ESSAYS IN INFORMATIONAL ECONOMICS

By

Jianrong Tian

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Economics)

at the
UNIVERSITY OF WISCONSIN MADISON

2016

Date of final oral examination: 05/04/2016

The dissertation is approved by the following members of the Final Oral Committee:
Lones Smith, Professor, Economics
William Sandholm, Professor, Economics
Dean Corbae, Professor, Finance
Marek Weretka, Associate Professor, Economics
Antonio Penta, Associate Professor, Economics

Acknowledgments

I am very grateful to Antonio Penta and Marek Weretka for their encouragement, scholarly advice, and invaluable discussions. Without their help, I could not have written the first chapter of my thesis. I would also like to thank Dean Corbae, Ju Hu, Daniel Quint, Marzena Rostek, William Sandholm, Ricardo Serrano-Padial, Lones Smith, Peter Sorensen, Nicholas Tenev, and Stephen Wolff for helpful discussions. Finally, thank Lones Smith for his financial support.

Abstract

In Chapter 1, I develop a theory of optimal interval division for capacity-constrained problems in which a continuum is divided into finitely many intervals. Examples include the location of public facilities by a social planner, the distribution of product characteristics by firms, coarse matching, and bounded memory. Optimal interval division refers to the problem of finding the interval partition of a given continuum maximizing an associated value, constrained by the number of classes. The value of each finite interval partition is derived by summation from a basic primitive that I call a cell function defined over all subintervals. I identify an important and common property of cell functions: they are submodular over the interval structure. This yields the following results. First, the maximum value exhibits decreasing marginal returns in the number of classes, and converges rapidly with an additional condition. Second, I uncover a novel sandwiching property: when allowing an extra class, the new optimal cut- offs are more spread in the sense that each new cut point lands in a different original partition element. Third, I show that a submodular optimal interval division problem can often be interpreted as scalar quantization with a general error function, despite arising from seemingly unrelated contexts like coarse matching.

In Chapter 2, I develop a multi-dimensional sufficient condition for the interval dominance order in Quah and Strulovici (2009), as well as a simple and unified proof by the method of convex cones. My new ingredient is the notion of a diagonally increasing (DI) path. DI paths admit an intuitive interpretation as the freedom of downwards switching within some contexts. Applications of this extension include optimal switching time and monotone pragmatics.

In Chapter 3 (with Lones Smith and Peter Srensen), we give a welfare analysis of the traditional herding model in which individuals ignore future informational gains to others in their discrete action choices. We explore what happens when individuals optimally internalize this gain. The resulting social planner's problem is a Bayesian optimal experimentation exercise. Herding is socially efficient, but occurs less readily. A simple reward scheme for selfish individuals can implement socially efficient behavior. Under a robust new information log-concavity condition, efficiency entails contrarian behavior individuals should optimally lean against taking the myopically more popular actions. Our sufficient condition also precludes informational cascades.

Contents

1 Optimal Interval Division

1.1 Introduction

Numerous economic problems involve a continuum divided into finitely many classes and choices or services that are class-targeted, due to various capacity constraints. For example, consider Hotelling City, in which residents are distributed along a (linear) interval. The mayor first distributes finitely many merchants along this interval, then the residents buy a commodity from the closest merchant. As a consequence, the residents are finitely partitioned into intervals so that everyone in a given subinterval buys from the same merchant. The Hotelling story corresponds to real problems like location of public facilities by a social planner or the distribution of finitely many product characteristics by firms. I discuss other examples later, including coarse matching [Wilson, 1989, McAfee, 2002], bounded memory [Dow, 1991], and mechanism design with limited communication [Rogerson, 2003, Chu and Sappington, 2007].[1]

In this paper, I develop a theory of optimal interval division for capacity-constrained problems. In my framework, the value of each finite interval partition of a given continuum is derived by summation from a basic primitive that I call a *cell function* defined over all the subintervals. Optimal interval division refers to the problem of finding the interval partition maximizing the associated value, constrained by the number of classes (i.e. partition elements).

For instance, in the Hotelling story, suppose that the mayor distributes the merchants to minimize the aggregate transportation cost. I recast the mayor's problem as an optimal interval division problem. For each line segment, I ask what is the negative minimum total transportation cost for its residents for one merchant to serve them. This gives the cell function. The value of each finite interval division of Hotelling City is a sum of the cell function over the partition elements.[2] We optimally choose the interval partition where the allowed number of partition elements is the number of the merchants. This determines an optimal division of Hotelling City. Finally, the merchants are located optimally in each segment.

[1] The interval partition structure also appears in Crawford and Sobel [1982]. The relation between optimal interval division and cheap talk is discussed later in the paper.

[2] I focus on continuous distributions, so there is no problem caused by subintervals sharing common endpoints.

I identify an important and common property on cell functions: they are *submodular* over the interval structure. Submodularity first implies that dividing an interval into two subinterval is always profitable: the sum of the two subinterval values weakly exceeds that of the interval. Submodularity further asserts that the marginal gain to dividing an interval decreases as the interval shrinks.

Submodularity allows me to explore two fundamental questions for optimal interval division. First, how does the maximum value change with the number of classes? I show that if the cell function is submodular, then the maximum value exhibits decreasing marginal returns in the number of classes. This result applies to a large class of economic situations and subsumes a few existing results as special cases.[3] For a different thrust, I provide a simple sufficient condition on cell functions under which the maximum value converges rapidly, more precisely, converges as the inverse square of the number of classes. The convergence result addresses the question posed by Rogerson [2003] for costly procurement.[4]

The second basic question is how do optimal cut-offs change with the number of classes? I uncover a novel *sandwiching property:* when allowing an extra class, the new optimal cut-offs are more spread compared with the original optimal ones in the sense that each new cut point lands in a different original partition element. This helps elucidate the effect of capacity constraints on economic behavior, or the welfare consequences of relaxing capacity constraints. For instance, in the Hotelling story, the sandwiching property implies that residents in extreme locations always benefit from more merchants, since they pay lower transportation costs. For residents with non-extreme locations, the welfare effect of more merchants is indeterminate.[5]

Finally, by a characterization theorem, I show that a submodular optimal interval division problem can often be interpreted as a Hotelling story with a general transportation cost function, despite arising

[3]As an easy corollary, the "greedy algorithm" [Fox, 1966] is commonly valid for the optimal assignment among multiple projects, e.g., assigning a fixed number of merchants among multiple linear cities.

[4]Comparing Wilson [1989] and McAfee [2002], Rogerson [2003] says: "the mathematical structure of these models is somewhat different than the mathematical structure of the contracting models, so the results do not transfer immediately. However, it seems possible that there is a more general result lurking underneath all of these results."

[5]For an example of the effects of capacity constraints on economic behavior, assume the continuum as a space of signal realizations of the net return of some risky asset. In this case the partition structure captures what is later recalled [Dow, 1991]. Each interval partition is a different possible memory scheme. Now assume that in period one, a small investor chooses a memory scheme with a given partition size. In period two, he sees the realized partition element and then decides how much to invest in the risky asset. I pose a behavioral question: How does memory capacity affect the investment behavior?

from seemingly unrelated contexts like coarse matching.

The remainder of the paper is organized as follows. I briefly review the existing literature. In Section 1.2, I introduce the model. Some important classes of cell functions are discussed. In Section 1.3, I introduce and interpret the notion of submodularity. Sufficient conditions for submodularity are provided for the important classes. Section 1.4 is the main part of the paper. In Section 1.4.1, I discuss the properties of decreasing marginal returns and quadratic convergence rate. In Section 1.4.2, I discuss the sandwiching property for optimal cut-offs. In Section 1.5, I briefly discuss the structure of cell functions. I summarize the results in Section 1.6.

Related Literature. A notion of submodular set functions is involved in discrete resource allocations and some known combinatorial optimization problems [Moulin, 1988, Gul and Stacchetti, 1999, Chade and Smith, 2006, Lehmann et al., 2006]. The interpretation of submodularity within our context differs from the traditional one for utilities over commodity bundles. More importantly, the basic questions and mathematical structures involved in the division problem are very different from those in resource allocation.

Cell functions relate closely to the notion of capacity in measure theory [Choquet, 1954]. The latter is widely involved in the economics of ambiguity [Schmeidler, 1989, Marinacci and Montrucchio, 2004, Lehrer and Teper, 2008]. Cell functions capture exactly what we need to know to find the optimal interval division, while capacity within this context contains redundant information.[6] In addition, for cell functions, submodularity and other properties are much easier to obtain. In many cases, simple calculus reveals much about cell functions. This is not true for capacity.

The theory of optimal interval division developed here applies to the following specific problems: scalar quantization in information theory [Oliver et al., 1948, Lloyd, 1982, Max, 1960], coarse matching [Wilson, 1989, McAfee, 2002, Shao, 2011], bounded memory [Dow, 1991], and mechanism design with limited communication [Rogerson, 2003, Chu and Sappington, 2007, Bergemann et al., 2012, Wong, 2014].

In information theory and quantization, Oliver, Pierce, and Shannon [1948] among others show

[6] The domain of capacity is commonly required to be a σ-algebra or an algebra or minimally, closed under the operations of union and intersection. In contrast, the domain of cell functions, i.e. the collection of all closed subintervals, is not closed even under the operation of union.

that the average distortion with the squared error measure by using a uniform quantizer and n classes tends to 0 asymptotically at an order of $1/n^2$.[7] Wilson [1989], Shao [2011], and Wong [2014] within various specific economic contexts rediscover similar results. Dow [1991] first observes decreasing marginal returns in a numerical example, but suspects its generality. Within the context of nonlinear pricing, Wong [2014] first shows that the monopolist's value exhibits decreasing marginal returns. McAfee [2002], Rogerson [2003] and Chu and Sappington [2007] among others show that the gain from very coarse schemes (two classes) may be large. The large gain from very coarse schemes relies on highly special closed-form assumptions.

It is interesting to note that Crawford and Sobel [1982] contains a sandwiching property for equilibrium cut-offs in their model. There it is a direct consequence of their monotone conditions on equilibrium equations (their M and M′ conditions). In contrast, the sandwiching property here arises as a basic feature from maximizing a coarse value with submodular cell functions.

Finally, in the context of observational learning, Smith, Sørensen, and Tian [2012] discuss how optimal cut-offs change with respect to prior beliefs and show that their objective function is supermodular in belief cut-offs. Their observation relies crucially on linear utilities. More importantly, their objective function is not the *coarse value* discussed here, since the partition size there has to be fixed.

1.2 The Model and Some Examples

1.2.1 The Model

Let $\mathcal{P} = \{[a, b] : 0 \leq a \leq b \leq 1\}$ be the collection of all the closed subintervals of $[0, 1]$, with a generic element denoted by I. We refer to each element in $\mathcal{P}$ as a *cell*. Let $\tilde{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

Definition 1. *A* cell function v *maps* $\mathcal{P}$ *to* $\tilde{\mathbb{R}}$ *with* $v([z, z]) = 0$ *for each* z *in* $[0, 1]$.

Let $X^0 = \{1\}$. For each $n \geq 1$, let $X^n = \{x \in [0, 1]^n : x_1 \leq \ldots \leq x_n\}$. Each x in X^n partitions the interval $[0, 1]$ into $n + 1$ (perhaps trivial) cells. To simplify notation, put $x_0 = 0$ and $x_{n+1} = 1$.

[7] The most important non-asymptotic results are the basic optimality conditions and iterative-descent algorithms for quantizer design, discussed by Lloyd [1982] and Max [1960] among others. See Gray and Neuhoff [1998] for an excellent review.

A cell function v induces a *coarse value* for each finite interval division of $[0,1]$ by summation, which we denote by V. Formally,

$$V(x) = \sum_{k=0}^{n} v([x_k, x_{k+1}]) \tag{1}$$

for each $n \geq 0$ and x in X^n.

Optimal interval devision refers to the constrained optimization problem

$$V^n = \sup_{x \in X^n} V(x). \tag{2}$$

In (2), the allowed number of partition elements n corresponds to the capacity constraint. For instance, in the Hotelling story, it is the number of merchants that the mayor affords to distribute. For each n, let $X^*(n) = \arg\max_{x \in X^n} V(x)$. We are interested in how V^n and $X^*(n)$ change with n.

1.2.2 Two Important Classes

I now discuss two basic examples of cell functions induced by decision problems and matching problems respectively. The latter is attributed to McAfee [2002].

A *decision problem* is defined as a collection $\{d\mu_t = u(\theta, t)dF\}_{t \in T}$, where F a distribution function over $[0,1]$ and for each t, $u(\theta, t)$ is in $L^1(dF)$. The decision problem is atomless if F is atomless. A cell function v is induced by an atomless decision problem $\{u(\theta, t)dF\}_{t \in T}$ if for each cell I, we have

$$v(I) = \sup_{t \in T} \int_I u(\theta, t)dF. \tag{3}$$

In (3), we may view T as a set of actions. For each t among T, $u(\theta, t)$ is the utility from action t, depending on the type θ. The value $v(I)$ arises from choosing an action t among T so as to maximize the total utility over I weighted by F. The Hotelling story discussed before is a special case of (3). Two additional examples are provided as below.

Example (Bounded Memory)**.** In Dow [1991], a consumer observes the price of an item from the first seller ($\theta \in [0,1]$) in period 1, but is able to recall it only among finitely many partition intervals in period 2, when he observes the price offered by the second seller ($p \in [0,1]$). In period 2 and based on his memory of the price offered by the first seller, the consumer decides whether to buy the item from the second seller or return to the first seller. In period 1 and before his observation of the price offered by the first seller, the consumer decides how to efficiently use his limited memory so as to minimize his expected payment of the item. Let F be the distribution function of the possible prices offered by the first seller and G, the one of the prices offered by the second seller and assume that prices are drawn independently. The negative expected payment for each memory scheme x in X^n is given by

$$V(x) = \sum_{k=0}^{n} W[(\int_{x_k}^{x_{k+1}} \theta dF)/(\int_{x_k}^{x_{k+1}} dF)] \int_{x_k}^{x_{k+1}} dF,$$

where

$$W(z) = -\int_0^1 \min\{z, p\} dG.$$

Note that W in () is convex. Thus by convex duality, we can rewrite () as

$$V(x) = \sum_{k=0}^{n} \sup_{t \in T} \int_{x_k}^{x_{k+1}} [\theta t + y(t)] dF,$$

where $\{\theta t + y(t)\}_{t \in T}$ is the collection of supporting lines of W. It is a coarse value with (3) as its cell function where $u(\theta, t)$ is linear in θ for each t.

Example (Costly Procurement)**.** In Bergemann et al. [2012] and Wong [2014], a continuum of types (θ in $[0,1]$) with the distribution function F is divided into finitely many classes, and the qualities of some service are class-targeted. The utility function for each type θ is given by $w(\theta, t) + p$ where $t \in [0,1]$ is the quality of the service at the choice of social planner, and p the transfer the agent receives. Assume that w is supermodular with $w(\theta, 0) = 0$. Let $c(t)$ with $c(0) = 0$ be the social cost incurred by providing the service at the quality of t. Suppose that the social planner only affords to provide at most n different qualities. The maximal surplus is achieved by maximizing the *upper coarse value* among X^n with (3) as its cell function where $u(\theta, t) = w(\theta, t) - c(t)$.[8]

[8] The maximal surplus is not achieved by maximizing the *coarse value* among X^{n-1} with (3) as its cell function where $u(\theta, t) = w(\theta, t) - c(t)$. Because the surplus from the extremely lower types is negative, by precluding them from service (the planner can do this by pricing properly), social welfare increases. The surplus loss by using coarse value instead of upper coarse value could be large when n is small.

Now I turn to coarse matching [Wilson, 1989, McAfee, 2002]. Consider a two-sided matching of men and women. Suppose that the types are random variables Y and Z, and without loss of generality that these are uniformly distributed on $[0,1]$. Let $u(y,z)$ over $[0,1]^2$ be the value arising from matching the type y of man with the type z of woman. Assume that u is integrable with respect to Lebesgue measure. Given cut-offs x in X^n, matching with $n+1$ priority classes means associating Y values in $[x_k, x_{k+1}]$ with Z values in $[x_k, x_{k+1}]$ randomly for each k. The social value of matching is given by the coarse value with the cell function

$$v_m(I) = \int_I \Big[\frac{\int_I u(y,z)dz}{m(I)}\Big] dy, \tag{4}$$

where $m(\cdot)$ denotes Lebesgue measure.

1.3 Cell Division and Submodularity

1.3.1 Submodularity

In this section, I identify a common and important effect of division of cells.

Definition 2. *A cell function v is* submodular *if for each cell I_1 and I_2 with $I_1 \cap I_2 \neq \emptyset$, we have*

$$v(I_1 \cap I_2) + v(I_1 \cup I_2) \leq v(I_1) + v(I_2). \tag{5}$$

For each cell $I = [a,b]$ and c in I, denote by $\Delta v(c,I) = [v([a,c]) + v([c,b])] - v(I)$ the marginal gain (of the cut-off c) to dividing the cell I into wo cells. Note that if the cell function v is submodular, then it is always profitable to divide cells, that is, $\Delta v(c,I) \geq 0$ for each cell I and c in I. The following observation offers an intuitive and equivalent interpretation of the property of submodularity.

