$ take place with (approximately) equal probability in a small neighborhood of the cutoff ψ, and there is no sharp distinction between any pair of states close to ψ.

Now as the third step, consider any sequence of equilibria $\left\{s^\lambda\right\}$ of the λ-game. By our first step above, we know that the equilibrium strategy s^λ is close to a step function with some cutoff ψ when λ is sufficiently small. As a necessary condition for s^λ to be an equilibrium of the λ-game, it must be that it is suboptimal for a player to translate this strategy in either direction, so that

$$\frac{d}{d\Delta} v_\lambda \left(T_\Delta s^\lambda, s^\lambda\right)\Big|_{\Delta=0} = \frac{d}{d\Delta}\left(u\left(T_\Delta s^\lambda, s^\lambda\right) - \lambda \cdot c\left(T_\Delta s^\lambda\right)\right)\Big|_{\Delta=0} = 0 \ .$$

But translation insensitivity implies that $\lambda \cdot \frac{d}{d\Delta} c\left(T_\Delta s^\lambda\right)\big|_{\Delta=0} \to 0$ as $\lambda \to 0$, since the latter term is bounded. And our second step implies that

$$\frac{d}{d\Delta} u\left(T_\Delta s^\lambda, s^\lambda\right)\Big|_{\Delta=0} \to g(\psi) \int\limits_{l=0}^{1} \pi(l,\psi)\, dl$$

as $\lambda \to 0$. Hence, for translations not to be optimal as we approach the limit, we must have $\int\limits_{l=0}^{1} \pi(l,\psi)\, dl = 0$ and thus $\psi = \theta^{**}$.

To see the importance of continuity in this argument and preview the converse result in the next section, suppose instead that s was $1_{\{\theta \geq \psi\}}$, the step function that is discontinuous at cutoff ψ. Then when comparing s and $T_\Delta s$, in the event that they induce different outcomes the player is sure that l, the proportion of others investing, does not belong to $(0,1)$ and takes distinct values at any pair of states on different sides of ψ no matter how close they are. Thus payoffs for $l \in (0,1)$ cannot player a role in determining the threshold ψ, and we end up with indeterminacy and multiple equilibria.

The usual intuition for Laplacian selection in global games is very different: for any signal observed, if noise is small, a player will have a uniform belief about the proportion of players with higher signals. Now if there is a threshold equilibrium, the threshold state must be associated with a Laplacian payoff of zero. This intuition is interim, i.e., based on a player's beliefs conditional on his signal. There is no such interim stage in our model, so there is no analogous intuition in this context. On the other hand, one can give an alternative intuition for Laplacian selection in global games based on an ex ante perspective used here.

4 Tightening Results and Continuous Stochastic Choice

Our main result gave natural and interpretable sufficient conditions for Laplacian selection: the mild translation insensitivity property and infeasible perfect discrimination. But the only observable implication of IPD is that players will choose continuous stochastic choice rules, and the optimality of continuous stochastic choice is all that is needed to obtain

Laplacian selection. In this section, we identify a weaker sufficient condition for continuous stochastic choice. We also show that Laplacian selection fails and there are multiple equilibria if discontinuous stochastic choice rules are optimal, which occurs when discontinuous stochastic choice rules are sufficiently cheap relative to continuous stochastic choice rules. Thus this section shows that the (dis-)continuity of optimal stochastic choice rules is the key - and observable - property that determines uniqueness (multiplicity) of equilibria.

4.1 A Converse

Yang (2015) showed that there is limit multiplicity if the cost of information is given by entropy reduction:

$$c_{ER}(s) = \mathbb{E}\left[H(s(\theta))\right] - H(\mathbb{E}\left[s(\theta)\right]) \ ,$$

where $H : [0,1] \to \mathbb{R}$ is given by

$$H(x) = x \ln x + (1-x)\ln(1-x) \ .$$

This cost functional is of interest because it is widely used across areas of applied economics. However, this cost functional has the feature that the distance between states does not affect the cost. For illustration, suppose the prior on θ is given by $\Pr[\theta = \theta_1] = p$ and $\Pr[\theta = \theta_2] = 1-p$. Then we have

$$c_{ER}(s) = p \cdot H(s(\theta_1)) + (1-p) \cdot H(s(\theta_2)) - H(p \cdot s(\theta_1) + (1-p) \cdot s(\theta_2)) \ .$$

Observe that effectively, θ_1 and θ_2 are just labels: given the values of $s(\theta_1)$ and $s(\theta_2)$, $c_{ER}(s)$ is independent of the values of θ_1 and θ_2, and thus independent of the distance between them. Later in this section we generalize this feature to cheap perfect discrimination (CPD) and formally show in Lemma 18 in the online appendix that the entropy reduction cost functional indeed falls into this category.

Yang (2015) also reported another class of cost functionals where there is also limit multiplicity. Say that a cost functional is *Lipschitz* if it is Lipschitz continuous with respect to the metric on SCRs introduced in Section 2.2, i.e., there exists a $K > 0$ such that

$$|c(s_1) - c(s_2)| < K \cdot \|s_1, s_2\|$$

for all $s_1, s_2 \in S$.

This condition directly captures the idea that changing an SCR at a small set of states results in a cost change of the same order, even if the SCR is discontinuous. The entropy reduction cost functional is not Lipschitz. This is because $\lim_{x \to 1} H'(x) = \infty$ and

$\lim_{x \to 0} H'(x) = -\infty$, so that the marginal cost of letting $s(\theta)$ approach 1 or 0 is infinite. We will report a sufficient condition for multiplicity that covers both the Lipschitz and the entropy reduction cost functionals.

We first introduce an operation on any monotonic SCR that makes it discontinuous at ψ and closer to the step function with cutoff ψ. In particular, for any $\psi \in (\theta_{\min}, \theta_{\max})$ and $\varepsilon \in (0, 1/2)$, define an operator $L_\psi^\varepsilon : S_M \to S_M$ such that

$$\left(L_\psi^\varepsilon s\right)(\theta) = \begin{cases} \max\left(1-\varepsilon, s(\theta)\right) & \text{if } \theta \geq \psi \\ \min\left(\varepsilon, s(\theta)\right) & \text{if } \theta < \psi \end{cases}. \tag{5}$$

Note that $L_\psi^\varepsilon s$ is discontinuous at ψ and jumps by a magnitude of at least $1 - 2\varepsilon > 0$ at ψ. It better discriminates event $\{\theta \geq \psi\}$ from its complement than does s unless $s(\theta) \geq 1 - \varepsilon$ for $\theta > \psi$ and $s(\theta) \leq \varepsilon$ for $\theta < \psi$, in which case $L_\psi^\varepsilon s = s$.

Whether it is optimal to replace s by $L_\psi^\varepsilon s$ when best responding to $1_{\{\theta \geq \psi\}}$ (i.e., whether $v_\lambda\left(L_\psi^\varepsilon s, 1_{\{\theta \geq \psi\}}\right) \geq v_\lambda\left(s, 1_{\{\theta \geq \psi\}}\right)$) will depend on the trade-off between the positive impact of replacing s by $L_\psi^\varepsilon s$ and the possibly negative impact on costs of doing so.

Definition 9 (CPD) *The cost functional satisfies cheap perfect discrimination if for any $\psi \in \mathbb{R}$ and $\varepsilon \in (0, 1/2)$, there exists a $\rho > 0$ and $K > 0$ such that*

$$\left|c\left(L_\psi^\varepsilon s\right) - c(s)\right| \leq K \cdot \left\|L_\psi^\varepsilon s, s\right\| \tag{6}$$

for all monotonic $s \in B_\rho\left(1_{\{\theta \geq \psi\}}\right)$.

The change in cost going from s to $L_\psi^\varepsilon s$ is bounded above by $\lambda \cdot \left|c\left(L_\psi^\varepsilon s\right) - c(s)\right|$ while the change in the base game payoff is of order $\left\|L_\psi^\varepsilon s, s\right\|$. The CPD condition requires that the cost responds at most linearly to the operation in a neighborhood of $1_{\{\theta \geq \psi\}}$. In other words, once CPD holds, it is inexpensive to sharply discriminate nearby states.

Since CPD only requires the Lipschitz property to hold for a special operation within a small neighborhood of the step function, it is implied by the Lipschitz property. It is straightforward to verify that the entropy reduction cost functional satisfies CPD, but the proof is tedious. We relegate the formal result and the proof to Lemma 18 in the Online Appendix.

As noted above, both entropy reduction and Lipschitz cost functionals satisfy CPD. CPD will also satisfied by more general posterior separable cost functionals (Caplin, Dean, and Leahy (2021)) and the likelihood separable cost functionals analyzed by Flynn and Sastry (2021b), who show that these cost functionals are much more tractable in many action games, although they also show that results track those with entropy reduction cost functionals in binary action games.[16]

[16] Flynn and Sastry (2021a) use this cost functional in their theory and empirical work on macroeconomic attention cycles.

The importance of the distance between the states is also well illustrated by the PS cost functional, in which the cost of distinguishing each pair of states θ and θ' is weighted by $\left|\theta' - \theta\right|^{-\alpha}$ so that it is less hard to distinguish nearby states than to distinguish distant states when α is smaller. Intuitively, if $\alpha = 0$, all pairs of states are treated identically in the cost functional and thus CPD is satisfied. We argue in the Online Appendix that CPD is satisfied when $\alpha = 0$ and if it is satisfied for some $\underline{\alpha} > 0$, it is also satisfied for $\alpha \in [0, \underline{\alpha}]$.

Proposition 10 *If cost functional $c(\cdot)$ satisfies cheap perfect discrimination, then for any $\theta^* \in (\theta_{\min}, \theta_{\max})$ and $\varepsilon \in (0, 1/2)$, there exists $\overline{\lambda} > 0$ such that whenever $\lambda \in \left[0, \overline{\lambda}\right]$, there is a symmetric equilibrium s^*_λ with $s^*_\lambda(\theta) \geq 1 - \varepsilon$ for all $\theta \geq \theta^*$ and $s^*_\lambda(\theta) \leq 1 - \varepsilon$ for all $\theta < \theta^*$.*

The proposition states that if CPD holds, for any threshold $\theta^* \in (\theta_{\min}, \theta_{\max})$, as λ vanishes, there is a sequence of equilibria uniformly converging to $1_{\{\theta \geq \theta^*\}}$, which is the SCR that perfectly discriminates event $\{\theta \geq \theta^*\}$ from its complement. Hence, the λ-game has infinitely many equilibria when λ is sufficiently small, a multiplicity result in sharp contrast to the limit unique equilibrium obtained in Proposition 8.

CPD requires that the incremental cost of choosing a discontinuous SCR $L^\varepsilon_\psi s$ over a continuous rule s is at most proportional to the incremental base game payoff and thus is negligible at small λ, making $L^\varepsilon_\psi s$ a better SCR than s.[17] Knowing that others' SCR jumps at ψ radically reduces the strategic uncertainty a player faces when ε is small (so that the jump is large). In particular, now a player is pretty sure that l, the fraction of others investing, exceeds $1 - \varepsilon$ for states above threshold ψ and otherwise falls below ε, making his base game payoff cross zero at exactly ψ. Since CPD holds, he then also prefers such an SCR that jumps at ψ from below ε to above $1 - \varepsilon$, confirming the equilibrium. This logic applies to all thresholds within $(\theta_{\min}, \theta_{\max})$ and results in multiple equilibria.

4.2 Continuous Stochastic Choice and a Strengthening of the Main Result

The proof of Proposition 8 relies on the property that SCRs used in equilibrium are continuous. The analysis of CPD cost functionals in the previous section provides a partial converse, showing that discontinuous equilibrium stochastic choice implies multiplicity. Infeasible perfect discrimination was a natural and interpretable property that immediately implies that continuous SCRs will be chosen in equilibrium. However, the proof of Proposition 8 makes clear that it is enough that continuous SCRs be chosen in equilibrium even if discontinuous SCRs are feasible. In this section, we introduce a condition - *expensive perfect discrimination* (EPD) - which is weaker than IPD but also sufficient for Laplacian

[17] Note that CPD is a weak restriction in our stochastic choice game in the sense that it just requires inequality (6) to hold in a small neighborhood of the step function rather than globally.

selection because it ensures continuous SCRs are chosen in equilibrium. This result thus helps close the significant gap between IPD and CPD.