Observation 1. *A cell function v is submodular if and only if for each cell I and I' with $I' \subset I$ and c in I', we have*

$$\Delta v(c, I') \leq \Delta v(c, I). \tag{6}$$

Proof. It is easy to verify that the property of submodualrity holds if and only if (6) holds for each c in

I' and $I' \subset I$ where I and I' share a common ending point. The general case follows from transitivity. *Q.E.D.*

In (6), the marginal gain to dividing a cell decreases as the cell shrinks. For instance, in the example of location of merchants, the larger the segment is, the larger the transportation cost saved by introducing an additional merchant. Observation 1 adapts to our context a related equivalence relation in combinatorial optimization [Lovasz, 1982]. The interpretation of submodularity within our context is different from the traditional one for utilities over commodity bundles [Moulin, 1988, Gul and Stacchetti, 1999].[9]

Now I turn to some standard notions in lattice analysis. Let "$\vee$" and "$\wedge$" be the join and meet operators over the n-dimensional Euclidean space $(\mathbb{R}^n, \geq)$.[10] Following Topkis [1978] or Milgrom and Shannon [1994], a subset $B \subseteq \mathbb{R}^n$ is a *sublattice* (of $\mathbb{R}^n$) if for each x and x' in B, we have both $x \wedge x' \in B$ and $x \vee x' \in B$. Note that each X^n as previously defined is a sublattice of $\mathbb{R}^n$. A function f over B is *supermodular over* B, if for each x and x' in B, we have $f(x \vee x') + f(x \wedge x') \geq f(x) + f(x')$.

Observation 2. *A cell function is submodular if and only if it is supermodular over* X^2 *where each* (x_1, x_2) *in* X^2 *denotes the cell* $[x_1, x_2]$.

Based on it, the following observation is immediate.

Observation 3. *The coarse value* V *is real valued and supermodular over each* X^n *if and only if its cell function is real-valued and submodular.*

1.3.2 Sufficient Conditions for Submodularity

I now show that with quite standard conditions, the cell functions for the two basic examples in Section 1.2 are submodular. A real-valued function $g(\theta)$ over $[0, 1]$ is *weakly single crossing* (WSC)

[9] The related notions for set functions in resource allocation are as follows. Let Ω be a finite set of items, and 2^Ω the power set of Ω, that is, the collection of all the possible bundles. A decision maker has a utility function u over 2^Ω. We say that u is submodular, if $u(A \cup B) + u(A \cap B) \leq u(A) + u(B)$ for each A and B in 2^Ω; u exhibits decreasing marginal returns if $u(B \cup \{c\}) - u(B) \leq u(A \cup \{c\}) - u(A)$ for each $A \subset B$ in 2^Ω and $c \notin B$. The property of decreasing marginal returns here refers to that the marginal utility by adding a new item to the bundle B increases as B shrinks. The classic equivalence relation states that u is submodular if and only if it exhibits decreasing marginal returns.

[10] For two points x and x' in $\mathbb{R}^n$, $x' \geq x$ if $x'_i \geq x_i$ for each i; $x' > x$ if $x' \geq x$ and $x' \neq x$. For x and x' in $\mathbb{R}^n$, $x \wedge x' = (\min\{x_1, x'_1\}, \cdots, \min\{x_n, x'_n\})$, and $x \vee x' = (\max\{x_1, x'_1\}, \cdots, \max\{x_n, x'_n\})$.

(in θ) if $\theta' > \theta$ and $g(\theta) > 0$ imply $g(\theta') \geq 0$. A decision problem $\{u(\theta,t)dF\}_{t\in T}$ is *WSC ordered* if for each $t \neq t'$, either $u(\theta,t) - u(\theta,t')$ is WSC, or $u(\theta,t') - u(\theta,t)$ is WSC.

Proposition 1 (Decision Problem)**.** *The cell function induced by a WSC ordered and atomless decision problem is submodular.*

Proof. Let $\{d\mu_t = u(\theta,t)dF\}_{t\in T}$ be a WSC ordered atomless decision problem. Fix two cells I_1 and I_2 with $I_1 \cap I_2 \neq \emptyset$. I show that for each t and t' in T, we have

$$\mu_t(I_1 \cap I_2) + \mu_{t'}(I_1 \cup I_2) \leq \sup_{t''\in T} \mu_{t''}(I_1) + \sup_{t''\in T} \mu_{t''}(I_2). \tag{7}$$

First consider the case $\mu_t(I_1 \cap I_2) \leq \mu_{t'}(I_1 \cap I_2)$. We have

$$\begin{aligned} \mu_t(I_1 \cap I_2) + \mu_{t'}(I_1 \cup I_2) &\leq \mu_{t'}(I_1 \cap I_2) + \mu_{t'}(I_1 \cup I_2) \\ &= \mu_{t'}(I_1) + \mu_{t'}(I_2). \end{aligned}$$

Thus (7) holds. Next, consider the case $\mu_t(I_1 \cap I_2) > \mu_{t'}(I_1 \cap I_2)$. By the weakly single crossing condition, we have either $\mu_t(I_1 \cap I_2^C) \geq \mu_{t'}(I_1 \cap I_2^C)$ or $\mu_t(I_2 \cap I_1^C) \geq \mu_{t'}(I_2 \cap I_1^C)$. Without loss of generality, consider the first case. We have

$$\begin{aligned} \mu_t(I_1 \cap I_2) + \mu_{t'}(I_1 \cup I_2) &= \mu_t(I_1 \cap I_2) + \mu_{t'}(I_1 \cap I_2^C) + \mu_{t'}(I_2) \\ &\leq \mu_t(I_1 \cap I_2) + \mu_t(I_1 \cap I_2^C) + \mu_{t'}(I_2) \\ &= \mu_t(I_1) + \mu_{t'}(I_2). \end{aligned}$$

Thus (7) holds. Since t and t' in (7) are arbitrary, $\sup_{t\in T} \mu_t(I)$ is submodular. *Q.E.D.*

Hopenhayn and Presscott [1992] shows that the supremum over some dimensions of a supermodular multi-dimensional function remains supermodular over the remaining dimensions. Proposition 1 is not implied by their observation.[11]

[11] Suppose that T is a subinterval of the reals. The integral $\int_{x_1}^{x_2} u(\theta,t)dF$ as a function of (x_1,x_2,t) is not supermodular over $X^2 \times T$ except the trivial case that all the actions are essentially the same. To see it, first consider the non-ordered

Lemma 1 (Matching)**.** *Suppose that u in (4) is supermodular. Then v_m is submodular.*

The proof of Lemma 1 is appendicized. The single crossing condition involved in Proposition 1 is satisfied for most common decision problems [Lehman, 1988, Milgrom and Shannon, 1994, Persico, 2000, Athey, 2002]. The condition of supermodularity for matching is standard in the matching literature [Becker, 1973]. Thus, *cell functions are commonly submodular.*

1.4 General Properties

This section is the main part of the paper. The two basic questions of optimal interval division (2) mentioned in the introduction are addressed. Also, the structure of cell functions is discussed with a characterization theorem.

1.4.1 Value

We say the division problem (2) exhibits decreasing marginal returns if for each $n \geq 1$, $V^{n+1} + V^{n-1} \leq 2V^n$ holds.

Theorem 1 (DMR)**.** *Let v be submodular. Then (2) exhibits decreasing marginal returns.*

Proof of Theorem 1: Given x in X^{n+1} and x' in X^{n-1}, let $x'' = (0, x', 1)$ in X^{n+1}. We have

$$\begin{aligned} V(x) + V(x') &= V(x) + V(x'') \\ &\leq V(x \vee x'') + V(x \wedge x'') \\ &\leq 2V^n. \end{aligned} \tag{8}$$

The first inequality in (8) follows from Observation 3. The second inequality in (8) holds since the last dimension of $x \vee x''$ equals 1 and the first dimension of $x \wedge x''$ equals 0, hence both $x \vee x''$ and $x \wedge x''$ are equivalent to some elements in X^n. Since x and x' are arbitrary, the desired result follows.

pair (x_1, x_2, t') and (x_1, x_2', t) with $x_2' > x_2$ and $t' > t$. Supermodularity requires $\int_{x_2}^{x_2'} u(\theta, t')dF \geq \int_{x_2}^{x_2'} u(\theta, t)dF$. Next, consider the non-ordered pair (x_1', x_2, t) and (x_1, x_2, t') with $x_1' > x_1$ and $t' > t$, supermodularity requires $\int_{x_1}^{x_1'} u(\theta, t')dF \leq \int_{x_1}^{x_1'} u(\theta, t)dF$. This implies that for each $x_2 > x_1$ in $[0, 1]$, we have $\int_{x_1}^{x_2} u(\theta, t')dF = \int_{x_1}^{x_2} u(\theta, t)dF$ and thus $u(\cdot, t) = u(\cdot, t')$ a.e. in F.

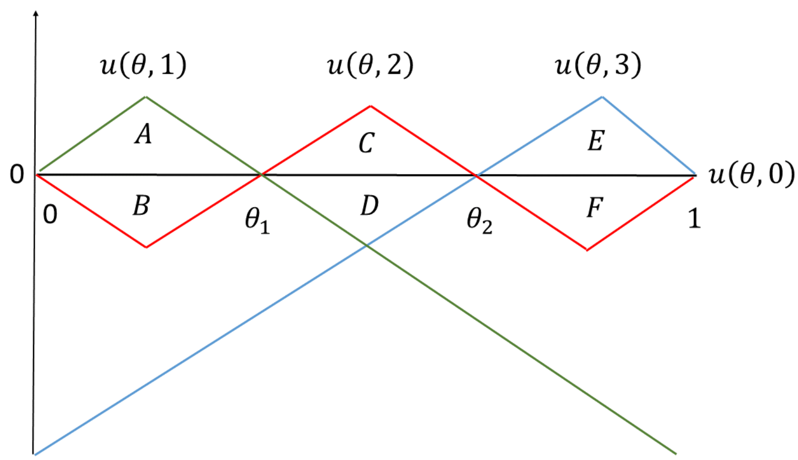

Figure 1: Failure of Decreasing Marginal Returns.

Q.E.D.

The intuition for Theorem 1 is clear from the preceding discussion of the submodular effect of cell division, i.e., Observation 1. Theorem 1 implies a large class of coarse values obeying decreasing marginal returns, including single crossing ordered decision problems by Proposition 1, coarse matching with supermodular primitive by Lemma 1, and the class of decision problems involved in Theorem 4 in Section 1.5.[12]

Example (Failure of Decreasing Marginal Returns)**.** Consider the decision problem (3). Assume that F is uniform and there are four actions: $T = \{0, 1, 2, 3\}$. Action 0 is safe with $u(\theta, 0) \equiv 0$. The utilities for the actions 1, 2, 3 are depicted respectively by the green, red, and blue lines in Figure 1. The areas of A, B, C, D, E, F are the same, denoted by $\Delta > 0$. In this case, we have $V^2 - V^1 = 2\Delta > V^1 - V^0 = \Delta$ and decreasing marginal returns fail. One may check that the cell function in this case is not submodular, because the actions 0 and 2 are not weakly single crossing ordered. On the other hand, if we drop action 0 and consider instead the set of actions $T' = \{1, 2, 3\}$, we have $V^2 - V^1 = 2\Delta = V^1 - V^0$ and decreasing marginal returns are restored, despite a knife-edge case. This is because the set of actions $T' = \{1, 2, 3\}$ is weakly single crossing ordered and the induced cell function now is submodular.

[12]Within the context of nonlinear pricing, Wong [2014] shows that the monopolist's value exhibits decreasing marginal returns. Wong [2014]'s observation is a special case of the single crossing ordered decision problems.

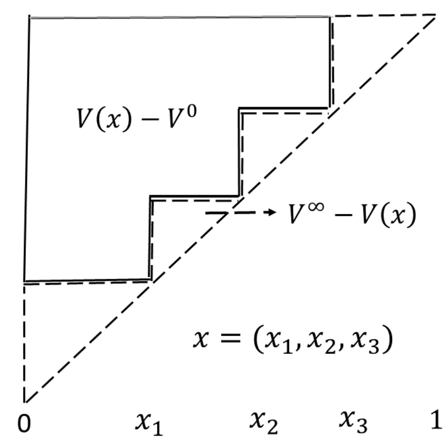

Figure 2: Relative Gain as Integral of v_{12}. The difference $V(x) - V^0$ is the area enclosed by the solid line with v_{12} as the density function. The difference $V^\infty - V^0$ is the area of the whole triangle X^2 with v_{12} as the density function.

Next, I turn to the problem of convergence. Assume that the cell function v as a function over X^2 is second-order cross differentiable and that v_{12} is integrable. If v is submodular, then v_{12} is positive (i.e. $v_{12} \geq 0$). The following observation says that the cross derivative function v_{12} precisely captures the potential gains in the involved division problem. Let $V^\infty = \lim_{n\to\infty} V^n$.

Lemma 2. *Suppose that v_{12} is positive. Then we have*

$$V(x) - V^0 = \sum_{k=1}^{n} \int_0^{x_k} \int_{x_k}^{x_{k+1}} v_{12} dz_1 dz_2, \tag{9}$$

for each $n \geq 1$ and x in X^n, and thus

$$V^\infty - V^0 = \int_{X^2} v_{12} dx_1 dx_2. \tag{10}$$

Proof. It is easy to verify directly that (9) holds, which gives (10) by $v_{12} \geq 0$. *Q.E.D.*

For each x in X^n, the difference $V^\infty - V(x)$ is simply the summation of the areas of the $n+1$ small triangles with v_{12} as the density function, as illustrated by Figure 2. Based on this observation, I now provide a simple sufficient condition on the cell function such that V^n converges at the rate of the inverse square of the number of classes. A function f over X^2 is *M-sublinear*, if $f(x_1, x_2) \leq M(x_2 - x_1)$

for each (x_1, x_2) in X^2.

Theorem 2 (Inverse-Square Convergence)**.** *Suppose that v_{12} is positive and M-sublinear. Then for each n, we have*

$$V^\infty - V^n < \frac{M}{6n^2}.$$

Proof. For each n, let $x(n) = (1/(n+1), \dots, k/(n+1), \dots, n/(n+1))$. By (9) and (10) in Lemma 2, we have

$$V^\infty - V^n \le V^\infty - V(x(n)) = \sum_{k=0}^{n} \int_{k/(n+1)}^{(k+1)/(n+1)} \int_{z_1}^{(k+1)/(n+1)} v_{12} dz_2 dz_1.$$

The desired result then follows from the observation that for each k,

$$\begin{aligned} \int_{k/(n+1)}^{(k+1)/(n+1)} \int_{z_1}^{(k+1)/(n+1)} v_{12} dz_2 dz_1 &\le \int_{k/(n+1)}^{(k+1)/(n+1)} \int_{z_1}^{(k+1)/(n+1)} M(z_2 - z_1) dz_2 dz_1 \\ &\le \frac{M}{6(n+1)^3}. \end{aligned}$$

Q.E.D.

Theorem 2 provides a simple unified perspective on the inverse-square convergence rate (see Appendix 1.7.2 for discussion).

1.4.2 Cut-offs

Two points x in X^n and x' in X^{n+1} are *sandwiched* if for each k, we have

$$x'_k < x_k < x'_{k+1}. \tag{11}$$

For instance, in the sequence of uniform cut-offs $\{(1/(n+1), \dots, n/(n+1))\}_{n \in \mathbb{N}}$, the cut-offs for each n and $n+1$ are sandwiched.

We say that x and x' are weakly sandwiched if both inequalities in (11) are replaced by "$\le$". Proposition 2 below implies that the weak sandwiching property holds commonly for extreme optimal cut-offs, based on Observation 3, and the fact that cell functions are commonly submodular.

Proposition 2 (A Weak Sandwiching Property). *Let v be real-valued, continuous, and submodular. Then for each $n \geq 1$, the greatest (least) element in $X^*(n)$ and the greatest (least) element in $X^*(n+1)$ are weakly sandwiched.*

Before addressing the sandwiching property, I shall now pause to discuss some regular properties of cell functions. For each cell I, let $X_I^2 = \{(x_1, x_2) \in I^2 : x_1 \leq x_2\}$. A cell function v has *nontrivial convex support*, if there is a cell $[b, c]$ with $b < c$ such that v is strictly supermodular over $X^2_{[b,c]}$ and modular over both $X^2_{[0,b]}$ and $X^2_{[c,1]}$. In this case, call $[b, c]$ the support of v, which we denote by $\text{supp}(v)$.

A cell function v is *regular* if 1) it is real-valued, continuous and submodular with nontrivial convex support; and 2) v over $X^2_{\text{supp}(v)}$ is partially differentiable and each v_i increases strictly in x_{-i}; and 3) $v(0, z)$ is both left-hand and right-hand differentiable in z over $(0, 1)$ with $v'(0, z+) \geq v'(0, z-)$ and the same also applies to $v(z, 1)$. A sufficient condition for regularity is that v as a function over X^2 is partially differentiable and each v_i increases strictly in x_{-i} or simply but more strongly $v_{12} > 0$. In Appendix 1.7.5, I show that regular cell functions are common for the important classes.

Theorem 3 (A Sandwiching Property). *Let v be regular. Then each x in $X^*(n)$ and x' in $X^*(n+1)$ are sandwiched for each $n \geq 1$.*

Crawford and Sobel [1982] derive a sandwiching property for equilibrium cut-offs (their Lemma 3), which is a consequence of their monotone assumptions on equilibrium equations (their M and M′ conditions).[13] In contrast, the sandwiching property here arises as a basic feature from maximizing supermodular coarse values. The example below illustrates the difference between optimal division and equilibrium analysis.

Example (Efficient Language vs Equilibrium Language). Consider a version of Crawford and Sobel [1982] where the sender and the receiver share the same preference [Lipman, 2009, Gerhard Jager and Riedel, 2011]. Player 1 observes the state θ, a random draw from $[0, 1]$ according to a continuous and positive density function f. He then chooses a message m from a set M with $|M| = n + 1$. Player

[13] To the best of my knowledge, the most general conditions for this kind of monotonicity, given by Szalay [2012], require smooth location forms of the utilities and log-concave densities.

2 observes this message but not θ and then chooses an action t from T. Assume that the messages themselves are costless. Both agents have the same utility function $u(\theta, t)$ which is twice continuously differentiable with $u_{22} < 0$ and $u_{12} > 0$. The set of efficient cut-offs corresponds to $\arg\max_{x \in X^n} V(x)$ and V is the coarse value with (3) as its cell function where F admits the density function f. By Lemma 9 in Appendix 9 and Theorem 3, the efficient cut-offs are sandwiched with partition sizes. The set of equilibrium cut-offs are the ones satisfying the indifference condition:

$$u(x_k, t^*(x_{k-1}, x_k)) = u(x_k, t^*(x_k, x_{k+1})),$$

where $t^*(x_{k-1}, x_k)$ solves $\max_{t \in T} \int_{x_{k-1}}^{x_k} u(\theta, t) f(\theta) d\theta$. For equilibrium cut-offs to be sandwiched, we need the M and M′ properties in Crawford and Sobel [1982].

1.5 Characterizing Submodular Cell Functions

In this subsection, I discuss briefly the structure of cell functions. Specifically, I show that submodular cell functions can often be induced by decision problems. This is based on a tight relation between cell functions and the notion of capacity in measure theory. The latter is first discussed by Choquet [1954].

First, I introduce some common properties of cell functions. A cell function is *monotone* if for each cell I_1 and I_2 with $I_1 \subset I_2$, $v(I_1) \leq v(I_2)$. In (3), if $u(\theta, t) \geq 0$, then the induced cell function is monotone. Let X^2 be endowed with the city block metric ρ, i.e., $\rho(x, x') = |x_1 - x'_1| + |x_2 - x'_2|$ for each x and x' in X.[14] A cell function v is *continuous* if it is continuous over (X^2, ρ); v is *Lipschitz continuous* if there exits some $M > 0$ such that for each x and x' in X^2, we have $|v(x) - v(x')| \leq M\rho(x, x')$.

Next, I adapt to our context the notion of *weakly filtering* which is involved in the characterization of submodular capacity [Anger, 1977]. Let $\mathcal{Q}$ be a subset of $\mathcal{P}$.

Definition 3 (Weakly Filtering). *A decision problem $\{d\mu_t = u(\theta, t)dF\}_{t \in T}$ is weakly filtering on $\mathcal{Q}$, if for each cell I and I' in $\mathcal{Q}$ with $I \subseteq I'$, we have both $\sup_{t \in T} \mu_t(I) = \mu_{t'}(I)$ and $\sup_{t \in T} \mu_t(I') = \mu_{t'}(I')$ for*

[14] The city block metric is just for convenience. All metrics strongly equivalent to ρ apply, including the Euclidean metric.

some t' in T.

A subset $\mathcal{Q}$ of $\mathcal{P}$ is *dense* in $\mathcal{P}$, if for each cell I in $\mathcal{P}$ and open set O with $I \subseteq O$, there exists a cell I' in $\mathcal{Q}$ with $I \subseteq I' \subseteq O$. The main result in this section is stated below.

Theorem 4 (Characterization)**.** *Let v be a real-valued cell function. Then*

1. *v is submodular, continuous and monotone if and only if it can be induced by some atomless decision problem $\{d\mu_n = u(\theta, n)dF\}_{n\in\mathbb{N}}$ weakly filtering on a dense subset of $\mathcal{P}$ such that $u(\theta, n) \geq 0$ and $\sup_{n\in\mathbb{N}} \mu_n([x_1, x_2])$ is finite and continuous over X^2;*
2. *v is submodular and Lipschitz continuous if and only if it can be induced by some decision problem $\{u(\theta, n)dm\}_{n\in\mathbb{N}}$ weakly filtering on a dense subset of $\mathcal{P}$ such that $\sup_{n\in\mathbb{N}} |u(\theta, n)| < M < \infty$.*

The following result applies Theorem 4 to coarse matching in McAfee [2002].

Corollary 1. *Suppose that u in (4) is supermodular and bounded. Then v_m can be induced by some decision problem with the properties stated in (2) of Theorem 4.*

Proof. If u in (4) is bounded, then v_m is Lipschitz continuous. The desired result then follows from Lemma 1 and Theorem 4. *Q.E.D.*

1.5.1 Proof of Theorem 4

Let $\mathfrak{K}$ be the collection of all the compact subsets of $[0, 1]$, and $\mathbb{R}_+ = \{z \in \mathbb{R} : z \geq 0\}$. Following Anger [1977] or Adamski [1977], a *capacity* c maps $\mathfrak{K}$ to $\mathbb{R}_+$ such that 1) $c(\emptyset) = 0$, and 2) $c(K) \leq c(K')$ if $K \subset K'$, and 3) for each K in $\mathfrak{K}$ and each $\epsilon > 0$, there exists an open neighborhood O of K, such that $c(K') - c(K) < \epsilon$ for every K' in $\mathfrak{K}$ with $K \subseteq K' \subset O$. A capacity c is submodular if for each A and B in $\mathfrak{K}$, we have $c(A \cap B) + c(A \cup B) \leq c(A) + c(B)$. Submodular capacity is the focus in many analyses [Choquet, 1954]. A cell function v *admits submodular capacity extension* if there exists a submodular capacity c such that v and c coincide at each cell.

Lemma 3 (Cell Function and Capacity)**.** *A cell function admits atomless submodular capacity extension if and only if it is continuous, submodular and monotone.*

A measure μ is *dominated* by a capacity c, if for each $K \in \mathfrak{K}$, we have $\mu(K) \leq c(K)$. Given a capacity c, let $\mathfrak{U}(c)$ be the collection of all the *positive* measures dominated by c. The next result, due to Anger [1977] and Adamski [1977] among others, says that a submodular capacity is upper envelopes of the positive measures dominated by it.

Lemma 4 (Anger [1977], Adamski [1977])**.** *Let c be a submodular capacity. Then $c(K) = \max_{\mu \in \mathfrak{U}(c)} \mu(K)$ for each $K \in \mathfrak{K}$. Moreover, for each $K \subset K'$ in $\mathfrak{K}$, there exists a μ in $\mathfrak{U}(c)$ such that $\mu(K) = c(K)$ and $\mu(K') = c(K')$.*

Proof of the only if part of (1) in Theorem 4: Let v be continuous, monotone and submodular, and let c be the submodular capacity extension of v in Lemma 3. By Lemma 4, for each cell I, we have

$$v(I) = \max_{\mu \in \mathfrak{U}(c)} \mu(I).$$

Now let $\mathcal{Q}$ be the collection of all the cells with rational ending points, which is dense in $\mathcal{P}$. For each pair of cells (I, I') in $\mathcal{Q}^2$ with $I \subseteq I'$, we choose a positive measure μ in $\mathfrak{U}(c)$ such that $\mu(I) = v(I)$ and $\mu(I') = v(I')$. Such a μ exists by Lemma 4. Denote this collection of measures by $\{\mu_n\}_{n \in \mathbb{N}}$, which is weakly filtering on $\mathcal{Q}$.

Next, I show that for each cell I, we have

$$v(I) = \sup_{n \in \mathbb{N}} \mu_n(I). \tag{12}$$

(12) holds if $I = [z, z]$ for some z. Now consider the non-degenerate cells. For each $\epsilon > 0$, there exists a cell I' in $\mathcal{Q}$ with $I' \subseteq I$ and $v(I') > v(I) - \epsilon$, due to continuity of v. Let $\mu_{n'}(I') = v(I')$. Since $\mu_{n'}$ is positive, we have $\mu_{n'}(I) \geq \mu_{n'}(I') > v(I) - \epsilon$ and so

$$v(I) \geq \sup_{n \in \mathbb{N}} \mu_n(I) > v(I) - \epsilon. \tag{13}$$

Since ϵ is arbitrary, (13) implies (12). The desired result then follows from (12), the continuity of v, and the fact that each μ_n is absolutely continuous with respect to $\sum_{n=1}^{\infty} (1/2^n) \mu_n$.

Q.E.D.

The notion of cell functions is perfectly suited for optimal interval division in the sense that it captures exactly what we need to know about the problem to find the optimal interval division. In contrast, capacity within this context contains redundant information. Furthermore, for cell functions, submodularity or other properties are much easier to obtain. In many cases, simple calculus reveals lots about cell functions – this is not true for capacity.

1.6 Conclusion

In this paper, I developed a theory of optimal interval division for capacity-constrained problems. My companion paper applies some of the analysis here to language and mechanism design with limited communication [Tian, 2015]. This paper focuses on scalar cases. Whether and how the results here extend to multidimensional contexts are important and deserve further efforts.

1.7 Appendix

1.7.1 Proof of Lemma 1

Proof. Let $x_1 < x_2' < x_2''$ in $[0,1]$. Since u is integrable over $[0,1]^2$, the cell function $v_m(x_1,x_2) = (\int_{[x_1,x_2]^2} u(y,z)dydz)/(x_2-x_1)$ as a function of x_2 is absolutely continuous and thus differentiable almost everywhere over $[x_2', x_2'']$. Next, by rewriting the integral $\int_{[x_1,x_2]^2} u(y,z)dydz$, we have

$$\int_{x_1}^{x_2} \Big\{ \int_{x_1}^{t} [u(y,t)+u(t,y)]dy \Big\} dt = v_m(x_1,x_2)(x_2-x_1). \tag{14}$$

Both sides of (14) are differentiable at x_2 if and only if v_m is differentiable there, with

$$\frac{\partial v_m(x_1,x_2)}{\partial x_2} = \frac{1}{(x_2-x_1)^2} \int_{[x_1,x_2]^2} w(x_2,y,z)dydz, \tag{15}$$

where

$$w(x_2,y,z) = u(y,x_2) + u(x_2,z) - u(y,z). \tag{16}$$

Since u is supermodular, the function w is increasing in (y, z) over $[0, x_2]^2$ for each x_2. Thus by (15) we have $v'_m(x'_1, x_2) \geq v'_m(x_1, x_2)$ for almost every x_2 in $[x'_2, x''_2]$ whenever $x_1 < x'_1 < x'_2$. So the difference

$$v_m(x_1, x''_2) - v_m(x_1, x'_2) = \int_{x'_2}^{x''_2} \frac{\partial v_m(x_1, x_2)}{\partial x_2} dx_2, \tag{17}$$

which follows from Theorem 7.20 in Rudin [1987], increases in x_1. This shows that the cell function v_m is supermodular over X^2. *Q.E.D.*

1.7.2 Applications of Theorem 2

I now apply Theorem 2 to the important classes. First consider decision problem (3). Suppose that in (3) F admits continuous and positive density function f, and T is a compact interval of the reals, and $u(\theta, t)$ over $[0, 1] \times T$ is twice continuously differentiable with $u_{22} < 0$ and u_2 strictly single crossing in θ for each t. For each $x_1 < x_2$ in $[0, 1]$, let $t^*(x_1, x_2)$ solve (3) for $I = [x_1, x_2]$. By the Envelope Theorem, we have

$$v_{12}(x_1, x_2) \quad = \quad u_2(x_2, t^*(x_1, x_2))f(x_2)\frac{\partial t^*(x_1, x_2)}{\partial x_1},$$

which is α-Lipschitz continuous. Thus by Theorem 2, the inverse-square convergence rate holds for decision problems with the conditions stated here.[15] Turning to matching (4), Shao [2011] generalizes Wilson [1989] and shows that if u in (4) is twice differentiable and $u_{12} \geq 0$ and is bounded, then the inverse-square convergence rate holds for coarse matching. One can verify that the condition in Shao [2011] implies that the cross derivative of v_m is α-Lipschitz continuous, and thus is a special case of Theorem 2.

[15] Bergemann et al. [2012] consider a special case with linear quadratic utilities. Wong [2014] considers the case with separable utilities. They both consider more general distribution functions.

1.7.3 Proof of Proposition 2

Lemma 5. *Let V be a coarse value with regular cell function. Then for each x in $X^*(n)$ and x' in $X^*(n+1)$, we have*

$$\{x \vee x'_{-(n+1)}, x \wedge x'_{-1}\} \subseteq X^*(n) \quad \textit{and} \quad \{(x'_1, x \vee x'_{-1}), (x \wedge x'_{-(n+1)}, x'_{n+1})\} \subseteq X^*(n+1).$$

Proof. By $V(x') - V(x \wedge x'_{-(n+1)}, x'_{n+1}) \geq 0$ and the supermodularity of V, we have $V(x \vee x'_{-(n+1)}, 1) - V(x, 1) \geq 0$ and so $x \vee x'_{-(n+1)} \in X^*(n)$. If $V(x') - V(x \wedge x'_{-(n+1)}, x'_{n+1}) > 0$, then $V(x \vee x'_{-(n+1)}) - V(x) > 0$, contradicting $x \in X^*(n)$. Thus $V(x') - V(x \wedge x'_{-(n+1)}, x'_{n+1}) = 0$ and $(x \wedge x'_{-(n+1)}, x'_{n+1}) \in X^*(n+1)$. The rest is similar and omitted.

Q.E.D.

Proof of Proposition 2. Since V is continuous and supermodular and each X^n is compact, each $X^*(n)$ is a nonempty compact sublattice of $\mathbb{R}^n$, by Corollary 2 in Milgrom and Shannon [1994]. Thus for each $n \geq 1$, both the greatest element and the least element in $X^*(n)$ exist (e.g., Corollary 2.3.2 in Topkis [1998]). Let x be the greatest element in $X^*(n)$, and x' the greatest element in $X^*(n+1)$. If $x \not\leq x'_{-1}$, we have $x' < (x'_1, x \vee x'_{-1})$, which contradicts that x' is the greatest element in $X^*(n+1)$, since $(x'_1, x \vee x'_{-1})$ is in $X^*(n+1)$ by Lemma 2. If $x'_{-(n+1)} \not\leq x$, we have $x < x \vee x'_{-(n+1)}$, which contradicts that x is the greatest element in $X^*(n)$, since $x \vee x'_{-(n+1)}$ is in $X^*(n)$ by Lemma 2. Thus $x'_k \leq x_k \leq x'_{k+1}$ for each k. The case of the least element is similar and omitted. *Q.E.D.*

1.7.4 Proof of Theorem 3

A subset of $\mathbb{R}^n$, say, Z, is *strictly totally ordered*, if for each $y \neq z$ in Z, either $y << z$ or $z << y$ where the notation "$z << y$" means $z_k < y_k$ for each k.

Lemma 6. *Let V be a coarse value with regular cell function. Then $\forall n \geq 1$, $X^*(n)$ is strictly totally ordered.*

Proof. Step 1. I first show that for each $x \leq x'$ in $X^*(n)$, either $x = x'$ or $x << x'$. By Lemma 7 below and Corollary 2 in Milgrom and Segal [2002], the FOC applies:

$$v_2(x_{k-1}, x_k) + v_1(x_k, x_{k+1}) = 0,\ k = 1, \dots, n. \tag{18}$$

If $x'_k = x_k$ for some k between 1 and n, then we have $x'_{k+1} = x_{k+1}$ and $x'_{k-1} = x_{k-1}$, by the condition that each v_i increases strictly in x_{-i}. Repeating this logic, we have $x' = x$.

Now let $x \neq x'$ in $X^*(n)$. Since $X^*(n)$ is a sublattice, $x \vee x' \in X^*(n)$. Since $x \vee x' \geq x$, Step 1 implies either $x \vee x' = x$ or $x << x \vee x'$. The former case implies $x' \leq x$ and thus $x' << x$ by Step 1 and $x \neq x'$. The latter case implies $x << x'$. *Q.E.D.*

Lemma 7. *Let V be a coarse value with a regular cell function v. Then for each x in $X^*(n)$ and k, x_k is in the interior of supp(v) with $x_k \neq x_j$ whenever $k \neq j$.*

Proof. First, $\Delta v(c, I) > 0$ for each non-degenerate $I \subseteq \text{supp}(v)$ and c in the interior of I. As a result, for each x in $X^*(n)$, we have $x_k \neq x_j$ for $k \neq j$ among $\{0, 1, \dots, n, n+1\}$.