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

Instead of precluding discontinuous SCRs by assigning infinite costs, EPD requires that it is cheap to approximate such SCRs with absolutely continuous ones relative to the degree of approximation. To see the intuition, consider a PS cost functional with $\alpha \in (1, 2)$ and $D(x_1, x_2) = |x_1 - x_2|$, under which we approximate $s_1 = 1_{\{\theta \geq 0\}}$ with its k-slope 0-threshold approximation (described in equation 4 and illustrated in Figure 2.3). Assume a uniform common prior over $[-A, A]$ for some $A > 0$. Straightforward (but tedious) calculation shows that as $k \to \infty$, $c(s_1) - c(s_2)$ and $\|s_1, s_2\|$ are of the order of $k^{\alpha-2}$ and k^{-1}, respectively. Since $\alpha < 2$, we know that $c(s_1) < \infty$ (this proved as Lemma 20 in the Online Appendix) so that s_1 is feasible, but $c(s_1) - c(s_2)$, the cost saving from using s_2 over s_1, converges slower to zero than does $\|s_1, s_2\|$, the degree of approximation, because $\alpha > 1$.

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

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

Proof. Suppose $s_1 \in S_\lambda(s)$ is not absolutely continuous. Then we can find an absolutely continuous s_2 such that

$$c(s_1) - c(s_2) > \frac{\overline{\pi}}{\lambda} \|s_1, s_2\| \ ,$$

where $\overline{\pi}$ is the uniform bound on $|\pi(l, \theta)|$. Then, the gain from replacing s_1 by s_2 is

$$\begin{aligned} & v_\lambda(s_2, s) - v_\lambda(s_1, s) \\ = & \int [s_2(\theta) - s_1(\theta)] \cdot \pi(s(\theta), \theta) \cdot g(\theta) \, d\theta + \lambda \cdot [c(s_1) - c(s_2)] \\ > & -\int |s_2(\theta) - s_1(\theta)| \cdot \overline{\pi} \cdot g(\theta) \, d\theta + \overline{\pi} \cdot \|s_1, s_2\| \\ = & \ 0, \end{aligned}$$

which contradicts the optimality of s_1. ■

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

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

This proposition in a strengthening of Proposition 8 - showing the same conclusion under a weaker assumption. But given Lemma 12 establishing that discontinuous stochastic choice rules are not chosen in equilibrium, the proof of 8 applies.[18]

5 Discussion

5.1 Submodular Costs and Monotonicity

All the results in this paper continue to hold without Assumption A8 (submodularity) *if* one restricts attention to equilibria in monotonic SCRs (symmetry follows from the fact we have a continuum of players). Submodularity of the cost functional was used only to establish supermodularity of the game, which in turn implies that "monotonic Laplacian selection" (i.e., Laplacian selection restricted to monotonic SCRs) implies Laplacian selection (i.e., Laplacian selection in all SCRs, as defined in Definition 5). If there were an independent argument establishing that the existence of largest and smallest monotonic equilibria, all our results would hold unchanged. One setting where this is true is in the work of Szkup and Trevino (2015) and Yang (2015) on global games with endogenous precision. As we will discuss in Section 5.2.3, the additive noise cost functional implicit in these works actually fails submodularity, but a separate argument establishes that we can restrict attention to monotonic strategies. Thus our results generalize these results for global games, modulo the different argument establishing that we can restrict attention to monotonic strategies.

In fact, we do not have general results or even examples establishing the necessity of submodularity of cost functionals for (global) Laplacian selection. An intuition why it might not be necessary comes from the observation that in the limit as $\lambda \to 0$, the payoffs of the game approaches supermodularity. Therefore, we conjecture that extra strong continuity assumptions *might* establish Laplacian selection even without submodular costs, but would be difficult to prove at the level of generality of the cost functionals we allow.

While we believe that Assumption A8 is intuitive, and well motivated in the body of the paper by the PS cost functional, there also exist cost functionals of interest that fail submodularity. In this section, we introduce one such cost functional, establish Laplacian selection (without restricting to monotonic SCRs), and provide a closed form solution of independent interest.

[18] The proposition would continue to hold if we replaced the EPD with the following local stochastic continuous choice property: for all $\psi \in (\theta_{\min}, \theta_{\max})$, there exists $\delta > 0$ such that $\arg\max_{\widetilde{s}} v_\lambda(\widetilde{s}, s)$ consists only of absolutely continuous functions for all $s \in B_\delta\left(1_{\{\theta \geq \psi\}}\right)$ and $\lambda \in \mathbb{R}_{++}$.

Example 14 ***[Max Slope Cost Functional]*** *The cost c_{slope} of an SCR $s \in S$ is $f(k)$, where k is the maximum slope of s and $f : \mathbb{R}_+ \cup \{\infty\} \to \mathbb{R}_+ \cup \{\infty\}$ is weakly increasing with $f(0) = 0$ and $f(k) < \infty$ for all $k \in \mathbb{R}_+$. Thus*

$$c_{slope}(s) = f\left(\sup_{\theta} \ |s'(\theta)|\right) . \tag{7}$$

If s is discontinuous at θ, then $s'(\theta)$ is understood to be infinity, and the cost is $c_{slope}(s) = f(\infty)$.[19]

The max slope cost functional satisfies Assumption A9 (feasible almost perfect discrimination), since the k-slope ψ-threshold SCR $\widehat{s}_{k,\psi}$ has finite cost and approaches $1_{\{\theta \geq \psi\}}$ as $k \to \infty$. Since translation does not change the slope of SCRs, the max slope cost functional satisfies translation insensitivity. It also satisfies infeasible perfect discrimination (IPD) when $f(\infty) = \infty$. The max slope cost functional is not submodular in SCRs.[20]

The slope is a natural measure of local attention to the state. In a very different class of evolutionary models, Robson (2001), Rayo and Becker (2007) and Netzer (2009), slope is used as a measure of attention in an analogous way.[21] We will discuss in the next section how the max slope cost functional is closely related the additive noise cost functional in global games and how the micro-founded Fisher cost functional also depends on the slope of the SCR, albeit in a richer way. In Morris and Yang (2016a), we discuss a rich parameterized class of cost functionals that are increasing in an average of the derivative of the SCR, where we can give explicit expressions for when the cost functional satisfies CPD, EPD and IPD.

The next proposition show that Laplacian selection holds in the regime change game (Example 1) with the max slope cost function, even when do *not* restrict attention to monotonic SCRs.

Proposition 15 *If the max slope cost functional satisfies IPD and the base game is the regime change game (Example 1), then for any $\delta > 0$, there exists $\overline{\lambda} > 0$ such that $\left\|s, 1_{\{\theta \geq t\}}\right\| \leq \delta$ whenever s is an equilibrium of the λ-game and $\lambda \in \left(0, \overline{\lambda}\right)$.*

Our proof of Proposition 15 relies on the fact that the simple functional form of the cost gives rise to piecewise linear SCRs in equilibrium, which allow piece-by-piece perturbations that make non-monotonic SCRs suboptimal.

The max slope cost function has the attraction that there is a simple closed form solution away from the limit. Morris and Yang (2016a) solve for equilibria with regime change game payoffs in the stochastic choice game with the max slope cost functional and generalizations of it. The following proposition is one implication of this analysis.

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

[20] An example in the Online Appendix that establishes the failure of submodularity of the additive noise cost functional also establishes the failure of submodularity of the max slope cost functional.

[21] We are grateful to a referee for pointing out this connection.

Proposition 16 *If the prior G is uniform on states $[0,1]$ and $f(\infty)=\infty$, then there exists a $\overline{\lambda}>0$ such that for all $\lambda \in \left(0,\overline{\lambda}\right)$, the λ-game with the regime-change payoff (Example 1) has a unique equilibrium $\widehat{s}_{k,\xi}$, where*

$$\xi = t + \left(t - \frac{1}{2}\right) k^{-1} \ ,$$

and

$$k = \underset{\widetilde{k}>0}{\operatorname{argmax}} \ \frac{1}{2}\left[G(1) - G(0)\right] t(1-t) \cdot \widetilde{k}^{-1} - \lambda \cdot f\left(\widetilde{k}\right) \ .$$

Moreover, in equilibrium the regime changes at threshold $\psi = t$.

Thus in the special case where $f(k) = k^{\gamma}$ with $\gamma \geq 1$, we have

$$k = \left[\frac{\widehat{g}t(1-t)}{2\gamma\lambda}\right]^{1/(1+\gamma)} \ .$$

5.2 Micro-Foundations For Cost Functionals

Our key properties of cost functionals (IPD, EPD and CPD) are rather abstract. In recent years, there has been much interest in establishing tractable parametric cost functionals of information that have natural micro-foundations. Here we review cost functionals with such micro-founded foundations and discuss how they relate to our abstract conditions.

5.2.1 Axiomatic Foundations

We have used pairwise-separable cost functionals to illustrate our results throughout the paper. As we noted above, these are variations on a class of cost functionals elegantly axiomatized by Pomatto, Strack, and Tamuz (2019). They propose four natural axioms on the cost of information: (1) Blackwell-consistency, (2) linearity in independent experiments, (3) linearity in probability, and (4) continuity. Blackwell-consistency requires that any experiment that is weakly more Blackwell informative than another is weakly more expensive. The linearity assumptions capture the idea that there is a constant cost of information (and imply that the cost of zero information is zero). They provide a representation for cost functionals satisfying those four axioms. The cost of information can be written as a weighted sum - across pairs of states - of the Kullback-Leibler divergences of the distributions over signals for those two states. The weight on a pair of states is naturally interpreted as the difficulty of distinguishing that pair of states.

Their representation is for a finite state model. As discussed earlier, our pairwise-separable cost functionals are inspired by this axiomatization. In particular, we assume the state space to be the real line and apply the functional form of Pomatto, Strack, and Tamuz (2019). We consider the case where the difficulty of distinguishing states θ and θ'

takes the form $(\theta-\theta')^{-\alpha}$ and allow the divergence between signal distributions to be given by an arbitrary decreasing-difference divergence (like the Kullback-Leibler divergence).

Pomatto, Strack, and Tamuz (2019) also provide a microfoundation for the difficulty function $(\theta-\theta')^{-2}$ by imposing additional axioms across cost functionals defined on arbitrary finite subsets of the real line. And they establish that there is a Lipschitz continuous choice property in all finite state models in this case. Thus our observation that - when $\alpha=2$ - there is infeasible perfect discrimination and thus continuous choice with the PS cost functional is a close analogue of their result in a continuous state analogue of their setting.[22]

5.2.2 Sequential Learning Foundations

An alternative foundation for cost functionals, pioneered by Hebert and Woodford (2021b), is to study cost functionals that arise from optimal stopping in a dynamic information acquisition problem. Strack (2016) considers what happens if players observe an exogenous drift diffusion process, with drift corresponding to the continuous state, where the cost is the expected expected stopping time. If the stopping time is bounded, Strack (2016) observes that the induced optimal SCR will always be continuous. This note does not discuss the cost of the stopping time, but *any* cost functional that gives rise to a bounded stopping time would thus have to satisfy EPD or IPD.[23]

Hebert and Woodford (2021a) introduce Fisher cost functionals that are consistent with sequential learning foundations when players choose what to learn and respect a neighborhood structure on states. And they explicitly construct a continuous state limit to derive a cost functional which, restricted to our SCRs, is given by

$$c_{Fisher}(s)=\int \frac{\left([g(\theta)\,s(\theta)]'\right)^2}{g(\theta)\,s(\theta)}+\frac{\left([g(\theta)\,(1-s(\theta))]'\right)^2}{g(\theta)\,(1-s(\theta))}d\theta\ .$$

This cost functional, like the max slope cost functional, depends on the slope of the cost functional, and thus satisfies IPD. It is submodular (in fact, also supermodular) as we establish in the Online Appendix. It is easy to verify that it satisfies A9 (feasible almost perfect discrimination) and translation insensitivity. So Laplacian selection holds.