Now suppose that $\text{supp}(v) = [b, 1]$ where $b > 0$. I show $x_1 > b$. Since v is modular over $[0, b]$, if we substitute by b the components of x which are strictly smaller than b, the new cut-off vector remains optimal. Thus it suffices to show that for each $x \in X^*(n)$, we have $x_1 \neq b$. Suppose the contrary that $x_1 = b$. Then b solves $\max_{z \in (0, x_2)} v(0, z) + v(z, x_2)$. Now I argue that for each e in (b, x_2), b also solves

$$\max_{z \in (0, e)} v(0, z) + v(z, e). \tag{19}$$

Suppose not. Since v is modular over $[0, b]$, there exists some q in (b, e) such that $v(0, q) + v(q, e) > v(0, b) + v(b, e)$, or equivalently $v(q, e) - v(b, e) > v(0, b) - v(0, q)$. Then by $q > b$, $x_2 > e$ and submodularity of v, we have $v(q, x_2) - v(b, x_2) > v(0, b) - v(0, q)$ or equivalently $v(0, q) + v(q, x_2) > v(0, b) + v(b, x_2)$. This is a contradiction. Thus (19) holds. By Corollary 2 in Milgrom and Segal [2002], FOC holds at b in (19) for each e in $(b, x_2]$. But this is impossible, since by regularity, $v_1(b, e)$ increases strictly in e over (b, c). The other cases are similar and omitted. *Q.E.D.*

Proof of Theorem 3: Suppose that v is regular with $\text{supp}(v) = [b, c]$. Let x in $X^*(n)$ and x' in $X^*(n+1)$. By Lemma 6 and $(x \wedge x'_{-(n+1)}, x'_{n+1}) \in X^*(n+1)$ in Lemma 2, we have $x \wedge x'_{-(n+1)} = x'_{-(n+1)}$

or equivalently $x'_{-(n+1)} \le x$. Next, by the submodularity of v, we have $v_1(x_n, c) \le v_1(x_n\pm, 1)$. Then we have $v_1(x_n, x_{n+1}) < v_1(x_n\pm, 1)$ by $x_{n+1} < c$ in Lemma 7. Step 1 in the proof of Lemma 6 then implies $x'_{-(n+1)} << x$. The argument for $x << x'_{-1}$ is similar and omitted. Thus $x'_k < x_k < x'_{k+1}$. *Q.E.D.*

Proof of Lemma 3

Proof. (Necessity). Let v be a cell function and c its submodular capacity extension. By continuity of c from above, the cell function $v(x_1, x_2)$ over X^2 is left continuous at x_1 and right continuous at x_2. To see that v is right continuous at x_1, suppose $x_1 < x_2$, and let $\{z_n\}$ be any sequence convergent to x_1 from above with each $z_n < x_2$. By submodularity and monotonicity of c, we have

$$v(x_1, x_2) - v(x_1, z_n) \le v(z_n, x_2) \le v(x_1, x_2). \tag{20}$$

Since $h(z) = v(x_1, z)$ as a function of z is right continuous at $z = x_1$, we have $\lim_{n\to\infty} v(x_1, z_n) = 0$ by $v(x_1, x_1) = 0$, and so $\lim_{n\to\infty} v(z_n, x_2) = v(x_1, x_2)$ by (20). Similarly, v is left continuous at x_2. The properties of submodularity and monotonicity are satisfied automatically.

(Sufficiency). (Step 1). Let $\mathfrak{O}$ be the collection of all the open sets which are *finite* unions of disjoint open intervals, including also the empty set. $\mathfrak{O}$ is closed under the operations of union and intersection. We first define a function $\tilde{c}$ over $\mathfrak{O}$ through summation of values of cells. That is, for each U as the union of disjoint open intervals $\{(a_n, b_n)\}_{1\le n\le N}$, let $\tilde{c}(U) = \sum_{n=1}^{N} v([a_n, b_n])$.

Observation 4. *1) $\tilde{c}((a, b)) = v([a, b])$ for each $a \le b$, and 2) for each cell I and each open set U in $\mathfrak{O}$ covering I, we have $\tilde{c}(U) \ge v(I)$ by the monotonicity of v and thus $v \ge 0$.*

I now show that $\tilde{c}$ is submodular over $\mathfrak{O}$. Let U_1 be the union of disjoint open intervals $\{(a_n, b_n)\}_{n\in N_1}$ and U_2 the union of $\{(a_n, b_n)\}_{n\in N_2}$. Choose the index sets N_1 and N_2 such that $N_1 \cap N_2 = \emptyset$. Denote $U_1 \cup U_2$ as the union of disjoint open intervals $\{(c_n, d_n)\}_{n\in N}$, and $U_1 \cap U_2$ as the union of $\{(e_n, f_n)\}_{n\in N_4}$. We show

$$\tilde{c}(U_1 \cup U_2) + \tilde{c}(U_1 \cap U_2) \le \tilde{c}(U_1) + \tilde{c}(U_2). \tag{21}$$

First, we can express each (c_n, d_n) as

$$(c_n, d_n) = \cup_{j=1}^{J_n}(a_{n_j}, b_{n_j}), \tag{22}$$

where 1) each $n_j \in N_1 \cup N_2$ and 2) if $J_n > 1$, then $b_{n_j} > a_{n_{j+1}}$ for each j, and if $n_j \in N_i$, then $n_{j+1} \in N_{-i}$ for $i = 1, 2$. Let $N_3 = \cup_{n \in N}\{n_1, \ldots, n_{J_n}\}$.

Applying the general submodular inequalities in Choquet [1954] to (22) successively, we have

$$v([c_n, d_n]) + \sum_{j=1}^{J_n - 1} v([a_{n_{j+1}}, b_{n_j}]) \leq \sum_{j=1}^{J_n} v([a_{n_j}, b_{n_j}]). \tag{23}$$

Summing (23) across n in N, we have

$$\tilde{c}(U_1 \cup U_2) + \sum_{n \in N: J_n > 1} \sum_{j=1}^{J_n - 1} v([a_{n_{j+1}}, b_{n_j}]) \leq \sum_{n \in N_3} v([a_n, b_n]). \tag{24}$$

Note that each $(a_{n_{j+1}}, b_{n_j})$ in the second summation on the left hand side of (24) equals some (e_n, f_n). Next, for each (e_n, f_n) that does not appear in the second summation on the left hand side of (24), we have $(e_n, f_n) = (a_n, b_n)$ for some n in $N_1 \cup N_2$ and $n \notin N_3$. This implies that (21) holds.

(Step 2). Define the function c over $\mathfrak{K}$ by covering and using elements in $\mathfrak{O}$, that is, $c(K) = \inf\{\tilde{c}(U) : U \in \mathfrak{O}, K \subseteq U\}$ for each K in $\mathfrak{K}$. The function c is well-defined because each compact set can be covered by some element in $\mathfrak{O}$. By the definition, c is monotone and continuous from above with $c(\emptyset) = 0$ and thus a capacity. In addition, by Observation 4 and continuity of v, c coincides with v on the cells.

I now show that c is submodular over $\mathfrak{K}$. For each K_1 and K_2 in $\mathfrak{K}$, let U_1 in $\mathfrak{O}$ cover K_1 and U_2 in $\mathfrak{O}$ cover K_2. Then $U_1 \cup U_2$ covers $K_1 \cup U_2$, and $U_1 \cap U_2$ covers $K_1 \cap K_2$. Thus we have

$$c(K_1 \cup K_2) + c(K_1 \cap K_2) \leq \tilde{c}(U_1 \cup U_2) + \tilde{c}(U_1 \cap U_2) \leq \tilde{c}(U_1) + \tilde{c}(U_2).$$

Since U_1 and U_2 are arbitrary open covers of K_1 and K_2 respectively, we have

$$c(K_1 \cup K_2) + c(K_1 \cap K_2) \leq c(K_1) + c(K_2). \tag{25}$$

Q.E.D.

Finishing the Proof of Theorem 4

Proof of the If Part of (1) in Theorem 4: Let $\{\mu_n\}_{n\in\mathbb{N}}$ satisfy the conditions in (1), and $v(I) = \sup_{n\in\mathbb{N}} \mu_n(I)$. First, we have $v([z, z]) = 0$ for each $z \in [0, 1]$ and so v is a cell function. The properties of continuity and monotonicity of v hold trivially. Before proving the property of submodularity, I first point out the following fact. Since v is continuous over X^2 and X^2 is compact, v is uniformly continuous over X^2. Then since $v([z, z]) = 0$ for each z in $[0, 1]$, we have that for each $\epsilon > 0$, there exists a $\delta > 0$ such that for each I with $m(I) < \delta$, $v(I) < \epsilon$. Since each μ_n is dominated by v, we have $\mu_n(I) < \epsilon$.

Now fix arbitrary numbers $a < b \leq c < d$ in $[0, 1]$ and $\epsilon > 0$. Choose $\delta > 0$ such that for each cell I with $m(I) < \delta$, we have $\mu_n(I) < \epsilon/4$ for each n in $\mathbb{N}$. Next, choose two cells $I_1 = [e, f]$ and $I_2 = [g, h]$ in $\mathcal{Q}$ such that 1) $e \leq a < g \leq b \leq c \leq h < d \leq f$, and 2) $m([e, f]) - m([a, d]) \leq \delta$ and $m([g, h]) - m([b, c]) \leq \delta$. Now let $\mu_{n'}$ in $\{\mu_n\}_{n\in\mathbb{N}}$ be such that $\mu_{n'}([e, f]) = v([e, f])$ and $\mu_{n'}([g, h]) = v([g, h])$. Then we have

$$\begin{aligned}
v([a, d]) + v([b, c]) &\leq v([e, f]) + v([g, h]) = \mu_{n'}([e, f]) + \mu_{n'}([g, h]) \\
&= \mu_{n'}([e, h]) + \mu_{n'}([g, f]) = \mu_{n'}([a, c]) + \mu_{n'}([b, d]) \\
&+ [\mu_{n'}([e, a]) + \mu_{n'}([g, b]) + \mu_{n'}([c, h]) + \mu_{n'}([d, f])] \\
&\leq v([a, c]) + v([b, d]) + \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we have $v([a, d]) + v([b, c]) \leq v([a, c]) + v([b, d])$. Since $[a, c]$ and $[b, d]$ are arbitrary two cells with nonempty intersection, the desired result follows. *Q.E.D.*

Lemma 8. *If a cell function is monotone, submodular and Lipschitz continuous, then it can be induced by some decision problem* $\{u(\theta, n)dm\}_{n\in\mathbb{N}}$ *with* $\sup_{n\in\mathbb{N}} |u(\theta, n)| < M < \infty$.

Proof. Let v satisfy the conditions in the above lemma with Lipschitz constant M, and $\{d\mu_n = u(\theta, n)d\mu\}_{n\in\mathbb{N}}$ the decision problem inducing v in part (1) of Theorem 4. For each μ_n, the fact that $\mu_n(I) \leq v(I) \leq Mm(I)$ for each cell I implies that μ_n is absolutely continuous with respect to the Lebesgue measure m with $d\mu_n/dm \leq M$. *Q.E.D.*

Proof of (2) in Theorem 4: The if part of (2) is same as the if part of (1). I show the only if part of (2). Consider the cell function $v'(I) = v(I) + M'm(I)$ with $M' > M$. As a summation of two submodular and Lipschitz continuous cell functions ($M'm(I)$ is modular and Lipschitz continuous), v' is submodular and Lipschitz continuous. For each cell $I \subset I'$, we have $v'(I') - v'(I) = [v(I') - v(I)] + M'[m(I') - m(I)] \geq (M' - M)[m(I') - m(I)] \geq 0$. Thus v' is monotone. By Lemma 8, v' can be induced by some decision problem $\{d\mu_n = u(\theta, n)dm\}_{n\in\mathbb{N}}$ weakly filtering on a dense subset $\mathcal{Q}$ of $\mathcal{P}$, with $\sup_{n\in\mathbb{N}} |u(\theta, n)| < M$ for some $M < \infty$. As a result, $v(I)$ can be induced by $\{[u(\theta, n) - M']dm\}_{n\in\mathbb{N}}$. *Q.E.D.*

1.7.5 Regular Cell Functions

A real-valued function $g(\theta)$ over $[0,1]$ is *single crossing* (SC) if $\theta' > \theta$ and $g(\theta) > 0$ imply $g(\theta') > 0$; *strictly single crossing* (SSC) if $\theta' > \theta$ and $g(\theta) \geq 0$ imply $g(\theta') > 0$; A utility function $u(\theta, t)$ over $[0,1] \times T$ where T is a nondegenerate and compact subinterval of the reals, is *regular* if 1) it is continuous and SC ordered and 2) partially differentiable in t with u_2 continuous in (θ, t), SSC in θ and strictly decreasing in t, and 3) $\max_{t\in T} u(\theta, t)$ has interior solution for θ over a non-degenerate subinterval of $[0,1]$.

Lemma 9 (Regular Cell Functions)**.** *The cell function v in (3) is regular if F admits continuous and positive density and u is regular. The cell function v_m in (4) is regular, if u in (4) is continuous and strictly supermodular.*

Proof. Suppose that $\max_{t\in T} u(\theta, t)$ admits interior solutions for each θ in (b, c) with $b < c$. Then v in (3) is continuous and submodular with non-trivial convex support $[b, c]$. The existence of the directional derivative of $v'(0, z\pm)$ follows simply from the continuity of u and Theorem 3 in Milgrom and Segal [2002]. Theorem 1 in Milgrom and Segal [2002] implies the right size comparison.

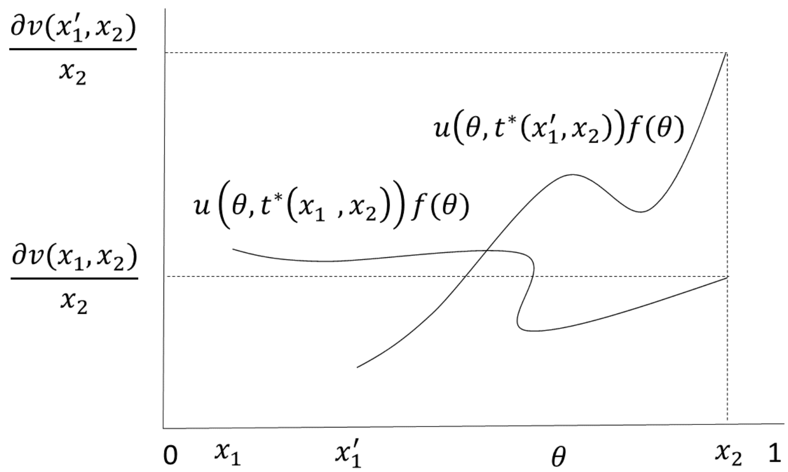

Figure 3: Illustration of (26) and (27) in the Proof of Lemma 9.

Next, for each cell $I = [x_1, x_2]$ with $b \leq x_1 < x_2 \leq c$, (3) also admits a unique interior solution, denoted by $t^*(x_1, x_2)$, which is continuous in (x_1, x_2) by the continuity of u_2. Since both $u(\theta, t)f(\theta)$ and $t^*(x_1, x_2)$ are continuous, the Envelope theorem applies, and so v over $X^2_{[b,c]}$ is partially differentiable with

$$\frac{\partial v(x_1, x_2)}{\partial x_2} = u(x_2, t^*(x_1, x_2))f(x_2).$$

Now I show that for each $x_1 < x_1' < x_2$ in $[b, c]$,

$$\frac{\partial v(x_1', x_2)}{\partial x_2} > \frac{\partial v(x_1, x_2)}{\partial x_2}, \tag{26}$$

by showing

$$u(x_2, t^*(x_1', x_2)) > u(x_2, t^*(x_1, x_2)). \tag{27}$$

Suppose the contrary: $u(x_2, t^*(x_1', x_2)) - u(x_2, t^*(x_1, x_2)) \leq 0$. First, we have $t^*(x_1', x_2) > t^*(x_1, x_2)$, since u_2 is SSC in θ. Then the function $u(\theta, t^*(x_1', x_2)) - u(\theta, t^*(x_1, x_2))$ is SC in θ, since u is SC ordered and $t^*(x_1', x_2) > t^*(x_1, x_2)$. Then since $u(x_2, t^*(x_1', x_2)) - u(x_2, t^*(x_1, x_2)) \leq 0$, we have $u(\theta, t^*(x_1', x_2)) - u(\theta, t^*(x_1, x_2)) \leq 0$ for each θ in $[0, x_2)$. This contradicts that $t^*(x_1', x_2)$ solves (3) for $I = [x_1', x_2]$. So (27) and thus (26) hold. The same also applies to $v_1(x_1, x_2)$.

The case of v_m follows easily from (14) and (15) in the proof of Lemma 1.

Q.E.D.

2 Comparative Statics for Cut-offs

2.1 Introduction

Milgrom and Shannon [1994] derived an important method of comparative statics, greatly extending the monotone methods introduced in Topkis [1978]. They imposed a strong single crossing property (SCP) on the objective function that guaranteed comparative statics valid for *every possible* lattice choice set. In an important extension, Quah and Strulovici [2009] (QS) relaxed the robustness condition, arguing that the choice sets in many cases are intervals. Their new condition, interval dominance Order (IDO), was thus weaker than the SCP.

In this note, I show that by taking path integral along the so-called *diagonally increasing* path, the scalar logic in QS's main sufficient conditions for IDO (Proposition 2 and 3) can be powerfully extended to cases like comparative statics for cut-offs.

2.2 QS's scalar result

Let X be a subset of the real line. All the intervals involved below are relative to X[16]. Following QS, for two real-valued functions f and g defined on X, we say that g dominates f in interval dominance order (denoted by $g \succeq_I f$) if $f(x'') - f(x') \geq (>)0 \Longrightarrow g(x'') - g(x') \geq (>)0$ for all $x'' > x'$ in X satisfying that $f(x'') \geq f(x)$ for all x in $[x', x'']$.

The Propositions 2 and 3 in QS are two important and easily verifiable sufficient conditions for interval dominance order (IDO) in scalar cases. I now provide a quick unified proof of them. Consider an integral inequality of Banks [1963]. Let f and ϕ be measurable functions over a measure space (X, μ). Assume ϕ bounded below, with f and $f\phi$ integrable. For $y \in \mathbb{R}$, define the *upper set* of ϕ by $U_\phi(y) = \{x \in X : \phi(x) \geq y\}$,[17] and put $L_\phi(y) = X - U_\phi(y)$.

Lemma 10 (Banks (1963)). *If* $\int_{U_\phi(y)} f d\mu \geq 0$, $\forall y \in \mathbb{R}$, *then* $\int_X f\phi d\mu \geq (\inf \phi) \int_X f d\mu$.

When $X \subseteq \mathbb{R}$ and ϕ is (weakly) increasing, the upper sets are simple: either $U_\phi(y) = [x, \infty)$ or $U_\phi(y) = (x, \infty)$ for some $x \in X$. But if $\int_{[x,\infty)} f d\mu \geq 0$ for all $x \in X$, then $\int_{(x,\infty)} f d\mu \geq 0$ for all

[16]For example, $[x', x''] = \{x \in X : x' \leq x \leq x''\}$, and $[x', \infty) = \{x \in X : x \geq x'\}$.
[17]All the results in this comment hold if we instead define $U_\phi(y) = \{x \in X : \phi(x) > y\}$.

$x \in X$, by the Dominated Convergence Theorem. Thus:

Corollary 2. *Assume $X \subseteq \mathbb{R}$ and ϕ increasing. If $\int_{[x,\infty)} f d\mu \geq 0$ for all $x \in X$, then $\int_X f\phi d\mu \geq (\inf \phi) \int_X f d\mu$.*

I now give a self-contained proof of Lemma 10, differing from Banks's proof for bounded ϕ, which appealed to a layer cake representation. Let I_A be the indicator function of A.

Proof. since any nonnegative bounded function lies in the closed convex cone generated (in the sup norm) by the indicator functions of the upper sets, we have $\int_X (\phi - \inf \phi) f d\mu \geq 0$, and thus $\int_X f\phi d\mu \geq (\inf \phi) \int_X f d\mu$. Next, since any unbounded ϕ is the limit of truncations $\phi I_{L_\phi(y)} + y I_{U_\phi(y)}$, we may apply the Dominated Convergence Theorem. *Q.E.D.*

Corollary 2 subsumes Lemmas 1 and 4 in QS as special cases, and thus affords a unified concise treatment of Proposition 2 and 3 in QS. Let (X, μ) be a positive measure space, where $X \subseteq \mathbb{R}$. Let $V_1(x) = c_1 + \int_{(-\infty,x)} u_1 d\mu$ and $V_2(x) = c_2 + \int_{(-\infty,x)} u_2 d\mu$, where u_1 and u_2 are integrable and c_1, $c_2 \in \mathbb{R}$.

Proposition 3 (Quah and Strulovici [2009])**.** *If $u_2 \geq u_1 \phi$ a.e.-μ for some strictly positive and increasing ϕ, then $V_2 \succeq_I V_1$.*

Proof. let $x'' > x'$ in X with $V_1(x'') \geq V_1(x)$ for all x in $[x', x'']$. I need to show that $V_2(x'') \geq V_2(x')$ and $V_2(x'') > V_2(x')$ if additionally $V_1(x'') > V_1(x')$.

$$V_2(x'') - V_2(x') = \int_{[x',x'')} u_2(z) d\mu \geq \int_{[x',x'')} \phi(z) u_1(z) d\mu. \tag{28}$$

For every x in $[x', x'')$, $\int_{[x,x'')} u_1(z) d\mu = V_1(x'') - V_1(x) \geq 0$. By Corollary 2, we have $\int_{[x',x'')} \phi(z) u_1(z) d\mu \geq \phi(x') \int_{[x',x'')} u_1(z) d\mu$, since ϕ is increasing. Thus by (28), $V_2(x'') - V_2(x') \geq \phi(x_1')(V_1(x'') - V_1(x'))$, since $\int_{[x',x'')} u_1(z) d\mu = V_1(x'') - V_1(x')$. This implies the desired result, given $\phi > 0$. *Q.E.D.*

2.3 Comparative statics for cut-offs

In this section, I extend the scalar logic in Proposition 3 to problems with multiple cut-offs. Let $(\mathbb{R}^n, \geq)$ be the n-dimensional Euclidean space, $\mathbf{1}$ the unit vector, and "$\vee$", "$\wedge$" the join and meet

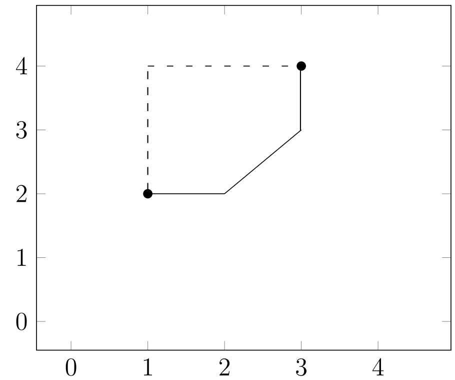

Figure 4: The DI path from $(1,2)$ to $(3,4)$ is the solid path; the dashed path is not DI. For an example in $\mathbb{R}^3$, the DI path from $(4,2,1)$ to $(5,3,6)$ is given by $\{(4,2,y) : y \in [1,2]\} \cup \{(4,y,y) : y \in [2,3]\} \cup \{(4,3,y) : y \in [3,4]\} \cup \{(y,3,y) : y \in [4,5]\} \cup \{(5,3,y) : y \in [5,6]\}$.

operators[18] . For each $x' > x$ in $\mathbb{R}^n$, the *diagonally increasing* (DI) path from x to x' is defined as the set $\{x \vee y\mathbf{1} \wedge x' : y \in \mathbb{R}\}$. Figure 1 illustrates the DI path from (1, 2) to (3, 4) in $\mathbb{R}^2$ by the solid line. A subset X of $\mathbb{R}^n$ is *diagonally monotone* (DM) connected if for every $x' > x$ in X, and $y \in \mathbb{R}$, we have $x \vee y\mathbf{1} \wedge x' \in X$. The following are three convenient classes of DM connected sets by definition: (a) $\{x \in I^n : x_1 \leq \cdots \leq x_n\}$ where I is an interval of $\mathbb{R}$, such as cut-off problems; (b) products of intervals; (c) budget sets. In the analysis below, I focus on cut-off problems with cut-offs. The results in this section applies more generally to DM connected sets with minor adaption.

2.3.1 Optimal Switching Time

I now give an intuitive interpretation of DM connectedness within the context of optimal switching time. Assume $n+1$ projects. An agent decides the optimal project switching timing, and must choose the projects in the precise order $1, 2, \ldots, n+1$. One is only allowed to choose one project at a time, and the decision to quit a project is irreversible. The choice set is given by $X = \{x \in [0,\infty)^n | x_1 \leq \cdots \leq x_n\}$, which rules that project k is implemented during the time interval $[x_{k-1}, x_k]$, if we interpret $x_0 = 0$ and $x_{n+1} = \infty$.

An important observation in this example is the *freedom of downwards switching.* That is, for each $x' > x$ in X, and each $t \in [0,\infty)$, it is always feasible for the agent to follow x' till time t, and deviate to

[18] For two points x and x' in $\mathbb{R}^n$, $x' \geq x$ if $x'_i \geq x_i$ for each i; $x' > x$ if $x' \geq x$ and $x' \neq x$. $x \wedge x' = (\min\{x_1, x'_1\}, \cdots, \min\{x_n, x'_n\})$, and $x \vee x' = (\max\{x_1, x'_1\}, \cdots, \max\{x_n, x'_n\})$

x since then[19]. If the agent does so, he actually uses the thresholds $x \vee t\mathbf{1} \wedge x'$. For example, let $n = 3$, $x = (1, 4, 7)$, and $x' = (2, 6, 8)$. Suppose that the agent plans to follow x' till $t = 5$ and to deviate to x since then. He will use project 1 over $[0, 2)$, project 2 over $(2, 5)$, project 3 over $(5, 7)$ and project 4 over $(7, \infty)$. This plan is equivalent to choosing the thresholds $(2, 5, 7) = (1, 4, 7) \vee 5\mathbf{1} \wedge (2, 6, 8)$. [20]

To summarize, in this example, for each $x' > x$ in X, and each $y \in \mathbb{R}^+$, the vector $x \vee y\mathbf{1} \wedge x'$ is the thresholds that one uses if he follows x' till $t = y$ and switches to x since then. The DI path from x to x' is the set of all the possible thresholds, due to the freedom of switching from x' to x at any time one wants. The property of DM connectedness of X is equivalent to the freedom of downwards switching for any $x' > x$ in X.

Following Quah and Strulovici [2007], for two real-valued functions f and g defined on X, we say that g dominates f in interval dominance order (denoted by $g \succeq_I f$) if $f(x'') - f(x') \geq (>) 0 \Longrightarrow g(x'') - g(x') \geq (>) 0$ for all $x'' > x'$ in X satisfying that $f(x'') \geq f(x)$ for all x in $[x', x'']$. Let $\geq_S$ be the common strong set order.

Proposition 4 (Optimal Switching time)**.** *If* $0 < \delta_1 < \delta_2$, *then*

$$\arg\max_{x \in X} V_{\delta_1}(x) \geq_S \arg\max_{x \in X} V_{\delta_2}(x).$$

Proof. I first show $V_{\delta_1} \succeq_I V_{\delta_2}$. Let $x'' > x'$ in X be such that $V_{\delta_2}(x'') \geq V_{\delta_2}(x)$ for all $x \in [x', x'']$. I need to show that $V_{\delta_1}(x'') \geq V_{\delta_1}(x')$ and $V_{\delta_1}(x'') > V_{\delta_1}(x')$ if additionally $V_{\delta_2}(x'') > V_{\delta_2}(x')$. Let $w'(t)$ and $w''(t)$ be the flow payoff associated respectively with x' and with x''. That is, $w'(t) = u^k(t)$, for $t \in [x'_{k-1}, x'_k]$, and $w''(t) = u^k(t)$, for $t \in [x''_{k-1}, x''_k]$. Then

$$V_{\delta_1}(x'') - V_{\delta_1}(x') = \int_0^\infty [w''(t) - w'(t)] e^{-\delta_2 t} e^{(\delta_2 - \delta_1)t} dt. \tag{29}$$

By previous discussion, for each $y \geq 0$, we have $x' \vee y\mathbf{1} \wedge x'' \in [x', x'']$ with $V_\delta(x' \vee y\mathbf{1} \wedge x'') =$

[19] I am very grateful of a referee for pointing out this observation to me. His comments motivate me to write this subsection.

[20] As comparison, "upwards switching" might not be feasible. That is, if one plans to follow x till some time t and switch to x' since then, he might revisit some already quited projects, which, however, is not allowed. For example, if one follows x till $t = 5$ and switch to x' since then. He will use project 1 over $[0, 1)$, project 2 over $(1, 4)$, project 3 over $(4, 5)$ and project 2 over $(5, 6)$, project 3 over $(6, 8)$ and project 4 over $(8, \infty)$. In this plan, the project 2 is quited first at $t = 4$, and then revisited at $t = 5$. Thus this specific upwards switching plan is not feasible.

$\int_0^y w''(t)e^{-\delta t}dt + \int_y^\infty w'(t)e^{-\delta t}dt$. Thus, for each $y \geq 0$,

$$\int_y^\infty [w''(t) - w'(t)]e^{-\delta_2 t}dt = V_{\delta_2}(x'') - V_{\delta_2}(x' \vee y\mathbf{1} \wedge x'') \geq 0. \tag{30}$$

Given $\delta_1 < \delta_2$, the function $e^{(\delta_2-\delta_1)t} \geq 1$ is strictly increasing. Thus by Corollary 2, (29) and (30) implies that $V_{\delta_1}(x'') - V_{\delta_1}(x') \geq \int_0^\infty [w''(t) - w'(t)]e^{-\delta_2 t}dt = V_{\delta_2}(x'') - V_{\delta_2}(x')$. It follows that $V_{\delta_1} \succeq_I V_{\delta_2}$.

Next, X is a sub-lattice of $\mathbb{R}^n$. Besides, each $\int_{x_{k-1}}^{x_k} u^k(t)e^{-\delta t}dt$ as a function of (x_{k-1}, x_k) is modular over $\{(x_{k-1}, x_k) : 0 \leq x_{k-1} \leq x_k\}$. V_δ is a summation of some modular functions, hence modular too. The desired result follows from the Theorem 2 in Quah and Strulovici [2007]. *Q.E.D.*

By Proposition 4, one will switch later for *all* projects when he is more patient. The key step, i.e., $V_{\delta_1} \succeq_I V_{\delta_2}$, cannot be derived by pairwise comparison using the scalar results in QS, i.e., their Proposition 2 or 5. To see the difficulty involved, let $n = 2$ and suppose that $(3, 4)$ is optimal for δ_1. We want to show that $(3, 4)$ is better than $(1, 2)$ for δ_2. By the scalar results in QS, $(3, 4)$ is better than $(1, 4)$ for δ_2. Hence we are done if we could further show that $(1, 4)$ is better than $(1, 2)$ for δ_2. But this could fail.

2.3.2 A Sufficient Condition for IDO

The interpretation of freedom of down-wards switching is lost when we go beyond problems like optimal switching time. However, DI paths still play a crucial role in understanding the monotone comparative statics for cut-offs.

For each $n \geq 1$, let $X^n = \{x \in [0, 1]^n : x_1 \leq \ldots \leq x_n\}$. Each x in X^n partitions the interval $[0, 1]$ into $n + 1$ (perhaps trivial) cells. To simplify notation, put $x_0 = 0$ and $x_{n+1} = 1$. Recall that in Tian(2015a), a cell function v is defined over X^2. A cell function v induces a *coarse value* for each finite interval division of $[0, 1]$ by summation, which we denote by V. Formally,

$$V(x) = \sum_{k=0}^{n} v([x_k, x_{k+1}])$$

for each $n \geq 0$ and x in X^n.

Let V_1 and V_2 be two coarse values respectively with v_1 and v_2 as their cell functions. I now give a sufficient conditions for V_1 and V_2 to be ordered in IDO over each X^n.

Proposition 5. *(IDO for Coarse Value). Suppose that for each $a \in I$, both $v_i(z, a)$ and $v_i(a, z)$ as functions of z are absolutely continuous with $v_2'(a, z) \geq \phi(z)v_1'(a, z)$ and $v_2'(z, a) \geq \phi(z)v_1'(z, a)$ for some common strictly positive and increasing function ϕ over $[0, 1]$. Then we have $V_2 \succeq_I V_1$ over X^n for each $n \geq 1$.*

The proof of Proposition 5 follows from taking path integral along the DI path, which I omit. Smith, Srensen, and Tian (2012) applied a version of Proposition 5 to get a comparative static result in belief cut-off rules. Tian (2015a) applied Proposition 5 to derive a general result about how the meaning of words changes with the prior belief in a model of efficient language, called monotone pragmatics.

3 Informational Herding, Optimal Experimentation, and Contrarianism

3.1 Introduction

In the informational herding model, an infinite sequence of individual agents must each choose an action from a common finite menu. Payoffs depend on an uncertain state of the world. Everyone has identical preferences, and a private signal. Individuals also observe the full history of prior actions. Bikhchandani et al. [1992] and Banerjee [1992] showed that a herd arises — eventually, all agents make the same choice, possibly unwise. Smith and Sørensen [2000] noted that a herd on an ex post suboptimal action can occur if and only if the private signals are of uniformly bounded precision.

In this paper, we undertake a general welfare analysis of this informational herding model, formally addressing the intuitive and oft-claimed notion that the equilibrium is inefficient. For observational learning involves an informational externality, since every action partially conveys a hidden private signal, and individuals do not account for the value of signaling information to successors. Shedding light on the informational herding externality, this paper completely characterizes the efficient forward-looking behaviour overseen by a social planner maximizing the discounted sum of expected utilities. We argue that this planner seeks to influence actions in a *contrarian* fashion, leaning against the popular actions, so that their lumpy actions will better reflect their private information.