[22] A subtlety is that they establish their continuous stochastic choice property when $\alpha>0$; in our setting $\alpha\geq 2$ is necessary for IPD but continuous stochastic choice may also arise under EPD and thus for some values of $a<2$. The apparent discrepancy is resolved if we normalize their finite state cost functional by the number of states when taking the limit to infinitely many states. Their sufficient condition continuous stochastic choice for $\alpha>0$ is then analogous to our sufficient condition for $\alpha\geq 2$.

[23] Morris and Strack (2018) do derive an explicit cost functional in the finite state version of this problem. However, with more than two states many distributions of posteriors are not feasible.

5.2.3 Additive Noise Foundations, Global Games and the Max Slope Cost Functional

We restricted information acquisition throughout to SCRs (which are equivalent to binary signal experiments). However, we noted in the introduction that given a cost functional on a richer class of experiments, we could always define an induced cost functional on SCRs to be the cost of the cheapest experiment that allows the player to replicate the SCR. We will discuss an example of this approach both to illustrate why focussing on SCRs is without loss of generality and to describe the relation to the global games literature.

Suppose that a player could observe a signal of the true state of the world $x = \theta + \frac{1}{k}\varepsilon$, where ε is distributed on $\mathbb{R}$ according to cumulative distribution function H, at cost $\widehat{c}(k)$, where $\widehat{c}(0) = 0$, $\widehat{c}(k)$ is increasing in k, and $\widehat{c}(k) \to \infty$ as $k \to \infty$. We will loosely refer to k as the precision of the signal and thus $\widehat{c}$ as the cost of precision. This was the cost of information in the global game results with costly private signals in Szkup and Trevino (2015) and Yang (2013).

Now suppose a player chooses behavioral strategy $b : \mathbb{R} \to [0, 1]$, with the interpretation that he chooses action 1 with probability $b(x)$ if the signal realization is x, and write B for the set of behavioral strategies. Thus fixing H, a signal of precision k and a behavioral strategy b induce an SCR

$$\widetilde{s}_{k,b}(\theta) = \int b\left(\theta + \frac{1}{k}\varepsilon\right) h(\varepsilon)\, d\varepsilon,$$

where h is the probability density function of ε. Now we define the induced additive noise cost functional on SCRs to be the cost of smallest precision that would allow that SCR to be induced in this way by some behavioral strategy.

$$c_{AN}(s) = \widehat{c}\left(\inf\left\{k \in \mathbb{R}_+ \,|\, s = \widetilde{s}_{k,b} \text{ for some } b \in B\right\}\right) .$$

Clearly, this cost functional satisfies IPD. We show in the Online Appendix that it is not submodular.

It turns out that there is a tight connection between the additive noise cost functionals and max slope cost functionals introduced in Section 5.1. Consider the special case of the additive noise cost functional where the distribution of noise is uniform on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Let b_ψ be the threshold behavioral strategy where a player chooses action 1 if and only if his signal is greater than ψ. Observe that in this case $\widetilde{s}_{k,b_\psi}$ is equal to $\widehat{s}_{k,\psi}$, the k-slope ψ-threshold approximation of the step function $1_{\{\theta \geq \psi\}}$. Now if the cost $\widehat{c}$ of precision in the uniform additive noise cost functional is the same as the cost f of the maximum slope in the max slope cost functional, the uniform additive noise cost functional will agree with the max slope cost functional discussed earlier on k-slope approximations. In fact, since threshold strategies end up being optimal in global games and k-slope approximations end up being optimal under the max slope cost functional, equilibrium analysis with the max slope cost functional and the uniform additive noise cost functional end up being identical.

5.2.4 Information Processing

Sims (2003) proposed modelling information processing constraints as reflecting limited capacity for processing information, subject to optimal coding. Results in information theory establish that entropy reduction measures the size of the channel required to process the information. Sims (2003) proposed endogenizing the choice of information structure by assuming that the decision maker was able to optimally choose the information structure subject to a constraint on entropy reduction. Such problems will naturally have a multiplier associated with the entropy reduction constraint. There is a natural dual to this problem where there is no constraint but a cost of entropy reduction corresponding to the multiplier. Entropy reduction has come to be widely used in applied economics in the last twenty years, even in cases where the cost of information does not literally refer to information processing cost. In those applications, the cost of information effectively captures any cost accrued to the decision maker from distinguishing states to facilitate state-contingent strategies. For example, besides physical information processing cost, it also includes mental cost of perceiving states.

The relevant information theory is developed in the case of a finite number of states and the coding theory foundations build in the property that state labels do not matter. Many economic applications extend the entropy reduction cost functional to infinite states, but the cost retains the property that state labels do not matter, and there is no special difficulty with choosing discontinuous SCRs. However, in many applications, states refer to values of particular economic parameters, where distant values are easier to distinguish than nearby ones either objectively in empirical investigations or subjectively in mental perception. We discuss evidence on this in the next section.

5.3 Evidence on Continuous Stochastic Choice and Laplacian Selection

This paper shows that if IPD holds, there is Laplacian selection. The key property and observable implication of IPD is continuous stochastic choice. In this section, we report experimental evidence in favor of continuous stochastic choice; and evidence that it gives rise to Laplacian selection, in an experiment designed to test out theory. We discuss each in turn.

Jazayeri and Movshon (2007) examine decision makers' ability to discriminate the direction of dots on the screen when they face a threshold decision problem. There is evidence that subjects are better at discriminating states on either side of the threshold, consistent with optimal allocation of scarce resources to discriminate. However, the ability to discriminate between states on either side of the threshold disappears as we approach the threshold, giving rise to continuous choice in our sense in this setting. The allocation of perceptual resources in this case is presumably at an unconscious neuro level.[24]

Caplin and Dean (2015) introduce an experimental framework where the allocation of resources (time) distinguish states is presumably a conscious choice. Subjects observe a screen with a mixture of balls of different colors, say red and blue. Subjects are asked to distinguish whether there are more red balls or blue balls. Subjects are better at distinguishing these two states when there is more at stake. Dewan and Neligh (2020) use this framework to address the relative cost of distinguishing nearby and distant states. They verify that it is harder to distinguish two mixtures of red and blue balls if the proportions are closer, verifying the key qualitative assumption in this paper. On the other hand, Dean and Neligh (2019) consider an alternative treatment where players are asked to distinguish which letter appears the most on a screen full of letters. There is arguably no natural order on states in this setting and there is less evidence that any pair of states are easier to distinguish than any other pair.[25] This is thus a finite state analogue of the cheap perfect discrimination property.

A recent paper of Goryunov and Rigos (2020) conducts a laboratory experiment to test our predictions and finds supportive results. The experiment uses the Line Treatment (LT) and the No Line Treatment (NLT) to generate a player's discontinuous and continuous SCRs, respectively. In the experiment, each randomly matched pair of subjects following the same treatment play a coordination game identical across pairs. Each subject first chooses a cutoff, marked by a vertical line across the horizontal state space on his/her computer screen. Then the state realizes and appears as a dot on the screen. When the dot appears, the vertical line shows only on the screen of a LT subject, enabling him/her to tell whether the state is above or below his/her chosen cutoff (i.e., to the right or left of the vertical line). For an NLT subject, his/her chosen vertical line does not show on the screen when the dot appears so that he/she may not be able to sharply tell whether the state is above or below his/her chosen cutoff. A subject's chosen cutoff indicates his/her intention to implement a strategy that takes the risky action (i.e., "invest" in our context) for the state below the cutoff and the safe action (i.e., "not invest") otherwise. A strategy is continuous if it induces a probability (proxied by frequency in the experiment) of taking risky action as a function of the state is continuous in the state. Intuitively, with the help of the vertical line, the LT (NLT) subjects are more (less) likely to implement discontinuous

[24] See Pomatto, Strack, and Tamuz (2019) and Hebert and Woodford (2021a) for further discussion of this evidence.

[25] This treatment was originally developed in work of Dean, Morris and Trevino on "Endogenous Information Structures and Play in Global Games."

strategies. It is shown that LT indeed induces discontinuous strategies, while NLT does not; and NLT does induce switching cutoffs close to the unique Laplacian selection equilibrium, consistent with our theory.

5.4 Alternative Interpretations of the Cost of Stochastic Choice Rules

Stochastic choice arises in our model because players have imperfect information about the state. With this motivation for costly stochastic choice rules, it would be natural to impose the assumptions that (i) the cost of any constant stochastic choice rule is zero; and (ii) the cost is weakly increasing in the Blackwell more informative order. But we did not assume these properties because our results did not require them.

Stochastic choice in game theoretic settings has been studied in a variety of other contexts (where assumptions (i) and (ii) might not be relevant). Stochastic choice arising from payoff perturbations were introduced in Harsanyi (1973) and is studied in important literatures on stochastic fictitious play (Fudenberg and Kreps (1993)) and quantal response equilibria (McKelvey and Palfrey (1995)). van Damme (1983) assumed that players faced control costs in reducing noise in their action choices. Flynn and Sastry (2021b) recently analyze stochastic choice games focussing on these interpretations, and with a likelihood separable cost functional failing property (i) above.

The standard modelling assumption in all these cases is that there is randomness in action choice that is independent across states. If there is a separable cost associated with reducing randomness in each state, these problems will fit the framework of this paper. However, the cost functional will satisfy cheap perfect discrimination.

If the cost of controlling actions at one state depended on the control at nearby states, the possibility of EPD and IPD cost functionals and Laplacian selection. However, we do not pursue alternative interpretations in this paper, but merely note that our results continue to apply with natural meaning under alternative interpretations.

5.5 Theoretical Extensions

5.5.1 Learning about Others' Information

A maintained assumption in our analysis is that players privately acquire information about the state only. Denti (2020) and Hoshino (2018) have initiated a literature on what would happen if players could also acquire information about others' information (and thus their actions).

Denti (2020) shows that - in a model with finite players and an entropy reduction cost function - Laplacian selection is restored if players can learn about others' information. A key property of entropy reduction is that it prevents optimal SCRs from attaining 0 or 1 as the marginal cost of doing so is infinite. So while players have an incentive to learn others' information, their equilibrium stochastic choice rules are continuous, they will not be able to perfectly coordinate on the state. This allows limit uniqueness arguments to go through. But if the information cost satisfies the Lipschitz property (also satisfying CPD) or there were a continuum of players, limit multiplicity would be restored. Under a Lipschitz cost functional, players could choose step functions like $1_{\{\theta \geq \psi\}}$, perfectly correlating their actions. With a continuum of players, aggregate actions/signals could perfectly reveal the state even if individual actions/signals did not perfectly reveal the state. Either way, the game reduces to the one studied in Section 4.1 and limit multiplicity would follow.

Hoshino (2018) shows that for some assumptions on the cost of information, there is limit uniqueness but any equilibrium behavior from the underlying complete information game can be uniquely selected if the information cost is chosen appropriately. Hoshino (2018) assumes finite states and does not introduce a distance between states, so there is no role for building in the idea that nearby states are hard to distinguish. In this sense, infeasible/expensive perfect discrimination is ruled out.

5.5.2 Potential Games

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

References

Angeletos, G.-M., C. Hellwig, and A. Pavan (2007). Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks. *Econometrica 75*(3), 711–756.

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

Caplin, A. and M. Dean (2015, July). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review 105*(7), 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2021). Rationally inattentive behavior: Characterizing and generalizing shannon entropy. Technical report, New York University, Columbia University and University of Michigan.

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

Dean, M. and N. Neligh (2019). Experimental tests of rational inattention. Technical report, Columbia University and Chapman University.

Denti, T. (2020). Unrestricted information acquisition. Technical report, Cornell University.

Dewan, A. and N. Neligh (2020). Estimating information cost functions in models of rational inattention. *Journal of Economic Theory 187.*

Flynn, J. and K. Sastry (2021a). Attention cycles. Technical report, M.I.T.

Flynn, J. and K. Sastry (2021b). Strategic mistakes. Technical report, M.I.T.