The essence of the informational herding model is that private information cannot be directly communicated or shared; only observed actions can signal the private information. So we assume the planner sees realized actions but not private signals, and can dictate to each agent how to map hidden private signals into observable actions from the finite menu.

The planner's dynamic decision problem is a Markovian function of the public belief on the state of the world, derived from the action history. The Bellman principle aptly captures how the planner trades off current payoffs against the benefits to later agents of more informative observational learning. The planner's desire to signal information to posterity may entail drastic current payoff sacrifices. An agent might have to take myopically dominated actions, or take actions in a myopically

[20] This section is joint with Lones Smith and Peter Sorøsen.

suboptimal order (see §3.5).

In Proposition 6, we first show that the planner's optimal policy induces each action for an interval of private signals, or equivalently, of posterior beliefs. Proposition 7 devises a new Vickrey-Clarke-Groves mechanism to implement the planner's optimum. This simple mechanism internalizes the externality caused by him each individual, rewarding him solely on the basis of his and his successor's actions. We show how this one-stage look-ahead scheme works because the successor's action is informative about the true state. Since we deduce in Lemma 11 that value functions are convex, Corollary 6 shows that the socially optimal incentive scheme in fact rewards anyone who is mimicked, and punishes any who are not. This scheme is reminiscent of the way that academia rewards authors for citations.

In exploring the optimal rule, we first investigate how the discount factor influences social learning. The touchstone of the herding model is the *cascade set*, to which public beliefs almost surely converge. In this limit, actions cease to reflect private information, and social learning stops. When the precision of private signals is uniformly bounded, at least two non-empty cascade sets have a non-empty interior. Proposition 8 shows that while extreme cascade sets have a nonempty interior, they strictly shrink in the discount factor; near perfect patience, the interior cascade sets vanish altogether, while the extreme sets converge to points. So when myopic agents first herd, the planner wishes that they would not. But herding eventually occurs with the slightest level of impatience. In other words, *the herding phenomenon does not owe to the selfishness of agents, but instead to the difficulties of signaling private information through finitely many actions.*

We next explore in Proposition 9 how the planner's optimal interval rule for an agent's posterior belief responds to the public belief, for any fixed discount factor. The cutoff posterior belief separating two actions is constant for myopic agents. But in the planner's problem, this cutoff rises in the public belief, discouraging taking the high action when the public beliefs favor it. We call this *contrarian* since individuals skew choices away from the publicly more likely actions. Contrarianism thus makes the continuation belief less responsive to the change of the public beliefs. For a rough intuition why this is socially optimal, assume a higher public belief. Then posterior beliefs lie above a given threshold level with greater chance. But since a more likely action conveys a weaker signal about private signals, the social planner deters this action by raising the cutoff posterior belief.

Our paper makes generally useful methodological contributions for experimentation and social learning. We introduce and heavily exploit convex duality as an analytic tool for value functions in experimentation. It allows us to capture future values using subtangents to the convex value function — a method which should prove computationally useful too.

Towards a general theory of finite action experimentation, Proposition 6 generalizes Gittins' indices to our finite action experimentation setting, where the payoffs of "arms" are not independent: The planner always chooses the action with the highest *welfare index*. As an application, note that a convex value function need not be differentiable. But Corollary 2 proves it is differentiable where it counts: at optimal continuation public beliefs.

Bayesian updating in the social learning world is somewhat subtle. If the public belief rises, one might think that after seeing a fixed high action, the posterior public belief should rise too. But this intuition fails, since willingness to take the high action no longer offers as strong an endorsement of the high state; however, the proof of our contrarianism result depends on securing the intuitive monotonicity. For the planner must trade-off an inference from seeing an action against the current individual's payoff loss at the marginal belief. By assuming a log-concave density for the unconditional distribution of the private log-likelihood ratio, we secure in Lemma 3 a monotone relation between the prior and posterior public belief. In light of Smith and Sørensen [2000], *no cascade starts in finite time in the standard herding model given this robust signal property* (Corollary 4).

A final technical contribution is that our proof methodology exploits recent methods in monotone comparative statics introduced in Quah and Strulovici [2007]. This is important since there are generally no conditions that ensure differentiable convex value functions.

Related Literature. Banerjee's proposed remedy for the externality was to exclude early agents from seeing others' actions, rendering independent signals of their actions.[21] The rationale for this corrective supposes that the welfare loss of the early agents was dwarfed by the gains to successors. Our analysis builds on the assumption that the planner attaches a positive weight to each agent's utility. In fact, our planner has Banerjee's policy at his disposal, but it is suboptimal. This

[21] This idea has been further explored by Sgroi [2002]. Closer to our spirit, Doyle [2002] considers the social planner's problem in the endogenous-timing herding model of Chamley and Gale [1994].

owes to our contrarianism result, as the team does better when the decision rule responds strictly to changes in the action history.

Vives [1993] explores a social learning model with a fundamentally different sequential structure, and Gaussian information. There, a continuum of privately informed agents act in every period, and then observe a noisy market price statistic summarizing the actions. Reminiscent of the informational herding externality, the more accurate is the historical signal about the state, the less current actions reveal about private information. Addressing this externality, Vives [1997] studies a team problem in the market setting. He proves that team members choose to reveal more of their private information.[22] Our more elaborate contrarianism comparative statics result here finds that teams shy away *more* from the *more* popular actions.[23] Vives also finds that the optimal long-run Gaussian precision growth is as low as in the selfish model. This may seem analogous to our finding that cascade sets have a non-empty interior in the team setting, but there is no clear logical connection. Vives' experimentation problem, eg., never yields incomplete learning.

The planner's optimum is a *team equilibrium* (Radner [1962]), where everyone maximizes the sum of discounted expected utilities. In an equilibrium among these altruistic agents, successors cannot fully interpret a deviation by an agent who chooses an unanticipated map of private signals into actions. They can observe only the action, but not the underlying map. This fact simplifies the analysis of the planner's optimum — for she can likewise ignore any previous attempt to change the optimal map from signals to actions.

In Dow [1991], a consumer first observes a signal realization, but in the next period can only recall the signal's partition interval. In the second and final period, another signal realization is seen, and a choice is made. The consumer's optimal determination of the first-period coarse partitioning of the signal is like our planner's problem. Our planner's Markovian decision problem can be analysed in a two-period setting similar to Dow's.

The incorrect herding outcome is intuitively related to the familiar failure of complete learning in

[22] In a related setting, Medrano and Vives [2001] describe behaviour that reveals less private information as 'contrarianism.' We find it more natural that contrarian behaviour leans against the public belief.

[23] Vives always employs the normal learning model, ruling out results like ours on the distributional shape's importance. On the other hand, that model allows the long-run properties of learning to be characterized by the speed with which the precision approaches infinity. Our analysis offers no analogy.

optimal experimentation. Rothschild's (1974) analysis of the two-armed bandit is a classic example: An impatient monopolist optimally experiments with two possible prices each period, with fixed uncertain purchase chances for each price. Rothschild showed that the monopolist (*i*) eventually settles down on one price, and (*ii*) selects the less profitable price with positive probability. This is analogous to (*i*) an action herd occurs, and (*ii*) with positive chance is ex-post incorrect. Yet, Easley and Kiefer [1988] prove that complete learning generically arises with finite state and action spaces. This is puzzling, since the herding outcome arises in a model with finite actions and states.

The formulation of our social planner's problem offers a resolution of this puzzle. Even though each agent chooses from a finite action set, our social planner has no access to private signals, and so cannot dictate the choice among any two actions. Rather, for each history, he chooses a continuously defined rule that maps agents' private beliefs into actions. In the myopic planner case with a zero discount factor, we obtain the original herding model. Hence, we can conclude that the herding outcome is formally equivalent to incomplete learning in an experimentation model with a continuous choice space.

The paper is organized as follows. We introduce the constrained efficient herding model in §3.2, and our convex duality formulation in §3.3. Section 3.4 characterizes optimal behavior, thereby introducing our novel welfare indexes; §3.5 offers some counterintuitive examples of optimal strategies that we must avoid. We show how cascade sets shrink as patience rises in §3.6, and finally motivate and explore contrarianism in §3.7. Many proofs are appendicized.

3.2 The Forward-Looking Herding Model

An infinite sequence of decision-makers (agents) $n = 1, 2, \ldots$ acts in that exogenous order, and share a common 50-50 prior belief over two states of the world $\omega \in \{L, H\}$.

The nth agent sees a random private signal σ_n about ω. The resulting posterior belief in state H is called his *private belief.* In state ω, the signals are i.i.d. across agents, with cdf F^ω. Assume that F^H and F^L are mutually absolutely continuous, with derivative $dF^H/dF^L = \sigma/(1-\sigma)$ and so common support $\text{supp}(F)$. Signals therefore never perfectly reveal the state: Accordingly, $F^H(\sigma) \leq F^L(\sigma)$, with inequality strictly inside $\text{supp}(F)$.

Abusing notation, agents choose from a finite action set $A \equiv \{1, \ldots, A\}$. Action a yields common payoff $u(a, \omega)$ in each state $\omega \in \{H, L\}$ for all agents. Action 1 is best in state L, and action A in state H. No two action payoffs are tied in either state, and payoffs obey increasing differences: $u(1, H) - u(1, L) < u(2, H) - u(2, L) < \cdots < u(A, H) - u(A, L)$. Still, an action might be dominated for all beliefs over $\{L, H\}$ if its payoff is low enough.

Before choosing, the n'th agent observes σ_n and the history of the $n-1$ predecessors' actions. He can compute the probability distribution over histories, based on correctly conjectured predecessors' strategies, and arrive at the *public belief* $\pi \in [0, 1]$ in state H. Conditioning next on the conditional independent private signal σ gives the *posterior belief* $\rho \in [0, 1]$:

$$\rho = r(\pi, \sigma) \equiv \frac{\pi\sigma}{\pi\sigma + (1 - \pi)(1 - \sigma)}. \tag{31}$$

The chosen action a maximizes the agent's *expected payoff* $\bar{u}(a, \rho) = (1-\rho)u(a, L) + \rho u(a, H)$. At public belief π, the private belief σ is distributed $F^\pi = \pi F^H + (1 - \pi)F^L$.

This paper modifies the herding model, and considers the game where everyone instead altruistically aims to maximize a measure of welfare. Specifically, we assume that an informationally constrained social planner observes the action history, but not the private signals. A *rule* ξ maps each agent's private beliefs σ to his action $a \in A$. Let Ξ be the set of all such rules. A *strategy* s_n for the player in period n is a map that assigns a rule to each action history, and $s = (s_1, s_2, \ldots)$ denotes a strategy profile. The social planner chooses s to maximize the expected average present value of utility stream u_n, obtaining value

$$v_\delta(\pi) = \sup_s E[(1 - \delta) \sum_{n=1}^{\infty} \delta^{n-1} u_n] \tag{32}$$

This informational herding model assumes $\delta = 0$. We study the altruistic case $\delta \in (0, 1)$.

3.3 Dynamic Programming and Convex Duality

Adapting Radner [1962], we call a perfect Bayesian equilibrium of this game a *team equilibrium*. Yet still there are many team equilibria, and in this paper we focus on the efficient team equilibrium, since *a social optimum is a team equilibrium for any discount factor* $\delta < 1$. To see why this is so, suppose that all but one agent uses a sequentially rational optimal strategy s, but that some agent n has a strictly better reply $\tilde{\xi}$ at a history. Then the planner can improve his value at that history by *fully* mimicking this deviation, i.e. by (i) using rule $\tilde{\xi}$ in the first period and (ii) continuing with s as if s_n had been applied at stage n with this history (as the team would not have detected the deviation). This one-shot deviation principle version contradicts optimality of the policy.

As is well-known, the socially optimal policy can be implemented as a dynamic optimization in a Markovian fashion, in which the public belief π is the state variable. A *Markov policy* Υ maps from public beliefs $\pi \in [0,1]$ into Ξ, so that $\Upsilon(\pi)$ is the rule taken at belief π. Given rule ξ, action a is taken with chance $\psi(a,\omega,\xi) = \int_{\xi^{-1}(a)} dF^{\omega}$ in state ω, and unconditionally with chance $\psi(a,\pi,\xi) = \int_{\xi^{-1}(a)} dF^{\pi}$. After seeing an agent's action a, the *posterior public belief* is $p(a,\pi,\xi)$. If $\psi(a,\pi,\xi) > 0$, then Bayes updating yields $p(a,\pi,\xi) = \pi\psi(a,H,\xi)/\psi(a,\pi,\xi)$, and the martingale property of beliefs implies $p(a,\pi,\xi) = \int_{\xi^{-1}(a)} r(\pi,\sigma)dF^{\pi}$. When $\psi(a,\pi,\xi) = 0$, Bayes-optimal behaviour does not pin down the off-path value of $p(a,\pi,\xi)$. All behaviour that we study in the rest of the paper is supported by any reasonable off-path belief: For instance, if $\psi(a,\pi,\xi) = 0$ then there exists $\sigma \in \text{supp}(F)$ such that $p(a,\pi,\xi) = r(\pi,\sigma)$.

By dynamic programming, the planner's value function solves the Bellman equation:

$$v(\pi) = \sup_{\xi\in\Xi}(T_{\xi}v)(\pi), \tag{33}$$

where the policy operator T_{ξ} in (33) is defined for any continuation value v:

$$(T_{\xi}v)(\pi) = \sum_{a=1}^{A} \psi(a,\pi,\xi)[(1-\delta)\bar{u}(a,p(a,\pi,\xi)) + \delta v(p(a,\pi,\xi))]. \tag{34}$$

Lemma 11 (Convexity and Monotonicity)**.** *The value function* $v_{\delta}(\pi)$ *is a bounded convex and continuous function of public beliefs, and weakly increases in the discount factor* δ*. All subtangents to* v_{δ} *have slopes bounded between* $u(1,H) - u(1,L)$ *and* $u(A,H) - u(A,L)$*.*

It follows that the value function v is the upper envelope of its supporting tangent lines, as described by the compact *tangent space to* v, denoted $\mathcal{T}_v \subset \mathbb{R}^2$, where tangents are parameterized by their slope and intercepts. Since $\bar{u}$ and τ_a are affine functions, and since $p(a,\pi,\xi) = \int_{\xi^{-1}(a)} r(\pi,\sigma)dF^\pi$, we can rewrite operator (34) as

$$(T_\xi v)(\pi) = \max_{(\tau_1,\ldots,\tau_A)\in\mathcal{T}_v^A} \sum_{a=1}^{A} \int_{\xi^{-1}(a)} [(1-\delta)\bar{u}(a,r(\pi,\sigma)) + \delta\tau_a(r(\pi,\sigma))]dF^\pi. \tag{35}$$

Exchange the order of maximization in (33) to obtain the dual problem:

$$v(\pi) = \max_{(\tau_1,\ldots,\tau_A)\in\mathcal{T}_v^A} \sup_{\xi\in\Xi} \sum_{a=1}^{A} \int_{\xi^{-1}(a)} [(1-\delta)\bar{u}(a,r(\pi,\sigma)) + \delta\tau_a(r(\pi,\sigma))]dF^\pi. \tag{36}$$

As an aside, convex duality offers a computational strategy for solving the dynamic programming problem. In the iterative process, given a value v_n, the next value v_{n+1} is obtained in principle by searching across all the possible rules. But the convex duality suggests an alternative faster way to compute v_{n+1}: The required tangent space is simply the set of all the left and right derivative lines to v_n.

In the multi-armed bandit (Bertsekas [1987], §6.5), an experimenter each period chooses one of A actions, with uncertain independent reward distributions. Gittins [1979] solved for optimal behaviour via index rules: Attach to each action the value of the problem with just that action and the largest fixed retirement reward yielding indifference. Then choose the action with the highest index.

We now argue that the policy in a stationary team equilibrium admits an analogous index rule: At public belief π and posterior ρ, an agent chooses the action a with the largest *welfare index* w — equal to the social payoff as privately gauged by the agent. The index expression (37) follows from (36) in the social optimum, but our proof derives it directly for any stationary team equilibrium.[24]

Proposition 6 (Optimal Behaviour via Index Rules)**.** *Fix a stationary team equilibrium* Υ*. For every public belief* π*, there exist affine functions* τ_a*, so that the altruistic agent with private posterior belief* ρ *who takes action* a *has expected average present value*

$$w(a,\pi,\rho) = (1-\delta)\bar{u}(a,\rho) + \delta\tau_a(\rho). \tag{37}$$

[24] Since by convex duality, any convex function is the upper envelope of all supporting tangents, we can extend our theory to almost any convex value function.

If Υ is a social optimum, then τ_a is a subtangent to v at continuation belief $p(a, \pi, \Upsilon(\pi))$.

Proof. Action a results in some state-contingent expected continuation value $\tau_a(\omega)$ in state ω, i.e. the discounted sum of future utilities. The agent's uncontingent continuation value is given by $\tau_a(\rho) \equiv \rho\tau_a(H) + (1-\rho)\tau_a(L)$. This gives the present value expression (37). In the social optimum, the continuation value is $\tau_a(p(a)) = v(p(a))$, where $p(a) = p(a, \pi, \Upsilon(\pi))$. Because the planner can always use the same subgame strategy starting at an arbitrary public belief p as is optimal at $p(a)$, we have $\tau_a(p) \le v(p)$. So the affine function τ_a is necessarily subtangent to v at $p(a)$. □

This result reduces the infinite horizon optimization to a standard decision problem with a piecewise linear payoff function (37). It adds to (36) that the planner chooses τ_a to be tangent to v at continuation belief $p(a, \pi, \xi)$. The function $\tau_a(\rho)$ plays a central role for us, and admits an economic interpretation — it is the expected value for an agent with private posterior belief ρ of the subgame starting at public posterior belief $p(a, \pi, \xi)$.

3.4 Cascade Sets and Implementation

A. Interval Rules and Cascade Sets. Call action $a \in A$ *active* at a public belief if it is taken with positive chance. The public belief π lies inside the *cascade set $C_a(\delta)$ for action a* when a is optimal for all private beliefs σ. The certain choice of action a is optimal iff $v_\delta(\pi) = \bar{u}(a, \pi)$. The union $C(\delta) = \cup_{a \in A} C_a(\delta)$ is the *cascade set.* Obviously, in a cascade on action $a \in A$, that action is active. Private beliefs are *unbounded* if supp(F) contains 0 and 1, and *bounded* if the support of private beliefs obeys supp$(F) \subseteq (0,1)$ Smith and Sørensen [2000] show that cascade sets are nonempty iff private beliefs are bounded. This result also obtains here. Part (d) below strengthens Lemma 11:

Lemma 1. (a) *The cascade set $C_a(\delta)$ for any action a is empty, a point, or a closed interval, with $0 \in C_1(\delta)$ and $1 \in C_A(\delta)$ for any $\delta \in [0,1)$, and $C(\delta) \neq [0,1]$.*

(b) *With bounded private beliefs, $C_1(\delta) = [0, \underline{\pi}(\delta)]$, $C_A(\delta) = [\bar{\pi}(\delta), 1]$, for $0 < \underline{\pi}(\delta) < \bar{\pi}(\delta) < 1$.*

(c) *With unbounded private beliefs, only the cascade sets $C_1(\delta)$ and $C_A(\delta)$ are nonempty.*

(d) *Cascade sets $C(\delta)$ weakly shrink in the discount factor δ.*

In the standard herding model, Smith and Sørensen [2000] show that limit beliefs belong to cascade sets. We verify this still holds here in Appendix §3.11.2-B.

A *belief interval rule* is described by a partition of supp(F) into possibly trivial or empty intervals $\mathcal{J}(\pi) = \{J_a(\pi)\}$ of private beliefs for each public belief π, with action a optimal (namely, $\Upsilon(\pi)(a) = 1$) iff $\sigma \in J_a(\pi)$. An equivalent interval rule partitions $[0, 1]$ into possibly empty intervals of posterior beliefs $\mathcal{I} = \{I_a\}$, with action a optimal iff $\rho \in I_a$. The indices coincide at the boundary between neighboring intervals: $w(a, \pi, \rho) = w(\tilde{a}, \pi, \rho)$ on the boundary ρ between I_a and $I_{\tilde{a}}$. These boundaries are called *thresholds*.

Corollary 1. *For each public belief π, an optimal interval policy $\mathcal{I}$, and thus $\mathcal{J}(\cdot)$, exists.*

Proof. By Proposition 6, the index value of each action is affine in the posterior belief. Since the agent chooses actions with the maximal index value, an optimal interval policy $\mathcal{I}$ exists for posteriors. Since the posterior belief $\rho = r(\pi, \sigma)$ rises in σ, any interval of posterior beliefs σ maps monotonely into a private belief interval. So $\mathcal{J}(\cdot)$ exists. □

By the convex duality logic in (36), an interval rule applies for all convex continuation value functions whose subtangents have bounded slope.

A convex value function is differentiable a.e. by convexity, but need not be everywhere differentiable; this immeasurably complicates our analysis. But our welfare index formulation implies that the value function must be smooth at any optimal continuation:

Corollary 2 (Differentiable Continuations)**.** *Given a convex private belief support* supp(F)*, the value function is differentiable at all active public beliefs $p(a, \pi, \xi)$, for $\pi \notin C(\delta)$.*

B. Implementation of the Social Optimum. The team problem assumes that every agent altruistically cares about posterity. If the agents are really selfish, can the planner implement the optimal solution using a transfer scheme? Since the planner cannot observe the private signals, transfers may only depend on the observed action history. Otherwise, in light of (37), the socially optimal behavior could be decentralized by awarding individuals transfers equal to $\delta\tau_a(\rho)/(1-\delta)$, depending on the posterior belief ρ.

When the planner's policy prescribes an interval rule that does not swap the myopic interval order, it suffices to reward the agent on the basis of his own actions. The planner can move the selfish agent's threshold between each pair of actions up (or down) by taxing (or subsidizing) the action taken above the threshold. But transfers based on the agent's own action can never reverse the myopic ordering of actions, and thus are not sufficient if the selfish optimal action ordering differs from the socially optimal action ordering.[25]

Let $M(a, \pi, \omega)$ be the maximal welfare later individuals can get with the observation of a current action a, minus the best they can get without it in state ω. A *pivot mechanism* rewards agents for their marginal contribution to social welfare, paying them $M(a, \pi, \omega)$.

Proposition 7. *The social optimum can be implemented by a mechanism whose transfers only depend on the public belief, an agent's action, and his successor's action. A unique such mechanism in this class is a pivot mechanism, for non-cascade continuation beliefs.*

Proof: Let $\tau(\pi, \omega)$ denote the ω-contingent continuation value of the subgame starting at public belief π. (From Proposition 6, there is a tangent τ to v at π such that $\tau(\pi, L) = \tau(0)$ and $\tau(\pi, H) = \tau(1)$.) The additional present value of current action a to later agents equals

$$M(a, \pi, \omega) = [\delta/(1-\delta)](\tau(p(a, \pi, \xi), \omega) - \tau(\pi, \omega)). \tag{38}$$

For cascade beliefs $\pi \in C(\delta)$, no transfer is needed since $C(\delta) \subseteq C(0)$.

We thus continue under assumption $\pi \notin C(\delta)$. Suppose first that no active action reaches the cascade set. Consider active action a. The successor takes some action b for the lowest private beliefs, with chance denoted $\psi(b, \omega)$. Give the agent transfer $t(a, b)$ when the successor chooses b, and $t(a, \neg b)$ otherwise. The expected transfer for a is $\psi(b, \omega)t(a, b) + (1 - \psi(b, \omega))t(a, \neg b)$. Lower beliefs are more likely in the low state: $\psi(b, H) < \psi(b, L)$; therefore, there exist a unique pair $t(a, b), t(a, \neg b)$ solving the two equations (for $\omega = H, L$):

$$\psi(a, b, \omega)t(a, b) + (1 - \psi(a, b, \omega))t(a, \neg b) = M(a, \pi, \omega). \tag{39}$$

[25] Bru and Vives [2002] consider IC mechanisms that cannot implement the optimum of Vives [1997].

Simple algebra confirms that $\bar{u}(a,\rho)+\rho M(a,\pi,H)+(1-\rho)M(a,\pi,L)$ is an affine transformation of the index $w(a,\pi,\rho)$, where the transformation only depends on π and δ. The transfer thus provides the right incentive and implements social optimum, and constitutes a pivot mechanism,[26] since each agent is paid the marginal contribution.

We deter agents from taking inactive actions with large negative transfers.

Suppose next that the continuation belief after action a lands in a cascade set. Then $\psi(b,H)=\psi(b,L)$, so system (39) might not be solvable. But by Claim 4 in §3.11.4, at most one cascade set, say $C_{a'}(\delta)$, is reached across all actions. We can in this case construct a valid non-pivot mechanism as follows. Let the transfer for all active actions leading to $C_{a'}(\delta)$ be 0. For all other active actions a, give the agent n a modified state contingent marginal contribution: $M'(a,\pi,\omega)=[\delta/(1-\delta)](\tau(p(a,\pi,\xi),\omega)-u(a',\omega))$. □

3.5 Communication Via Action Choices

We make two observations about the planner's actions choices, in the spirit of **?**.[27]

Lesson 1: Dominated actions may be socially valuable. Informational herding describes learning filtered through a finite mesh action screen. If agents could more precisely convey their private information with more actions, then welfare intuitively rises. This effect can be so strong that using myopically dominated actions might be efficient.

For an example, assume bounded beliefs and altruism ($\delta>0$). Suppose that action A dominates $A-1$, with $u(A,\omega)=u(A-1,\omega)+\varepsilon$. We will show that for small enough $\varepsilon>0$, action $A-1$ is optimally taken with positive chance for some public beliefs.

To see this, suppose that the planner never uses action $A-1$. Then the value function v is affine on the cascade set $C_A=[\bar{\pi},1]$, but not on any extension of C_A to the left, and so is strictly convex at $\bar{\pi}$. At belief $\bar{\pi}$, action A is optimal for all private beliefs. The alternative rule x that maps all private beliefs below 1/2 into $A-1$, and others into A, induces $p(A-1,\bar{\pi},\xi)<\bar{\pi}<p(A,\bar{\pi},\xi)$. Since the

[26]While there exist simpler mechanisms, a pivot mechanism is quite intuitive and has an interesting implication we pursue later. **?** focus on flow marginal contribution. But in our case, each agent enters just once.

[27]We generalize **?**'s **?** Proposition 2, which assumes perfect patience and a simple second-period value function. His Example 3 shows that a multiplicity of optimal solutions can arise in these problems.

value function is strictly convex at $\bar{\pi}$, the expected continuation value exceeds $v(\bar{\pi})$ by some $\eta > 0$. This policy change produces a myopic loss less than ε, and therefore improves on the optimal policy when $\delta\eta > (1-\delta)\varepsilon$.

Lesson 2: Actions might optimally be taken in an "unnatural" order. Specifically, the *natural order* requires that if actions $a' > a$ are both active, then interval $I_{a'}(\delta)$ lies above $I_a(\delta)$. We argue that this holds if the discount factor is low enough — for intuitively, the dynamic optimization is then well-approximated by the myopic one. Define the payoff slope differences $\Delta_a \equiv (u(a,H)-u(a,L)) - (u(a-1,H)-u(a-1,L))$, for actions $a = 2,\ldots,A$. By our action ordering, $\Delta_a > 0$ for all a. Define the sum $\Delta \equiv (u(A,H)-u(A,L)) - (u(1,H)-u(1,L))$ and minimum $\underline{\Delta} = \min_{\{2,\ldots,A\}} \Delta_a$.

Corollary 3. *If $\delta < \frac{\underline{\Delta}}{\Delta + \underline{\Delta}}$, then for any public belief π not in a cascade set, the optimal interval policy is consistent with the natural order. With two actions, this holds for $\delta < 0.5$.*

Proof: By Proposition 6, it is optimal to choose the action with highest welfare index $w(a,\pi,\rho)$. Since $w(a,\pi,\rho)$ is linear in ρ, it suffices that $\frac{\partial}{\partial\rho}w(a,\pi,\rho)$ strictly increase in a. Given v_δ convex, the slope of any subtangent line τ of v_δ is sandwiched as follows:

$$u(1,H) - u(1,L) \leq v'(0+) \leq \frac{\partial\tau}{\partial\rho} \leq v'(1-) \leq u(A,H) - u(A,L)$$

This inequality allows us to bound the difference of welfare indices (37) from below:

$$\frac{\partial w(a+1,\pi,\rho)}{\partial\rho} - \frac{\partial w(a,\pi,\rho)}{\partial\rho} \geq (1-\delta)\Delta_{a+1} - \delta\Delta$$

This is strictly positive when $\delta < \Delta_{a+1}/(\Delta + \Delta_{a+1})$. Finally, $\underline{\Delta} = \Delta_2$ for $A = 2$. □

The premise of Corollary 3 is needed. In an example in the Supplemental Appendix, an unnatural action order is in fact optimal with a high enough discount factor.

3.6 Shrinking Cascade sets Via Patience

We prove in §3.11.2 that, as in Smith and Sørensen [2000], beliefs converge almost surely to the cascade sets. Similarly, learning is incomplete here, and so incorrect herds arise with positive probability iff private beliefs are bounded. We now argue that, with bounded beliefs, cascades shrink in the discount factor, vanishing in the patience limit. The chance of incorrect herds falls; in the limit of a very patient social planner, all herds are correct.

Proposition 8. *Assume bounded beliefs.*

(*a*) *(Cascades) Non-empty cascade sets strictly shrink when $\delta < 1$ rises: For all actions a, if $\delta_2 > \delta_1$ and $C_a(\delta_1) \neq \emptyset$, then $C_a(\delta_2) \subset C_a(\delta_1)$. For large enough $\delta < 1$, all cascade sets disappear except for $C_1(\delta)$ and $C_A(\delta)$, while $\lim_{\delta\to 1} C_1(\delta) = \{0\}$ and $\lim_{\delta\to 1} C_A(\delta) = \{1\}$.*

(*b*) *(Herds) A herd almost surely starts, and the chance it is incorrect vanishes as $\delta \uparrow 1$.*

This result formalizes the gut feeling of many economists on the inefficiency of cascades. The proof exploits the planner's indifference about actively experimenting at the edge of a cascade set. As a result, he strictly prefers to do so if he is slightly more patient, and as a result, the cascade set shrinks. Intuitively, when the planner is actively experimenting, the planner enjoys a higher value of information (expected gain in his continuation value) when he is more patient. The proof requires that we strengthen Lemma 11.

Lemma 2 (Strict Value Monotonicity)**.** *The value function increases strictly in $\delta < 1$ outside the cascade sets: If $\delta_2 > \delta_1$, then $v_{\delta_2}(\pi) > v_{\delta_1}(\pi)$ for all public beliefs $\pi \notin C(\delta_2)$.*

Since cascade sets change with the discount factor, we can now finish the discussion in §3.2, and explain why an equilibrium need not be a social optimum. For there exists a suboptimal team equilibrium when $\delta > 0$ in which everyone acts myopically. For herding on action a is a team equilibrium for public beliefs $\pi \in C_a(0) \setminus C_a(\delta)$. If every successor considers agent n's action uninformative, since π is in a cascade set for $\delta = 0$, the best that agent n can do is to maximize his current payoff. Since $\pi \notin C_a(\delta)$, this rule is suboptimal.

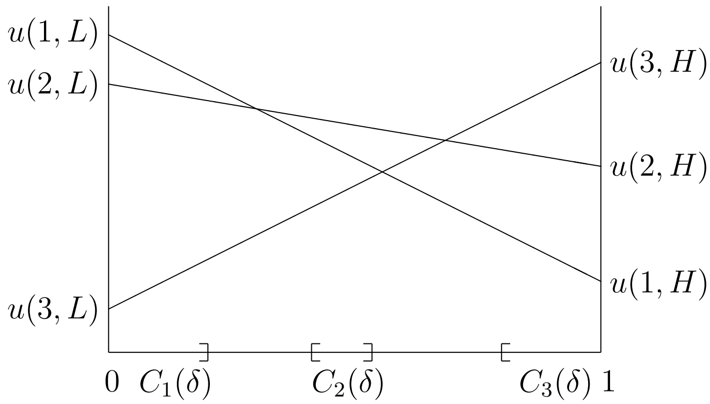

Figure 5: **Static Payoffs, Bellman Values, and Cascade Sets.** By Proposition 8, each cascade set $C_\delta(\pi)$ shrinks as the discount factor δ rises.

3.7 Contrarianism

3.8 An Illustrative Example: The Professor and his Student

We first consider a fully-solvable two period example that captures the essence of the short-run optimality of contrarian behavior in a stylized setting. Assume two actions 1 and 2 with payoffs $u(2,H) = u(1,L) = 1, u(1,H) = u(2,L) = -1$.

A professor and a student share a common prior π on state H, and observe conditionally iid signals σ with state-dependent cdf's $F^H(\sigma) = \sigma^2$ and $F^L(\sigma) = 2\sigma - \sigma^2$. The professor sees the signal, and takes an action; his student observes his action, but not his signal. Subject to this restriction, the professor selflessly acts to maximize his student's expected payoff. So the problem is formally one of pure information transmission, as in Dow [1991].