Frankel, D. M., S. Morris, and A. Pauzner (2003). Equilibrium selection in global games with strategic complementarities. *Journal of Economic Theory 108*(1), 1–44.

Fudenberg, D. and D. M. Kreps (1993). Learning mixed equilibria. *Games and Economic Behavior 5*(3), 320 – 367.

Goryunov, M. and A. Rigos (2020). Discontinuous and continuous stochastic choice and coordination in the lab. Technical report, Lund University.

Harsanyi, J. C. (1973, Dec). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory 2*(1), 1–23.

Hebert, B. and M. Woodford (2021a). Neighborhood-based information costs. *American Economic Review forthcoming.*

Hebert, B. and M. Woodford (2021b). Rational inattention when decisions take time. Technical report, Stanford University and Columbia University.

Hoshino, T. (2018). Rational inattention in games. Technical report, Pennsylvania State University.

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

Mathevet, L. (2008). A contraction principle for finite global games. *Economic Theory 42*(3), 539.

Mathevet, L. and J. Steiner (2013). Tractable dynamic global games and applications. *Journal of Economic Theory 148*(6), 2583 – 2619.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*(1), 6 – 38.

Monderer, D. and L. S. Shapley (1996). Potential games. *Games and Economic Behavior 14*(1), 124 – 143.

Morris, S. and H. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L. Hansen, and S. Turnovsky (Eds.), *Advances in Economics and Econometrics: Proceedins of the Eight World Congress of the Econometric Society*, pp. 56–114. Cambridge: Cambridge University Press.

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 (2004). Coordination risk and the price of debt. *European Economic Review 48*(1), 133–153.

Morris, S. and P. Strack (2018). The wald problem and the equivalence of sequential sampling and static information costs. Unpublished working paper, Princeton University and University of California at Berkeley.

Morris, S. and T. Ui (2005). Generalized potentials and robust sets of equilibria. *Journal of Economic Theory 124*(1), 45 – 78.

Morris, S. and M. Yang (2016a). Coordination and continuous choice. Technical report, Princeton University and Duke University; https://ssrn.com/abstract=2889861.

Morris, S. and M. Yang (2016b). Coordination and the relative cost of distinguishing nearby states. Technical report, Princeton University and Duke University; https://ssrn.com/abstract=2889861.

Netzer, N. (2009). Evolution of time preferences and attitudes toward risk. *The American Economic Review 99*(3), 937–955.

Pomatto, L., P. Strack, and O. Tamuz (2019). The cost of information. Unpublished working paper, California Institute of Technology and University of California at Berkeley.

Rayo, L. and G. Becker (2007). Evolutionary efficiency and happiness. *Journal of Political Economy 115*(2), 302–337.

Robson, A. J. (2001). The biological basis of economic behavior. *Journal of Economic Literature 39*(1), 11–33.

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

Strack, P. (2016). A note on the smoothness of choice probabilities in the ddm model. Available at SSRN: https://ssrn.com/abstract=3297255.

Sun, Y. (2006). The exact law of large numbers via fubini extension and characterization of insurable risks. *Journal of Economic Theory 126*(1), 31 – 69.

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

van Damme, E. (1983). *Refinements of the Nash Equilibrium Concept*. Springer.

Van Zandt, T. and X. Vives (2007). Monotone equilibria in bayesian games of strategic complementarities. *Journal of Economic Theory 134*(1), 339 – 360.

Weinstein, J. and M. Yildiz (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica 75*(2), 365–400.

Woodford, M. (2009). Information-constrained state-dependent pricing. *Journal of Monetary Economics 56*(Supplement 1), S100–S124.

Yang, M. (2013). Coordination with flexible information acquisition. Working paper 2013, Duke University.

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

6 Appendix

To prove Proposition **8**, we will first introduce some notation and a lemma about best responses. As argued in Subsection 2.4, it suffices to focus on symmetric equilibria in monotonic SCRs, which are best responses among monotonic strategies to themselves. To study these equilibria, we write

$$S_\lambda(s) = \underset{\widetilde{s}\in S_M}{\arg\max}\, v_\lambda(\widetilde{s}, s)$$

for a player's best monotonic response in the λ-game if he thinks that other players will follow strategy $s \in S_M$. Assumptions A1 and A4 imply, for any monotone SCR s, that there exists a threshold $\theta_s \in \mathbb{R}$ such that $\pi(s(\theta), \theta) > 0$ if $\theta > \theta_s$ and $\pi(s(\theta), \theta) < 0$ if $\theta < \theta_s$. We will show that it is optimal for players to choose strategies that are close to a step function jumping at θ_s when the cost parameter λ is small.

Lemma 17 (Optimal Best Responses) *The essentially unique best response to s if $\lambda = 0$ is a step function at θ_s, i.e.,*

$$S_\lambda(s) = \left\{1_{\{\theta\geq\theta_s\}}\right\} \ .$$

Moreover, for any $\rho > 0$, there exists a $\overline{\lambda} > 0$ such that $S_\lambda(s) \subset B_\rho\left(1_{\{\theta\geq\theta_s\}}\right)$ for all $s \in S_M$ and $\lambda < \overline{\lambda}$.

Proof. When $\lambda = 0$, the player can choose any SCR for free and the optimal SCR is $1_{\{\theta\geq\theta_s\}}$.

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

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

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

$$|\pi(s(\theta), \theta)| \geq |\pi(s(\theta_s), \theta)| \geq z(\delta) \ , \tag{8}$$

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

Let $\overline{g} = \sup_{\theta \in \mathbb{R}} g(\theta) < \infty$ and choose $\delta = \frac{\rho}{4\overline{g}}$. For any $s_\lambda \in S_\lambda(s)$, note that

$$\begin{aligned}
& \int_{-\infty}^{\infty} \pi(s(\theta), \theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - s_\lambda(\theta)\right] g(\theta)\, d\theta \\
\geq \; & \int_{\theta \notin [\theta_s - \delta, \theta_s + \delta]} \pi(s(\theta), \theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - s_\lambda(\theta)\right] g(\theta)\, d\theta \\
\geq \; & \int_{-\infty}^{\infty} z(\delta) \cdot \left|1_{\{\theta \geq \theta_s\}} - s_\lambda(\theta)\right| g(\theta)\, d\theta - \int_{\theta_s - \delta}^{\theta_s + \delta} z(\delta) \cdot \left|1_{\{\theta \geq \theta_s\}} - s_\lambda(\theta)\right| g(\theta)\, d\theta \\
\geq \; & z(\delta) \cdot \left[\left\|1_{\{\theta \geq \theta_s\}}, s_\lambda\right\| - 2 \cdot \overline{g} \cdot \delta\right] = z(\delta) \cdot \left[\left\|1_{\{\theta \geq \theta_s\}}, s_\lambda\right\| - \frac{\rho}{2}\right] \; . \qquad (9)
\end{aligned}$$

The first inequality holds since $\pi(s(\theta), \theta)$ and $\left[1_{\{\theta \geq \theta_s\}} - s_\lambda(\theta)\right]$ always have the same sign and thus

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

and the second inequality follows (8).

Let $\varepsilon = z\left(\frac{\rho}{4\overline{g}}\right) \cdot \frac{\rho}{4\overline{\pi}}$, where $\overline{\pi}$ is the uniform upper bound of $\pi(l, \theta)$. For each $\theta_l \in [\theta_{\min}, \theta_{\max}]$, according to Assumption A9, there exists an SCR $\widetilde{s}_l \in B_\varepsilon\left(1_{\{\theta \geq \theta_l\}}\right)$ such that $c(\widetilde{s}_l) < \infty$. Note that $\left\{1_{\{\theta \geq \theta_l\}} : \theta_s \in [\theta_{\min}, \theta_{\max}]\right\}$ is sequentially compact and thus compact in S under the L^1-metric. Since $\left\{B_\varepsilon\left(1_{\{\theta \geq \theta_l\}}\right) : \theta_l \in [\theta_{\min}, \theta_{\max}]\right\}$ is an open cover of $\left\{1_{\{\theta \geq \theta_l\}} : \theta_l \in [\theta_{\min}, \theta_{\max}]\right\}$, it has a finite sub-cover. Let

$$\left\{B_\varepsilon\left(1_{\{\theta \geq \theta^1\}}\right), B_\varepsilon\left(1_{\{\theta \geq \theta^2\}}\right), \cdots, B_\varepsilon\left(1_{\{\theta \geq \theta^N\}}\right)\right\}$$

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

$$\begin{aligned}
& \int_{-\infty}^{\infty} \pi(s(\theta), \theta) \cdot \left[1_{\{\theta \geq \theta_s\}} - \widetilde{s}^n(\theta)\right] g(\theta)\, d\theta \\
\leq \; & \overline{\pi} \cdot \left\|1_{\{\theta \geq \theta_s\}}, \widetilde{s}^n\right\| < z\left(\frac{\rho}{4\overline{g}}\right) \cdot \frac{\rho}{4} \; . \qquad (10)
\end{aligned}$$

Combining (9) and (10) leads to

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

where the left hand side is the sacrificed expected return from using $\widetilde{s}^{n}$ instead of $s_{\lambda}\left(\theta\right)$. The optimality of s_{λ} implies

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

where the second inequality follows the definition of $\overline{c}$. The above two inequalities imply

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

i.e.,

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

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

Proof of Proposition 8.

Proof. We prove the limit uniqueness result from just the absolute continuity of equilibrium SCRs for $\lambda>0$ (together with translation insensitivity). IPD is thus sufficient but not necessary. In other words, it does not matter in the proof if some feasible SCRs are not absolutely continuous, provided that they will not be chosen in equilibrium. Hence, the proof here is more general than just for the stated result of the proposition.

By Lemma 17, we have

$$\lim_{\lambda\to 0}\sup_{\widetilde{s}\in S_{\lambda}(s)\text{ and }s\in S_M}\left\|\widetilde{s},1_{\{\theta\geq\theta_s\}}\right\|=0. \tag{11}$$

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

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

which by Assumption A4 induces a threshold θ_{λ}^{*} such that $\pi\left(\widehat{s}_{\lambda}^{*}\left(\theta\right),\theta\right)>0$ if $\theta>\theta_{\lambda}^{*}$ and $\pi\left(\widehat{s}_{\lambda}^{*}\left(\theta\right),\theta\right)<0$ if $\theta<\theta_{\lambda}^{*}$. By (11),

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

Since

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

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

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

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

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

$$\begin{aligned} & \int_{-\infty}^{\infty} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) \cdot \frac{ds^*_{i,\lambda}(\theta)}{d\theta} \cdot g(\theta)\, d\theta - \lambda \cdot \left.\frac{dc\left(T_\Delta s^*_{i,\lambda}\right)}{d\Delta}\right|_{\Delta=0} \\ = & \int_{-\infty}^{\infty} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) \cdot g(\theta)\, ds^*_{i,\lambda}(\theta) - \lambda \cdot \left.\frac{dc\left(T_\Delta s^*_{i,\lambda}\right)}{d\Delta}\right|_{\Delta=0} = 0. \end{aligned}$$

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

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

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

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

i.e.,

$$\left| \int_{-\infty}^{\infty} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) \cdot g(\theta)\, d\widehat{s}^*_\lambda(\theta) \right| < K\varepsilon\ . \tag{12}$$

Since the density function $g(\theta)$ is continuous on $[\theta_{\min}, \theta_{\max}]$, it is also uniformly continuous on $[\theta_{\min}, \theta_{\max}]$. For the same reason, $\Pi(\theta)$ is also uniformly continuous on $[\theta_{\min}, \theta_{\max}]$. Hence, for any $\varepsilon > 0$, we can find an $\eta > 0$ such that $\left|g(\theta) - g\left(\theta'\right)\right| < \varepsilon$ and $\left|\Pi(\theta) - \Pi\left(\theta'\right)\right| <$

ε for all $\theta, \theta' \in [\theta_{\min}, \theta_{\max}]$ and $\left|\theta - \theta'\right| < 2\eta$. Without loss of generality, we can choose $\eta < \varepsilon$. By (11), 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 (12), we obtain

$$\begin{aligned}
& \left| \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) \cdot g(\theta)\, d\widehat{s}^*_\lambda(\theta) \right| \\
< \;& \int_{-\infty}^{\theta^*_\lambda - \eta} \left|\pi\left(\widehat{s}^*_\lambda(\theta), \theta\right)\right| \cdot g(\theta)\, d\widehat{s}^*_\lambda(\theta) + \int_{\theta^*_\lambda + \eta}^{\infty} \left|\pi\left(\widehat{s}^*_\lambda(\theta), \theta\right)\right| \cdot g(\theta)\, d\widehat{s}^*_\lambda(\theta) + K\varepsilon \\
\leq \;& 2L\overline{g}\varepsilon + K\varepsilon\ , \qquad (13)
\end{aligned}$$

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

$$\left| g(\theta^*_\lambda) \cdot \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) d\widehat{s}^*_\lambda(\theta) - \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) \cdot g(\theta)\, d\widehat{s}^*_\lambda(\theta) \right| < L\varepsilon. \qquad (14)$$

Inequalities (13) and (14) imply

$$\left| \int_{\theta^*_\lambda - \eta}^{\theta^*_\lambda + \eta} \pi\left(\widehat{s}^*_\lambda(\theta), \theta\right) d\widehat{s}^*_\lambda(\theta) \right| < \frac{2L\overline{g} + K + L}{\underline{g}} \varepsilon\ , \qquad (15)$$

where $\underline{g} = \inf_{\theta \in [\theta_{\min}, \theta_{\max}]} g(\theta) > 0$ since g is continuous and strictly positive on $[\theta_{\min}, \theta_{\max}]$ by assumption.

Next note that

$$\begin{aligned}
\left| \int_{\widehat{s}^*_\lambda(\theta^*_\lambda - \eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda + \eta)\, ds - \int_{\widehat{s}^*_\lambda(\theta^*_\lambda - \eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda + \eta)} \pi(s, \theta^*_\lambda - \eta)\, ds \right| \;&\leq\; \left|\Pi(\theta^*_\lambda + \eta) - \Pi(\theta^*_\lambda - \eta)\right| + 4L\varepsilon \\
&<\; \varepsilon + 4L\varepsilon\ , \qquad (16)
\end{aligned}$$

where the first inequality follows because Lemma 17 implies 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 $\Pi(\theta)$ on $[\theta_{\min}, \theta_{\max}]$.

Further note that Assumption A2 implies

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

which together with (15) and (16) implies

$$\begin{aligned} -\left(\frac{2L\overline{g}+K+L}{\underline{g}}+4L+1\right)\varepsilon &< \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi\left(s,\theta^*_\lambda-\eta\right)ds \\ &\leq \int_{\widehat{s}^*_\lambda(\theta^*_\lambda-\eta)}^{\widehat{s}^*_\lambda(\theta^*_\lambda+\eta)} \pi\left(s,\theta^*_\lambda+\eta\right)ds < \left(\frac{2L\overline{g}+K+L}{\underline{g}}+4L+1\right)\varepsilon \end{aligned} \tag{17}$$

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

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

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

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

Therefore, we have

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

which implies

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

according to Assumptions A5 and A6. ■

Proof of Proposition 10.

Proof. Without loss of generality, we only need to consider ε sufficiently small. Let

$$N^\varepsilon_{\theta^*}=\left\{s\in S_M:\Pr\left(\left|s\left(\theta\right)-1_{\left\{\theta\geq\theta^*\right\}}\right|\leq\varepsilon\right)=1\right\}.$$

Suppose the aggregate SCR $\overline{s}\in N^\varepsilon_{\theta^*}$. Since ε is sufficiently small, the cutoff $\theta_{\overline{s}}$ induced by $\overline{s}$ is θ^*, as defined in Assumption A4. That is, $\pi\left(\overline{s}\left(\theta\right),\theta\right)>0$ for $\theta>\theta^*$ and $\pi\left(\overline{s}\left(\theta\right),\theta\right)<0$

for $\theta < \theta^*$. It suffices to show that when λ is sufficiently small, any best response $s^*_{i,\ \lambda}$ to $\overline{s}$ also belongs to $N^{\varepsilon}_{\theta^*}$.

Since $\theta^* \in (\theta_{\min}, \theta_{\max})$ and ε is sufficiently small, we have

$$\inf\left(\left\{\pi\left(1-\varepsilon,\theta\right) : \theta > \theta^*\right\}\right) > 0$$

and

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

Let

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

Since the cost functional satisfies cheap perfect discrimination, we can choose $\rho > 0$ and $K > 0$ such that

$$\left|c\left(L^{\varepsilon}_{\theta^*}s\right) - c\left(s\right)\right| \leq K \cdot \left\|L^{\varepsilon}_{\theta^*}s, s\right\|$$

for all $s \in B_\rho\left(1_{\{\theta \geq \theta^*\}}\right)$. By Lemma 17, there exists a $\lambda_1 > 0$ such that $S_\lambda\left(\overline{s}\right) \in B_\rho\left(1_{\{\theta \geq \theta^*\}}\right)$ for all $\lambda \in (0, \lambda_1)$. Let $\overline{\lambda} = \min\left(\lambda_1, \frac{b}{K}\right)$. We show that any $s \in S_M \backslash N^{\varepsilon}_{\theta^*}$ is strictly dominated by $L^{\varepsilon}_{\theta^*}s$ when $\lambda \in \left(0, \overline{\lambda}\right)$. The benefit from choosing $L^{\varepsilon}_{\theta^*}s$ instead of s is

$$\begin{aligned}
& \int_{-\infty}^{\infty} \pi\left(\overline{s}\left(\theta\right), \theta\right) \cdot \left[\left(L^{\varepsilon}_{\theta^*}s\right)\left(\theta\right) - s\left(\theta\right)\right] g\left(\theta\right) d\theta - \lambda \cdot \left[c\left(L^{\varepsilon}_{\theta^*}s\right) - c\left(s\right)\right] \\
\geq \ & \int_{-\infty}^{\theta^*} - b \cdot \left[\left(L^{\varepsilon}_{\theta^*}s\right)\left(\theta\right) - s\left(\theta\right)\right] g\left(\theta\right) d\theta + \int_{\theta^*}^{\infty} b \cdot \left[\left(L^{\varepsilon}_{\theta^*}s\right)\left(\theta\right) - s\left(\theta\right)\right] g\left(\theta\right) d\theta \\
& -\lambda \cdot \left[c\left(L^{\varepsilon}_{\theta^*}s\right) - c\left(s\right)\right] \\
= \ & b \cdot \left\|L^{\varepsilon}_{\theta^*}s, s\right\| - \lambda \cdot \left[c\left(L^{\varepsilon}_{\theta^*}s\right) - c\left(s\right)\right] \\
> \ & \left(b - \lambda K\right) \cdot \left\|L^{\varepsilon}_{\theta^*}s, s\right\| > 0,
\end{aligned}$$

where the last inequality holds because $\lambda < \frac{b}{K}$ and $\left\|L^{\varepsilon}_{\theta^*}s, s\right\| > 0$ by construction.

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

Proof of Proposition 13.

Proof. Let $\left\{s^*_{i,\lambda}\right\}_{i\in[0,1]}$ denote an equilibrium of the λ-game. By Lemma 12, any equilibrium strategy $s^*_{i,\lambda}$ is absolutely continuous. Since the proof of Proposition 8 just makes use of the absolute continuity of equilibrium SCRs for $\lambda > 0$, the desired result follows the same argument. ■

Proof of Proposition 15.

Proof. Let s_λ be an equilibrium strategy when the cost parameter takes value $\lambda > 0$. We need to show that $\left\| s_\lambda, 1_{\{\theta \geq t\}} \right\| \to 0$ as $\lambda \to 0$. Since the cost of s_λ is determined by its largest slope and $f(\infty) = \infty$, the largest slope of s_λ is finite and s_λ is piecewise linear. If s_λ is nondecreasing, then the desired result is obtained as a special case of Proposition 8, which shows that monotonic equilibria converge to the Laplacian selection as $\lambda \to 0$. Now it is sufficient to focus on the case that s_λ is not monotonic. Let θ^* be the lowest state at which the regime switches, i.e., $\theta^* = \inf\{\theta : s_\lambda(\theta) \geq 1 - \theta\}$.[26] Let k be the maximal slope of s_λ. As the best response to the regime change at θ^*, $s_\lambda(\theta)$ must rise from 0 with slope k at some state $\theta_1 < \theta^*$ and reach the critical mass at θ^*, so that $k(\theta^* - \theta_1) = 1 - \theta^*$, which implies

$$\theta^* = \frac{k\theta_1 + 1}{k+1} \ . \tag{18}$$

Since s_λ is not monotonic, there exists a $\theta_2 > \theta^*$ such that $s_\lambda(\theta) = k(\theta - \theta_1)$ for $\theta \in [\theta_1, \theta_2]$ and s_λ starts to decrease at θ_2. Suppose s_λ decreases with slope $-k'$ in a right neighborhood of θ_2.[27] By the continuity of s_λ, there exists a small $\epsilon > 0$ such that $s_\lambda(\theta_2 + \epsilon) \geq 1 - \theta_2 - \epsilon$. Thus the regime does not change back in $[\theta_2, \theta_2 + \epsilon]$. Now we consider a small deviation from s_λ. Let

$$\widetilde{s}_\lambda = \begin{cases} 0 & \text{if } \theta < \theta_1 + \epsilon \\ k(\theta - \theta_1 - \epsilon) & \text{if } \theta_1 + \epsilon \leq \theta \leq \theta_2 + \frac{k\epsilon}{k+k'} \\ s_\lambda(\theta) & \text{if } \ \theta > \theta_2 + \frac{k\epsilon}{k+k'} \end{cases} .$$

Note that $\widetilde{s}_\lambda(\theta) \leq s_\lambda(\theta)$ holds globally and the inequality is strict in $\left(\theta_1, \theta_2 + \frac{k\epsilon}{k+k'}\right)$. For a player deviating to $\widetilde{s}_\lambda$, the reduction of investment probability in (θ_1, θ^*) where the regime does not change results in an expected gain relative to using s_λ

$$\begin{aligned} & t \cdot \left[\frac{k(\theta^* - \theta_1)^2}{2} - \frac{k(\theta^* - \theta_1 - \epsilon)^2}{2} \right] \cdot g(\theta^*) + O\left(k^{-1}\right) \\ = \quad & k\epsilon t \cdot \left[(\theta^* - \theta_1) - \frac{\epsilon}{2} \right] \cdot g(\theta^*) + O\left(k^{-1}\right) \ , \end{aligned}$$

where the second term $O\left(k^{-1}\right)$ comes from the variation of the density $g(\theta)$ in (θ_1, θ^*). Similarly, the reduction of investment probability in $[\theta^*, \theta_2 + \frac{k\epsilon}{k+k'})$ where the regime changes results in an expected loss relative to using s_λ

$$\begin{aligned} & (1-t) \cdot \left[(\theta_2 - \theta^*) k\epsilon + \frac{k\epsilon}{2} \frac{k\epsilon}{k+k'} \right] \cdot g(\theta^*) + O\left(k^{-1}\right) \\ = \quad & k\epsilon(1-t) \cdot \left[(\theta_2 - \theta^*) + \frac{k\epsilon}{2(k+k')} \right] \cdot g(\theta^*) + O\left(k^{-1}\right) \ , \end{aligned}$$

[26] Note that such θ^* exists since by definition the regime must change in $[0, 1]$.

[27] Actually, since the cost of s_λ depends only on its largest slope, we should have $k' = k$. But this is not important for the proof.

where, again, the second term $O\left(k^{-1}\right)$ comes from the variation of the density $g(\theta)$ in $[\theta^*, \theta_2 + \frac{k\epsilon}{k+k'})$. Hence, the net expected payoff from the deviation is

$$k\epsilon\left[t\left(\theta^*-\theta_1\right)-t\frac{\epsilon}{2}-(1-t)\left(\theta_2-\theta^*\right)-(1-t)\frac{k\epsilon}{2\left(k+k'\right)}\right]\cdot g\left(\theta^*\right)+O\left(k^{-1}\right) ,$$

which should be non-positive due to the optimality of s_λ. Since $\epsilon > 0$ can be chosen arbitrarily small, and in the limit where $\lambda \to 0$, we have $k \to \infty$, this requires

$$t\left(\theta^*-\theta_1\right)-(1-t)\left(\theta_2-\theta^*\right)\leq 0 . \tag{19}$$

Note that

$$\begin{aligned}
& t\left(\theta^*-\theta_1\right)-(1-t)\left(\theta_2-\theta^*\right) \\
= & \left(\theta^*-\theta_1\right)-(1-t)\left(\theta_2-\theta_1\right) \\
= & \frac{1-\theta_1}{k+1}-(1-t)\left(\theta_2-\theta_1\right) \\
\geq & \frac{1-\theta_1}{k+1}-\frac{1-t}{k} ,
\end{aligned} \tag{20}$$

where the second equality follows (18), and the inequality follows the fact that $\theta_2-\theta_1 \leq k^{-1}$. Combining (19) and (20) leads to $\theta_1 \geq t-\frac{1-t}{k}$. Since $\theta^* > \theta_1$, we have

$$\theta^* > t-\frac{1-t}{k} . \tag{21}$$

Let θ^{**} be the highest state at which the regime switches, i.e., $\theta^{**}=\sup\left\{\theta : s_\lambda(\theta)<1-\theta\right\}$. Then similar argument leads to

$$\theta^{**} < t+\frac{t}{k} ,$$

and thus

$$t-\frac{1-t}{k} < \theta^* < \theta^{**} < t+\frac{t}{k} .$$

By construction, θ^* and θ^{**} are the lowest and highest states at which the regime switches. Therefore, as $\lambda \to 0$, k goes to infinity and the regime switches at states arbitrarily close to t. This proves the desired result. ■

Proof of Proposition 16.

Proof. Since the cost only depends on the maximal slope of the SCR, the equilibrium SCR must take the piecewise linear form (4). Let $\widehat{s}_{k,\xi}$ denote the equilibrium SCR and ψ the state at which the regime changes. Then the optimality of $\widehat{s}_{k,\xi}$ implies

$$\xi = \psi + \frac{1}{k}\left(t - \frac{1}{2}\right) . \tag{22}$$

By Lemma 17, the slope k can be arbitrarily large when λ is sufficiently small. Note that as a regime-change threshold, ψ belongs to $(0,1)$. Since k is large, ξ belongs to $(0,1)$ and $\widehat{s}_{k,\xi}(\theta) = 1_{\{\theta \geq \psi\}}$ for $\theta \in \mathbb{R}\backslash(0,1)$. Then, a player's expected payoff from playing $\widehat{s}_{k,\xi}$ is

$$\begin{aligned}
& \int \widehat{s}_{k,\xi}(\theta) \cdot \left(1_{\{\theta \geq \psi\}} - t\right) \cdot g(\theta)\, d\theta - \lambda \cdot f(k) \\
= & \int_{\mathbb{R}\backslash(0,1)} \widehat{s}_{k,\xi}(\theta) \cdot \left(1_{\{\theta \geq \psi\}} - t\right) \cdot g(\theta)\, d\theta + [G(1) - G(0)] \int_0^1 \widehat{s}_{k,\xi}(\theta) \cdot \left(1_{\{\theta \geq \psi\}} - t\right) d\theta - \lambda \cdot f(k) \\
= & \int_1^{\infty} \left(1_{\{\theta \geq \psi\}} - t\right) \cdot g(\theta)\, d\theta + [G(1) - G(0)] \left[(1-\psi)(1-t) - \frac{1}{2} t (1-t) k^{-1}\right] - \lambda \cdot f(k),
\end{aligned}$$

where the first equality follows that G is uniform over $[0,1]$, and the second equality follows that $\widehat{s}_{k,\xi}(\theta) = 1_{\{\theta \geq \psi\}}$ for $\theta \in \mathbb{R}\backslash(0,1)$. Hence, it is straightforward to see that k solves

$$\max_{k>0} -\frac{1}{2}[G(1) - G(0)]\, t(1-t) \cdot k^{-1} - \lambda \cdot f(k) .$$

Next, by definition, the equilibrium regime-change threshold ψ is pinned down by

$$\begin{aligned}
1 - \psi &= \widehat{s}_{k,\xi}(\psi) \\
&= \frac{1}{2} + k(\psi - \xi) \\
&= 1 - t ,
\end{aligned}$$

where the last equality follows (22). Hence, the equilibrium regime-change threshold is $\psi = t$.

Finally, (22) immediately implies

$$\xi = t + \left(t - \frac{1}{2}\right) k^{-1} .$$

■

7 Online Appendix: Properties of Cost Functionals

In this section, we collect together proofs of properties of cost functionals mentioned in main body of the paper.

7.1 Entropy Reduction Cost Functional

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

Proof. For any SCR s, the associated entropy reduction is

$$c(s) = \mathbf{E}\left[H(s(\theta))\right] - H\left[\mathbf{E}(s(\theta))\right] \ ,$$

where $H : [0,1] \rightarrow \mathbb{R}$ is given by

$$H(x) = x \ln x + (1-x)\ln(1-x) \ .$$

Now let $p_1(s) = \mathbf{E}(s(\theta))$ denote the unconditional probability that action 1 is chosen under SCR s. Note that this cost functional is convex and Fréchet differentiable at s with derivative

$$H'(s(\theta)) - H'(p_1(s)) \ .$$

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

$$c\left(L_\psi^\varepsilon s\right) - c(s) \leq \int_{A(s)} \left[H'\left(L_\psi^\varepsilon s(\theta)\right) - H'\left(p_1\left(L_\psi^\varepsilon s\right)\right)\right]\left(L_\psi^\varepsilon s(\theta) - s(\theta)\right) dG(\theta)$$

and

$$c\left(L_\psi^\varepsilon s\right) - c(s) \geq \int_{A(s)} \left[H'(s(\theta)) - H'(p_1(s))\right]\left(L_\psi^\varepsilon s(\theta) - s(\theta)\right) dG(\theta) \ ,$$

Hence,

$$\left|c\left(L_\psi^\varepsilon s\right) - c(s)\right| \leq \max\left(\begin{array}{c} \left|\int_{A(s)} \left[H'(s(\theta)) - H'(p_1(s))\right]\left(L_\psi^\varepsilon s(\theta) - s(\theta)\right) dG(\theta)\right|, \\ \left|\int_{A(s)} \left[H'\left(L_\psi^\varepsilon s(\theta)\right) - H'\left(p_1\left(L_\psi^\varepsilon s\right)\right)\right]\left(L_\psi^\varepsilon s(\theta) - s(\theta)\right) dG(\theta)\right| \end{array}\right) \ .$$

Since $H'(x)$ is increasing in x, for all $\theta \in A(s)$, both $|H'(s(\theta)) - H'(p_1(s))|$ and $\left|H'\left(L^{\varepsilon}_{\psi} s(\theta)\right) - H'\left(p_1\left(L^{\varepsilon}_{\psi} s\right)\right)\right|$ are bounded above by

$$K = \max\left(\left|H'(1-\varepsilon) - H'(\xi)\right|, \left|H'(1-\xi) - H'(\varepsilon)\right|\right) .$$

Therefore,

$$\begin{aligned} \left|c\left(L^{\varepsilon}_{\psi} s\right) - c(s)\right| &\leq \int_{A(s)} K \cdot \left|L^{\varepsilon}_{\psi} s(\theta) - s(\theta)\right| dG(\theta) \\ &= K \cdot \left\|L^{\varepsilon}_{\psi} s, s\right\| . \end{aligned}$$

This concludes the proof. ■

7.2 The Pairwise-Separable Cost Functional

Lemma 19 *The PS cost functional satisfies A9 (feasible almost perfect discrimination).*

Proof. It suffices to show that $c_{PS}(\widehat{s}_{k,\psi}) < \infty$, i.e., the integral

$$\int_{\theta} \int_{\theta'} \left|\theta' - \theta\right|^{-\alpha} D\left(\widehat{s}_{k,\psi}(\theta), \widehat{s}_{k,\psi}(\theta')\right) h\left(\theta, \theta'\right) d\theta' d\theta$$

exists.

Let

$$A = \left\{\left(\theta, \theta'\right) \in \mathbb{R}^2 : -k^{-1} \leq \theta - \theta' \leq k^{-1}\right\}$$

and

$$A_1 = \left\{\left(\theta, \theta'\right) \in \mathbb{R}^2 : \theta \geq \psi + k^{-1}/2 \text{ and } \theta' \geq \psi + k^{-1}/2, \text{ or } \theta \leq \psi - k^{-1}/2 \text{ and } \theta' \leq \psi - k^{-1}/2\right\} .$$

First note that $\left|\theta' - \theta\right|^{-\alpha}$ is bounded on $\mathbb{R}^2 \backslash A$, thus the integral over $\mathbb{R}^2 \backslash A$ exists. Second, since $D\left(\widehat{s}_{k,\psi}(\theta), \widehat{s}_{k,\psi}(\theta')\right) = 0$ on A_1, we just need to show that the integral over $A \backslash A_1$ exists. Let

$$B_1 = \left\{\left(\theta, \theta'\right) \in A \backslash A_1 : -k^{-1}/2 \leq \theta' \leq k^{-1}/2 \text{ and } 0 \leq \theta - \theta' \leq k^{-1}\right\} ,$$

$$B_2 = \left\{\left(\theta, \theta'\right) \in A \backslash A_1 : -k^{-1}/2 \leq \theta' \leq k^{-1}/2 \text{ and } 0 \leq \theta' - \theta \leq k^{-1}\right\} ,$$

$$B_3 = \left\{\left(\theta, \theta'\right) \in A \backslash A_1 : -k^{-1}/2 \leq \theta \leq k^{-1}/2 \text{ and } 0 \leq \theta' - \theta \leq k^{-1}\right\} ,$$

and

$$B_4 = \left\{\left(\theta, \theta'\right) \in A \backslash A_1 : -k^{-1}/2 \leq \theta \leq k^{-1}/2 \text{ and } 0 \leq \theta - \theta' \leq k^{-1}\right\} .$$

Then $A \backslash A_1 = B_1 \cup B_2 \cup B_3 \cup B_4$. We next show that the integral over B_1 exists. Similar calculations can show the existence of the integral over B_2, B_3 and B_4, and are thus omitted.

By definition of a PS cost functional, $D(x_1, x_2)$ is bounded on $[0,1] \times [0,1]$ and $D(x_1, x_2) = O\left(|x_1 - x_2|^\beta\right)$ as $|x_1 - x_2| \to 0$. So there exists a $K > 0$, such that

$$D(x_1, x_2) \le K \cdot |x_1 - x_2|^\beta \tag{23}$$

on $[0,1] \times [0,1]$. Now

$$\begin{aligned}
& \int_{B_1} \left|\theta - \theta'\right|^{-\alpha} D\left(\widehat{s}_{k,\psi}(\theta), \widehat{s}_{k,\psi}(\theta')\right) h(\theta, \theta')\, d\theta' d\theta \\
\le\ & \int_{B_1} \left|\theta - \theta'\right|^{-\alpha} K \cdot \left|\widehat{s}_{k,\psi}(\theta) - \widehat{s}_{k,\psi}(\theta')\right|^\beta h(\theta, \theta')\, d\theta' d\theta \\
=\ & \int_{B_1} \left(\theta - \theta'\right)^{-\alpha} K \cdot \left(\frac{1}{2} + k(\theta - \psi) - \frac{1}{2} - k(\theta' - \psi)\right)^\beta h(\theta, \theta')\, d\theta' d\theta \\
\le\ & K k^\beta \overline{h} \int_{B_1} \left(\theta - \theta'\right)^{\beta - \alpha} d\theta' d\theta\ ,
\end{aligned}$$

for some $\overline{h} > 0$, where the first inequality is implied by (23), the equality is implied by the definition of $\widehat{s}_{k,\psi}$ and the last inequality is true because $\theta \ge \theta'$ on B_1 and $\frac{h(\theta,\theta')}{g(\theta)g(\theta')}$ is bounded above in the definition of PS cost functionals.

Now changing variables from (θ, θ') to (t, t') such that $t = \theta$ and $t' = \theta - \theta'$, we have

$$\begin{aligned}
& \int_{B_1} \left(\theta - \theta'\right)^{\beta - \alpha} d\theta' d\theta \\
=\ & \int_0^{k^{-1}} (t')^{\beta - \alpha} \int_{-k^{-1}/2 + t'}^{k^{-1}/2 + t'} dt \cdot dt' \\
=\ & k^{-1} \int_0^{k^{-1}} (t')^{\beta - \alpha} dt'\ .
\end{aligned}$$

This integral exists since $\beta - \alpha + 1 > 0$. Therefore, $c_{PS}(\widehat{s}_{k,\psi}) < \infty$. ■

Proposition 20 *The PS cost functional satisfies IPD if and only if $\alpha \ge 2$.*

Proof. Let s be a non-decreasing discontinuous SCR and $s\left(\widehat{\theta}_-\right) < s\left(\widehat{\theta}_+\right)$ for some $\widehat{\theta} \in [\theta_{\min}, \theta_{\max}]$.[28] Let

$$s_{\widehat{\theta}}(\theta) = \begin{cases} s\left(\widehat{\theta}_+\right) & \text{if } \theta > \widehat{\theta} \\ s\left(\widehat{\theta}_-\right) & \text{if } \theta \le \widehat{\theta} \end{cases} \tag{24}$$

[28] We can focus on $\widehat{\theta} \in [\theta_{\min}, \theta_{\max}]$ because the possible $\widehat{\theta}$s of equilibrium SCRs are always in $[\theta_{\min}, \theta_{\max}]$ due to Assumption A3.

and

$$A = \min\left[D\left(s\left(\widehat{\theta}_-\right), s\left(\widehat{\theta}_+\right)\right), D\left(s\left(\widehat{\theta}_+\right), s\left(\widehat{\theta}_-\right)\right)\right].$$

Note that $A > 0$ since $s\left(\widehat{\theta}_-\right) \neq s\left(\widehat{\theta}_+\right)$. Then we have

$$\begin{aligned}
c_{PS}(s) &= \int_\theta \int_{\theta'} \left|\theta' - \theta\right|^{-\alpha} D\left(s(\theta), s(\theta')\right) h\left(\theta, \theta'\right) d\theta' d\theta \\
&\geq \int_\theta \int_{\theta'} \left|\theta' - \theta\right|^{-\alpha} D\left(s_{\widehat{\theta}}(\theta), s_{\widehat{\theta}}(\theta')\right) h\left(\theta, \theta'\right) d\theta' d\theta \\
&= D\left(s\left(\widehat{\theta}_-\right), s\left(\widehat{\theta}_+\right)\right) \int_{-\infty}^{\widehat{\theta}} \int_{\widehat{\theta}}^{\infty} \left(\theta' - \theta\right)^{-\alpha} h\left(\theta, \theta'\right) d\theta' d\theta \\
&\quad + D\left(s\left(\widehat{\theta}_+\right), s\left(\widehat{\theta}_-\right)\right) \int_{\widehat{\theta}}^{\infty} \int_{-\infty}^{\widehat{\theta}} \left(\theta - \theta'\right)^{-\alpha} h\left(\theta, \theta'\right) d\theta' d\theta \\
&\geq 2A \cdot \int_{-\infty}^{\widehat{\theta}} \int_{\widehat{\theta}}^{\infty} \left(\theta' - \theta\right)^{-\alpha} h\left(\theta, \theta'\right) d\theta' d\theta \, , \qquad (25)
\end{aligned}$$

where the first inequality follows the monotonicity of s in θ, and the second inequality follows the definition of A. Since g is continuous and strictly positive on $[\theta_{\min}, \theta_{\max}]$, it has a strictly positive lower bound on $[\theta_{\min}, \theta_{\max}]$. Since $\frac{g(\theta)g(\theta')}{h(\theta,\theta')}$ is bounded above, $h\left(\theta, \theta'\right)$ has a strictly positive lower bound on $[\theta_{\min}, \theta_{\max}] \times [\theta_{\min}, \theta_{\max}]$. Hence, $\int_{-\infty}^{\widehat{\theta}} \int_{\widehat{\theta}}^{\infty} \left(\theta' - \theta\right)^{-\alpha} h\left(\theta, \theta'\right) d\theta' d\theta$ is integrable if and only if $2 - \alpha > 0$. Therefore, $\alpha \geq 2$ implies $c_{PS}(s) = \infty$ and thus IPD. For the converse, consider an SCR $s_{\widehat{\theta}}(\cdot)$ defined by (24) such that $D\left(s\left(\widehat{\theta}_-\right), s\left(\widehat{\theta}_+\right)\right) = D\left(s\left(\widehat{\theta}_+\right), s\left(\widehat{\theta}_-\right)\right) \equiv A > 0$. Immediate from the previous derivation of (25) we obtain that $c_{PS}\left(s_{\widehat{\theta}}\right) = \infty$ if $\alpha \geq 2$ and $c\left(s_{\widehat{\theta}}\right) < \infty$ if $\alpha < 2$. Then, IPD implies $c_{PS}\left(s_{\widehat{\theta}}\right) = \infty$ and thus $\alpha \geq 2$. ■

The following lemmas show that CPD is satisfied if $\alpha = 0$ and it is easier to be satisfied at lower values of α. Since the PS cost functional is continuous in α, there exists some $\widehat{\alpha} \in [0, \min(2, \beta + 1))$ such that CPD is satisfied for $\alpha \in [0, \widehat{\alpha}]$. Due to the technicalities associated with the PS cost functional and the generality of the definitions of CPD and EPD, however, we do not obtain an analytical bound $\widehat{\alpha}$ between CPD and EPD.

Lemma 21 *The PS cost functional satisfies CPD at $\alpha = 0$.*

Proof. When $\alpha = 0$, the cost functional becomes

$$c_{PS}(s) = \int_\theta \int_{\theta'} D\left(s(\theta), s(\theta')\right) h\left(\theta, \theta'\right) d\theta' d\theta \, .$$

Hence, by the triangle inequality,

$$\begin{aligned}
\left|c_{PS}\left(L_{\psi}^{\varepsilon}s\right)-c_{PS}\left(s\right)\right| &= \left|\int_{\theta}\int_{\theta'}\left[D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),\left(L_{\psi}^{\varepsilon}s\right)(\theta')\right)-D\left(s(\theta),s(\theta')\right)\right]h\left(\theta,\theta'\right)d\theta'd\theta\right| \\
&\leq \int_{\theta}\int_{\theta'}\left|D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),\left(L_{\psi}^{\varepsilon}s\right)(\theta')\right)-D\left(s(\theta),s(\theta')\right)\right|h\left(\theta,\theta'\right)d\theta'd\theta \\
&\leq \int_{\theta}\int_{\theta'}\left|D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),\left(L_{\psi}^{\varepsilon}s\right)(\theta')\right)-D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),s(\theta')\right)\right|h\left(\theta,\theta'\right)d\theta'd\theta \\
&\quad+\int_{\theta}\int_{\theta'}\left|D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),s(\theta')\right)-D\left(s(\theta),s(\theta')\right)\right|h\left(\theta,\theta'\right)d\theta'd\theta\ . \qquad (26)
\end{aligned}$$

Since $\frac{\partial D(x_1,x_2)}{\partial x_1}$ and $\frac{\partial D(x_1,x_2)}{\partial x_2}$ exist on $[0,1]\times[0,1]$,[29] there exists a $K>0$ such that $\left|D\left(x_1',x_2\right)-D\left(x_1,x_2\right)\right|\leq K\cdot\left|x_1'-x_1\right|$ and $\left|D\left(x_1,x_2'\right)-D\left(x_1,x_2\right)\right|\leq K\cdot\left|x_2'-x_2\right|$ for all $x_1,x_2\in[0,1]$. Hence,

$$\left|D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),\left(L_{\psi}^{\varepsilon}s\right)(\theta')\right)-D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),s(\theta')\right)\right|\leq K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta')-s(\theta')\right|$$

and

$$\left|D\left(\left(L_{\psi}^{\varepsilon}s\right)(\theta),s(\theta')\right)-D\left(s(\theta),s(\theta')\right)\right|\leq K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta)-s(\theta)\right|\ .$$

Plugging the above two inequalities into (26), we obtain

$$\begin{aligned}
&\left|c_{PS}\left(L_{\psi}^{\varepsilon}s\right)-c_{PS}\left(s\right)\right| \\
\leq\ & \int_{\theta}\int_{\theta'}K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta')-s(\theta')\right|h\left(\theta,\theta'\right)d\theta'd\theta+\int_{\theta}\int_{\theta'}K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta)-s(\theta)\right|h\left(\theta,\theta'\right)d\theta'd\theta \\
\leq\ & \int_{\theta}\int_{\theta'}K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta')-s(\theta')\right|K'g(\theta')g(\theta)d\theta'd\theta+\int_{\theta}\int_{\theta'}K\cdot\left|\left(L_{\psi}^{\varepsilon}s\right)(\theta)-s(\theta)\right|K'g(\theta')g(\theta)d\theta'd\theta \\
=\ & KK'\cdot\int_{\theta}\left\|L_{\psi}^{\varepsilon}s,s\right\|g(\theta)d\theta+KK'\cdot\int_{\theta'}\left\|L_{\psi}^{\varepsilon}s,s\right\|g(\theta')d\theta' \\
=\ & 2KK'\cdot\left\|L_{\psi}^{\varepsilon}s,s\right\|\ ,
\end{aligned}$$

where the second inequality follows because $\frac{h(\theta,\theta')}{g(\theta)g(\theta')}$ is bounded above by some $K'>0$. Therefore, c_{PS} satisfies CPD when $\alpha=0$. ■

Lemma 22 *If the PS cost functional satisfies CPD at some $\alpha\geq 0$, then it satisfies CPD at all $\alpha'\in[0,\alpha]$.*

Proof. To avoid confusion, let $c_{PS}^{\alpha}(\cdot)$ denote the PS cost functional with parameter α. Since $c_{PS}^{\alpha}(\cdot)$ satisfies CPD, for any $\psi\in\mathbb{R}$ and $\varepsilon\in(0,1/2)$, there exists a $\rho>0$ and $K>0$ such that

$$\left|c_{PS}^{\alpha}\left(L_{\psi}^{\varepsilon}s\right)-c_{PS}^{\alpha}\left(s\right)\right|\leq K\cdot\left\|L_{\psi}^{\varepsilon}s,s\right\|$$

[29] The proof goes through under a weaker condition that $\frac{\partial}{\partial x_i}D(x_1,x_2)$ exists for all $x_i\in(0,1)$ and $x_j\in[0,1]$, $i,j\in\{1,2\}$, $i\neq j$.

for all monotonic $s \in B_\rho\left(1_{\{\theta\geq\psi\}}\right)$. Without loss of generality, we can choose a sufficiently small $\rho > 0$. Then by the construction of operator L^ε_ψ, there exists an interval $[\theta_1, \theta_2]$ such that for any monotonic $s \in B_\rho\left(1_{\{\theta\geq\psi\}}\right)$, $L^\varepsilon_\psi s$ and s differ only in $[\theta_1, \theta_2]$. Fix a $z > 0$. Then

$$
\begin{aligned}
& \left|c^{\alpha'}_{PS}\left(L^\varepsilon_\psi s\right) - c^{\alpha'}_{PS}\left(s\right)\right| \\
= & \left|\int_\theta\int_{\theta'} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
\leq & \left|\int_{\mathbb{R}^2\setminus[\theta_1-z,\theta_2+z]\times[\theta_1-z,\theta_2+z]} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
& + \left|\int_{[\theta_1-z,\theta_2+z]\times[\theta_1-z,\theta_2+z]} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
= & \left|\int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)}\int_{[\theta_1,\theta_2]} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
& + \left|\int_{[\theta_1,\theta_2]}\int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
& + \left|\int_{[\theta_1-z,\theta_2+z]\times[\theta_1-z,\theta_2+z]} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \quad (27)
\end{aligned}
$$

where the second equality follows the fact $L^\varepsilon_\psi s$ and s differ only in $[\theta_1, \theta_2]$. Since $\frac{\partial D(x_1,x_2)}{\partial x_1}$ and $\frac{\partial D(x_1,x_2)}{\partial x_2}$ exist on $[0,1]\times[0,1]$,[30] there exists a $K_1 > 0$ such that $|D(x_1', x_2) - D(x_1, x_2)| \leq K_1 \cdot |x_1' - x_1|$ and $|D(x_1, x_2') - D(x_1, x_2)| \leq K_1 \cdot |x_2' - x_2|$ for all $x_1, x_2 \in [0,1]$. Then, the first term in the right hand side of (27) is

$$
\begin{aligned}
& \left|\int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)}\int_{[\theta_1,\theta_2]} \left|\theta'-\theta\right|^{-\alpha'}\left[D\left(\left(L^\varepsilon_\psi s\right)(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right] h\left(\theta,\theta'\right) d\theta' d\theta\right| \\
\leq & \int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)}\int_{[\theta_1,\theta_2]} \left|\theta'-\theta\right|^{-\alpha'}\left|D\left(s(\theta), \left(L^\varepsilon_\psi s\right)(\theta')\right) - D\left(s(\theta), s(\theta')\right)\right| h\left(\theta,\theta'\right) d\theta' d\theta \\
\leq & K' \int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)}\int_{[\theta_1,\theta_2]} z^{-\alpha'} K_1 \cdot \left|\left(L^\varepsilon_\psi s\right)(\theta') - s(\theta')\right| g(\theta')\, d\theta' g(\theta)\, d\theta \\
\leq & z^{-\alpha'} K' K_1 \cdot \int_{(-\infty,\theta_1-z)\cup(\theta_2+z,\infty)} \left\|L^\varepsilon_\psi s, s\right\| g(\theta)\, d\theta \\
\leq & z^{-\alpha'} K' K_1 \cdot \left\|L^\varepsilon_\psi s, s\right\| \ ,
\end{aligned}
$$

where the first inequality holds because $\left(L^\varepsilon_\psi s\right)(\theta) = s(\theta)$ for $\theta \in (-\infty, \theta_1 - z) \cup (\theta_2 + z, \infty)$, and the second inequality follows that $\left|\theta'-\theta\right|^{-\alpha'} \leq z^{-\alpha'}$ for $\theta \in (-\infty, \theta_1 - z) \cup (\theta_2 + z, \infty)$

[30] The proof goes through under a weaker condition that $\frac{\partial}{\partial x_i} D(x_1, x_2)$ exists for all $x_i \in (0,1)$ and $x_j \in [0,1]$, $i, j \in \{1,2\}$, $i \neq j$.

and $\theta' \in [\theta_1, \theta_2]$, and that $\frac{h(\theta,\theta')}{g(\theta)g(\theta')}$ is bounded above by some $K' > 0$. By a symmetric argument, the second term in the right hand side of (27) is also bounded by $z^{-\alpha'} K' K_1 \cdot \left\| L_{\psi}^{\varepsilon} s, s \right\|$. Since $\alpha - \alpha' \geq 0$, $\left|\theta' - \theta\right|^{\alpha - \alpha'}$ is bounded for $\left(\theta, \theta'\right) \in [\theta_1 - z, \theta_2 + z] \times [\theta_1 - z, \theta_2 + z]$, then there is a $K_2 > 0$ such that the third term in the right hand side of (27) is

$$
\begin{aligned}
& \left| \int_{[\theta_1 - z, \theta_2 + z] \times [\theta_1 - z, \theta_2 + z]} \left|\theta' - \theta\right|^{\alpha - \alpha'} \left|\theta' - \theta\right|^{-\alpha} \left[D\left(\left(L_{\psi}^{\varepsilon} s\right)(\theta), \left(L_{\psi}^{\varepsilon} s\right)\left(\theta'\right)\right) - D\left(s(\theta), s\left(\theta'\right)\right)\right] h\left(\theta, \theta'\right) d\theta' d\theta \right| \\
\leq \; & K' K_2 \cdot \left| \int_{[\theta_1 - z, \theta_2 + z] \times [\theta_1 - z, \theta_2 + z]} \left|\theta' - \theta\right|^{-\alpha} \left[D\left(\left(L_{\psi}^{\varepsilon} s\right)(\theta), \left(L_{\psi}^{\varepsilon} s\right)\left(\theta'\right)\right) - D\left(s(\theta), s\left(\theta'\right)\right)\right] g\left(\theta'\right) g(\theta)\, d\theta' d\theta \right| \\
\leq \; & K' K_2 \cdot \left| c_{PS}^{\alpha}\left(L_{\psi}^{\varepsilon} s\right) - c_{PS}^{\alpha}(s)\right| \\
\leq \; & K' K_2 K \cdot \left\| L_{\psi}^{\varepsilon} s, s \right\| \; .
\end{aligned}
$$

Hence, (27) becomes

$$
\begin{aligned}
& \left| c_{PS}^{\alpha'}\left(L_{\psi}^{\varepsilon} s\right) - c_{PS}^{\alpha'}(s) \right| \\
\leq \; & 2 z^{-\alpha'} K' K_1 \cdot \left\| L_{\psi}^{\varepsilon} s, s \right\| + K' K_2 K \cdot \left\| L_{\psi}^{\varepsilon} s, s \right\| \\
= \; & \left(2 z^{-\alpha'} K_1 + K_2 K \right) K' \cdot \left\| L_{\psi}^{\varepsilon} s, s \right\| \; .
\end{aligned}
$$

Therefore, $c_{PS}^{\alpha'}$ satisfies CPD. ■

7.3 The Fisher Cost Functional

Lemma 23 *The Fisher cost functional satisfies sub-modularity.*

Proof. Let s_1 and s_2 be two SCRs. It is straightforward to see that $c_{Fisher}(s_2 \vee s_1) + c_{Fisher}(s_2 \wedge s_1) = c_{Fisher}(s_1) + c_{Fisher}(s_2)$. Let $A = \{\theta \in \mathbb{R} : s_2(\theta) \geq s_1(\theta)\}$ and $B = \{\theta \in \mathbb{R} : s_2(\theta) < s_1(\theta)\}$. Then,

$$
\begin{aligned}
& c_{Fisher}(s_2 \vee s_1) + c_{Fisher}(s_2 \wedge s_1) \\
= & \int_A \frac{\left([g(\theta)\, s_2(\theta)]'\right)^2}{g(\theta)\, s_2(\theta)} + \frac{\left([g(\theta)\,(1 - s_2(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_2(\theta))} d\theta + \int_B \frac{\left([g(\theta)\, s_1(\theta)]'\right)^2}{g(\theta)\, s_1(\theta)} + \frac{\left([g(\theta)\,(1 - s_1(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_1(\theta))} d\theta \\
& + \int_A \frac{\left([g(\theta)\, s_1(\theta)]'\right)^2}{g(\theta)\, s_1(\theta)} + \frac{\left([g(\theta)\,(1 - s_1(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_1(\theta))} d\theta + \int_B \frac{\left([g(\theta)\, s_2(\theta)]'\right)^2}{g(\theta)\, s_2(\theta)} + \frac{\left([g(\theta)\,(1 - s_2(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_2(\theta))} d\theta \\
= & \int_A \frac{\left([g(\theta)\, s_1(\theta)]'\right)^2}{g(\theta)\, s_1(\theta)} + \frac{\left([g(\theta)\,(1 - s_1(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_1(\theta))} d\theta + \int_B \frac{\left([g(\theta)\, s_1(\theta)]'\right)^2}{g(\theta)\, s_1(\theta)} + \frac{\left([g(\theta)\,(1 - s_1(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_1(\theta))} \\
& + \int_A \frac{\left([g(\theta)\, s_2(\theta)]'\right)^2}{g(\theta)\, s_2(\theta)} + \frac{\left([g(\theta)\,(1 - s_2(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_2(\theta))} d\theta + \int_B \frac{\left([g(\theta)\, s_2(\theta)]'\right)^2}{g(\theta)\, s_2(\theta)} + \frac{\left([g(\theta)\,(1 - s_2(\theta))]'\right)^2}{g(\theta)\, 1\,(-s_2(\theta))} d\theta d\theta \\
= & c_{Fisher}(s_1) + c_{Fisher}(s_2) \ .
\end{aligned}
$$

■

7.4 The Additive Noise Cost Functional

Here we show that the additive noise cost functional c_{AN} is not submodular, by constructing a counterexample. Suppose ε is uniform on $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Let $b_\psi = 1_{\{x \geq \psi\}}$ be the step function behavioral strategy where a player invests if and only if his signal is above ψ. Then the induced stochastic choice rule $\widetilde{s}_{k,b_\psi}$ is equal to the slope k threshold approximation of $1_{\{\theta \geq \psi\}}$, i.e.,

$$
\widetilde{s}_{k,b_\psi}(\theta) = \int_{-1/2}^{1/2} b_\psi\left(\theta + \frac{1}{k}\varepsilon\right) d\varepsilon = \int_{-1/2}^{1/2} 1_{\varepsilon \leq k(\theta - \psi)} = \widehat{s}_{k,\psi}(\theta)
$$

Since k is the maximum slope of $\widehat{s}_{k,\psi}$, we have

$$
\frac{d\widetilde{s}_{k,b}(\theta)}{d\theta} \leq k \ , \tag{28}
$$

where the inequality is an equality if and only if the behavioral strategy is the switching strategy b_ψ for some switching cutoff ψ. Now consider $\widetilde{s}_{k_1,b_\psi}$ and $\widetilde{s}_{k_2,b_\psi}$, where $k_2 > k_1 > 0$. Note that $\widetilde{s}_{k_1,b_\psi}$ and $\widetilde{s}_{k_2,b_\psi}$ intersect at $(\psi, 1/2)$, so that

$$
\left(\widetilde{s}_{k_1,b_\psi} \vee \widetilde{s}_{k_2,b_\psi}\right)(\theta) = \begin{cases} \widetilde{s}_{k_1,b_\psi}(\theta) & \text{if } \theta < \psi \\ \widetilde{s}_{k_2,b_\psi}(\theta) & \text{if } \theta \geq \psi \end{cases}
$$

and

$$\left(\widetilde{s}_{k_1,b_\psi} \wedge \widetilde{s}_{k_2,b_\psi}\right)(\theta) = \begin{cases} \widetilde{s}_{k_2,b_\psi}(\theta) & \text{if } \theta < \psi \\ \widetilde{s}_{k_1,b_\psi}(\theta) & \text{if } \theta \geq \psi \end{cases} .$$

So k_2 is the maximal slope of both $\widetilde{s}_{k_1,b_\psi} \vee \widetilde{s}_{k_2,b_\psi}$ and $\widetilde{s}_{k_1,b_\psi} \wedge \widetilde{s}_{k_2,b_\psi}$. Inequality (28) thus implies $c_{AN}\left(\widetilde{s}_{k_1,b_\psi} \vee \widetilde{s}_{k_2,b_\psi}\right) = c(k_2)$ and $c_{AN}\left(\widetilde{s}_{k_1,b_\psi} \wedge \widetilde{s}_{k_2,b_\psi}\right) = c(k_2)$. Therefore,

$$\begin{aligned} c_{AN}\left(\widetilde{s}_{k_1,b_\psi}\right) + c_{AN}\left(\widetilde{s}_{k_2,b_\psi}\right) &= \widehat{c}(k_1) + \widehat{c}(k_2) \\ &< 2\widehat{c}(k_2) \\ &= c_{AN}\left(\widetilde{s}_{k_1,b_\psi} \vee \widetilde{s}_{k_2,b_\psi}\right) + c_{AN}\left(\widetilde{s}_{k_1,b_\psi} \wedge \widetilde{s}_{k_2,b_\psi}\right) , \end{aligned}$$

a violation of submodularity.