If the student starts with a continuation public belief p, then she takes action 2 exactly when her signal $\sigma \geq 1 - p$. Now, $\sigma \geq 1 - p$ with chance $1 - F^H(p) = 1 - (1-p)^2$ in state H and with chance $F^L(1-p) = p^2$ in state L. Hence, the student's value function is

$$V_S(p) = p(1 - 2(1-p)^2) + (1-p)(1 - 2p^2) = 1 - 2p + 2p^2$$

The professor clearly employs a private belief threshold rule $\bar{\sigma} = \bar{\sigma}(\pi)$: He either chooses action 1 for discouraging signals $\sigma < \bar{\sigma}$, and action 2 if $\sigma \geq \bar{\sigma}$, or the exact opposite. For recalling the message of §3.5, the discount factor is one, and either action ordering is possible. We assume the natural ordering for simplicity. He seeks to maximize $V(\pi) = E[V_S(P)|\pi]$, where P is his student's realized public belief, and the expectation is taken *ex ante*. Since $\pi = E[P|\pi]$ by the martingale property of beliefs, we have

$$V(\pi) = E[V_S(P)|\pi] = E(1 - 2P + 2P^2|\pi) = 1 - 2\pi + 2\pi^2 + 2E[(P-\pi)^2|\pi]$$

Then the professor's optimal value $V(\pi)$ exceeds the student value $V_S(\pi) = 1 - 2\pi + 2\pi^2$ by twice the variance of beliefs. We now compute this term. Given the threshold rule, a different continuation public belief $P = p_1$ or $P = p_2$ arises after each of the two professorial actions 2 and 1. Bayes rule reveals the formulas $p_1(\bar{\sigma}) = [\pi\bar{\sigma}^2]/[\pi\bar{\sigma}^2 + (1-\pi)(2\bar{\sigma} - \bar{\sigma}^2)]$ and $p_2(\bar{\sigma}) = [\pi(1-\bar{\sigma}^2)]/[\pi(1-\bar{\sigma}^2) + (1-\pi)(1 - 2\bar{\sigma} + \bar{\sigma}^2)]$. We can explicitly compute:

$$E[(P-\pi)^2|\pi] = \frac{\pi - p_1}{p_2 - p_1}(p_2 - \pi)^2 + \frac{p_2 - \pi}{p_2 - p_1}(\pi - p_1)^2 = (p_2 - \pi)(\pi - p_1) \tag{40}$$

Only this term in $V(\pi)$ depends on $\bar{\sigma}$. Maximizing (40) over $\bar{\sigma}$ yields *private signal threshold* $\bar{\sigma}(\pi) = (\pi - 1 + \sqrt{\pi - \pi^2})/(2\pi - 1)$ if $\pi \neq 1/2$, with limit $\bar{\sigma}(1/2) = 1/2$ by l'Hopital's rule.

The professor's Bayesian *posterior belief threshold* is $\theta(\pi) = [\pi - \sqrt{\pi - \pi^2}]/[2\pi - 1]$. Illustrating our later short-run contrarianism result, $\theta(\pi)$ is increasing in π. So the professor optimally communicates the state of the world by acting in a "contrarian" fashion. To wit, he leans against the public belief, so that the professor chooses action 2 less often with more public confidence in state H. It is easy to verify that the posterior threshold $\theta(\pi)$ exceeds the myopic posterior belief threshold $\theta(\pi) = 1 - \pi$ exactly when $\pi > 1/2$.

3.9 Monotone Posterior Beliefs: Cascades Cannot Start Late

We now provide a robust condition on the private signal distribution yielding *posterior monotonicity*: holding fixed a rule's interval policy $\mathcal{I}$, the continuation public belief $p(a, \pi, \mathcal{I})$ strictly rises in the current public belief π, for all active actions a.

Given the equi-likely states, the *unconditional private belief distribution* is described by the function $F = (F^H + F^L)/2$. When the density $f = F'$ exists, Bayesian updating implies a simple "no introspection condition" [Smith and Sørensen, 2000] for densities: $f^H(\sigma) = 2\sigma f(\sigma)$ and $f^L(\sigma) = 2(1-\sigma)f(\sigma)$. Associate to private signal σ the log-likelihood ratio $\ell = \Lambda(\sigma) \equiv \log(\sigma/(1-\sigma))$, with inverse $\mathcal{S}(\ell) = e^\ell/(1+e^\ell)$. *In the rest of the paper, we maintain the following (novel) regularity condition:*

(LC): The log-likelihood ratio density $\phi(\ell) \equiv f(\mathcal{S}(\ell))\mathcal{S}'(\ell)$ exists, and is log-concave.

Assumption (LC) is violated by atomic distributions, but common continuous distributions are log-concave (see Marshall and Olkin [1979], §18.*B*.2.*d*), including that in §3.8.

The posterior ρ in (31) depends on the public π and private signal σ. Observe that $\Lambda(\rho) = \Lambda(\pi) + \Lambda(\sigma)$. We denote the density of the posterior belief by $g(\rho|\pi)$.

Lemma 3 (Posterior Monotonicity)**.** *The posterior belief density $g(\rho|\pi)$ is strictly log-supermodular, given (LC). Posterior monotonicity obtains for any active action a.*

Proof. Let $\phi^\omega(\ell)$ be the density over the private log likelihood ratio ℓ in state ω. Observe that $\phi(\ell) = (\phi^L(\ell) + \phi^H(\ell))/2 = (1+e^\ell)\phi^L(\ell)/2$, since the no introspection condition implies $\phi^H(\ell) = e^\ell \phi^L(\ell)$. Hence, $\log \phi^L(\ell) = \log \phi(\ell) - \log(1+e^\ell) + \log 2$ is strictly concave, since ϕ is log-concave by condition (LC). As a result, the unconditional density $h(\ell|\pi)$ over the posterior log likelihood ratios ℓ given prior belief π is strictly log-supermodular, since:

$$h(\ell|\pi) = (1-\pi)\phi^L(\ell - \Lambda(\pi)) + \pi\phi^H(\ell - \Lambda(\pi)) = (1-\pi)(1+e^\ell)\phi^L(\ell - \Lambda(\pi))$$

Then $g(\rho|\pi)$ is strictly log-supermodular since the map $\ell \mapsto \rho$ strictly increases. Finally, a is taken when $\rho \in I_a$. Continuation public belief $p(a,\pi,\mathcal{I}) = \int_{I_a} \rho g(\rho|\pi) d\rho / \int_{I_a} g(\rho|\pi) d\rho$ strictly rises in π — the denominator is positive since a is active. □

In general, the optimal interval policy $\mathcal{I}$ depends on π, and we do not claim that posterior belief monotonicity always holds in equilibrium. To see the role played by Assumption (LC) in Lemma 3, we now offer an example in which (LC) fails and posterior monotonicity fails too. We slightly modify our signal family example in §3.8, punching a hole in its support. Choose $b \in (1/2, 1)$, and define the density $f(\sigma) = 1/(2-2b)$, for $\sigma \le 1-b$ and $\sigma \ge b$, and $f(\sigma) = 0$ otherwise. Let $f^H(\sigma) = 2\sigma f(\sigma)$

and $f^L(\sigma) = 2(1-\sigma)f(\sigma)$. Assume that "Buy" is optimal if the posterior belief exceeds 1/2. Given a public prior belief $\pi > b$, the posterior likelihood ratio after seeing "Buy" is

$$LR(\pi) \equiv \frac{\pi}{1-\pi} \frac{\frac{1+b}{2} + \int_{1-\pi}^{1-b} \frac{\sigma}{1-b} d\sigma}{\frac{1-b}{2} + \int_{1-\pi}^{1-b} \frac{1-\sigma}{1-b} d\sigma}$$

Provided $b > (1+2\sqrt{2})/7$, we see that $LR(\pi)$ is decreasing on $(b, b+\epsilon)$ for some $\epsilon > 0$.

When $\delta = 0$, every action a is taken for posterior beliefs in a fixed interval $[\theta_1, \theta_2]$. In their "bounded beliefs example", Smith and Sørensen [2000] found that public beliefs can transition into a cascade set if and only if the posterior public belief after an action is not monotone in the prior belief. Hence:

Corollary 4 (No Cascades)**.** *Given (LC), for $\delta = 0$, cascades can't start after period one.*

It is instructive to observe that the multinomial signal examples with cascades in the seminal paper by Bikhchandani et al. [1992] violate assumption (LC).

3.9.1 Contrarian Behavior and its Applications

We now consider short run contrarian behavior, i.e., that individuals increasingly lean against actions increasingly favored by popular beliefs. We take two non-cascade public belief realizations $\pi < \pi'$, and generalize the finding from the professor-student example that the posterior belief threshold separating a pair of actions satisfies $\theta(\pi) < \theta(\pi')$. The generalization is to more than two actions, to the infinite-horizon model, and to the possibility of multiple optimal rules. We maintain the log-concavity assumption.

We have observed that, in general, the optimal action ordering depends on the public belief. But the threshold comparison $\theta(\pi) < \theta(\pi')$ is meaningful only when the same action ordering is optimal at both π and π', including that the set of active actions is identical.

Fixing one such action order, re-label the A active actions so that higher actions are taken at higher beliefs. Let θ_a denote the threshold between posterior beliefs leading to actions a and $a+1$, and define the *threshold vector* $\theta = (\theta_1, \ldots, \theta_{A-1})$. The *threshold space* $\Theta(\pi) \subset \mathbb{R}^{A-1}$ is the set of vectors θ where $r(\min \operatorname{supp}(F), \pi) < \theta_1 < \cdots < \theta_{A-1} < r(\max \operatorname{supp}(F), \pi)$, so all A actions are active.

For an interval rule defined by vector θ, the chance of action a is denoted $\psi(a, \pi, \theta)$, and the posterior public belief is $p(a, \pi, \theta)$.

Let $\Theta^*(\pi) \subset \Theta(\pi)$ be the set of optimal threshold vectors. We formally define that behaviour is *contrarian* if, for any pair $\pi < \pi'$ which have this identical optimal action ordering, $\Theta^*(\pi')$ is higher than $\Theta^*(\pi)$ in the strong set order.[28] This coincides with an intuitive notion of first order stochastic dominance: at the higher π', any set of lower actions $\{1, \ldots, a\}$ is taken for a higher set of posterior beliefs $[0, \theta_a]$. Behaviour is *strictly contrarian* if, for all $\theta \in \Theta^*(\pi)$ and $\theta' \in \Theta^*(\pi')$, we have $\theta' \gg \theta$ (all coordinates are higher).

Proposition 9 (Contrarianism)**.** *Assume (LC). Behaviour is contrarian.*

Figure **??** depicts the key step of the proof. The interior threshold between actions a and $a+1$ satisfies index function indifference, $w(a, \pi, \theta_a) = w(a+1, \pi, \theta_a)$, and that $w(a, \pi, \rho)$ down-crosses $w(a+1, \pi, \rho)$ at θ_a. By (37), the public belief affects w through the tangent to the value function. Tangents move sideways up along the convex value function, and predicts that $w(a, \pi', \theta_a) > w(a + 1, \pi', \theta_a)$, as illustrated.[29] Down-crossing yields the conclusion that $\theta'_a > \theta_a$.

Assumption (LC) guarantees updating monotonicity of public beliefs, and so a monotone tangent difference (44). Figure **??** further illustrates this part of the proof. We show in Appendix 3.11.5 that contrarianism can fail without monotone public beliefs.

Proposition 9 yields weak contrarian behavior. We can strengthen the proof to strict contrarianism if the value function is strictly convex (so the tangents strictly change) and the actions are taken in the natural order.

Corollary 5 (Strictly Contrarian)**.** *If signals obey (LC) and all actions are taken in the natural order, then behavior is strictly contrarian outside the cascade sets.*

We need here the value function to be strictly convex at a continuation belief. The value function is instead affine on $[\underline{z}, \bar{z}]$ if there exists an optimal strategy which is constant on $[\underline{z}, \bar{z}]$; in particular,

[28]Recall that Y' dominates Y in the strong set order if $y' \in Y', y \in Y \Rightarrow y \vee y' \in Y', y \wedge y' \in Y$.

[29]In fact, the proof works for any convex continuation value function — it needs not arise from our infinite horizon dynamic optimization. Proposition 9 is valid for the two-period professor-student problem in §3.8, for instance.

on cascade sets.[30] But outside cascade sets, it is not affine:[31]

Lemma 4 (Strict Value Convexity)**.** *If the private belief support* supp(F) *is convex and all actions are taken in the natural order, then the value function v is strictly convex outside the cascade sets.*

Lemma 4 yields another intriguing implication. Motivating individuals to behave altruistically as a function of their private posterior belief, requires a transfer monotonicity which is stronger than the planner's subtle contrarian change of the belief threshold. The transfers of the pivot mechanism in Proposition 7 reward being mimicked by successors.

Corollary 6 (Mimicry)**.** *Assume the natural action ordering in the binary action world. The transfers are ranked $t(a,a) > t(a,\neg a)$ whenever both π and $p(a,\pi,\xi)$ are not in $C(\delta)$.*

Proof: For the sake of argument, consider $a = 1$, with $p(1,\pi,\xi) \notin C(\delta)$ by assumption. First consider the case where also $p(2,\pi,\xi) \notin C(\delta)$. By equation (39),

$$t(1,1) - t(1,2) = \frac{M(1,\pi,L) - M(1,\pi,H)}{\psi(1,L) - \psi(1,H)}. \tag{41}$$

Now, $\psi(1,L) > \psi(1,H)$, as in (39). Thus, the fraction shares the sign of the numerator. From the definition of M in Proposition 7, $\tau(\pi,\rho)$ is a subtangent line of the value function v at π. Then $M(1,\pi,L) - M(1,\pi,H) = \partial\tau(\pi,\rho)/\partial\rho - \partial\tau(p(1,\pi,\xi),\rho)/\partial\rho$. By the natural action ordering, we have $p(1,\pi,\xi) < \pi$ and thus $\partial\hat{\tau}(\pi,\rho)/\partial\rho - \partial\hat{\tau}(p(1,\pi),\rho) > 0$, since v_δ is strictly convex at π by Lemma 4. Then $t(1,1) - t(1,2) > 0$.

Finally, when $p(2,\pi,\xi) \in C(\delta)$ the logic is the same, substituting M in (41) by M'. $\square$

3.9.2 The Detailed Proof of Contrarianism with Two Actions

While Proposition 9 is valid for belief pairs π, π' that can be far apart, our proof relies on a local argument which we explain in this subsection. By assumption, at both π and π' there exist optima with the same fixed action order.

[30] A strategy, started at $\pi \in [\underline{z}, \bar{z}]$, yields some state-contingent expected values v^H and v^L. As in the proof of Proposition 6, $\tau(\rho) = (1-\rho)v^L + \rho v^H$ is tangent to v at π. If v is affine, then $v(\pi) = \tau(\pi)$, and the strategy is optimal for all $\pi \in [\underline{z}, \bar{z}]$. Conversely, if the strategy is optimal, $v(\rho) = \tau(\rho)$ for all $\rho \in [\underline{z}, \bar{z}]$.

[31] Note that (LC) implies a convex belief support.

We explore the comparative statics properties of the constrained Bellman equation for any $\pi \notin C(\delta)$ where the planner can choose only among rules $\xi \in \Xi$ which maintain this same action order, but takes for granted the value function v from the unconstrained problem. So, granted the convex function v, define the Bellman function as in (34),

$$B(\theta|\pi) = \sum_{a=1}^{2} \psi(a,\pi,\theta)[(1-\delta)\bar{u}(a,p(a,\pi,\theta)) + \delta v(p(a,\pi,\theta))]. \tag{42}$$

Solutions to this constrained problem define the optimizer set $\Theta^*(\pi)$.

To prove Proposition 9, it suffices that $\Theta^*(\pi)$ increase in the strong set order. We wish to apply Theorem 1 in **?**. But they deliver our conclusion under the assumption that $B(\cdot|\pi')$ exceeds $B(\cdot|\pi)$ in their interval dominance order. Their Proposition 2 yields a sufficient condition — that there exist an increasing and strictly positive function $\alpha(\theta)$ such that, almost everywhere, $B_\theta(\theta|\pi') \geq \alpha(\theta) B_\theta(\theta|\pi)$.

So motivated, we derive an expression for $B_\theta(\theta|\pi)$.

Lemma 5. *The Bellman function B is differentiable almost everywhere with*

$$B_\theta(\theta|\pi) = g(\theta,\pi)\left(w(1,\pi,\theta) - w(2,\pi,\theta)\right), \tag{43}$$

and B is absolutely continuous: $B(\theta'|\pi) - B(\theta|\pi) = \int_\theta^{\theta'} B_\theta(\tilde{\theta}|\pi)\, d\tilde{\theta}$ for $\theta, \theta' \in \Theta(\pi)$.

The next result is a useful fact about tangents to a convex function (refer to Figure **??**).[32]

Lemma 6 (Changing Tangents to a Convex Function)**.** *Fix $z_1 < z_2 < z_3$ and a convex function v. Let τ_i be a tangent function to v at z_i. Then $\tau_2(z_1) \geq \tau_3(z_1)$ (and $\tau_1(z_3) \leq \tau_2(z_3)$), with strict inequality unless v is affine on $[z_2, z_3]$ (and on $[z_1, z_2]$).*

Returning to our main line of argument, suppose some $\theta \in \Theta^*(\pi)$ and $\theta' \in \Theta^*(\pi')$ are inversely ordered $\theta' < \theta$ — otherwise, we're done. Since $r(\sigma,\pi)$ is an increasing function of π, the open interval $\Theta(\pi)$ rises in π. So $[\theta', \theta] \subset \Theta(\pi) \cap \Theta(\pi')$. We first argue that the index difference $\Delta(\tilde{\theta},\pi) \equiv w(1,\pi,\tilde{\theta}) - w(2,\pi,\tilde{\theta})$ in (43) weakly increases in the public belief π when $\tilde{\theta} \in [\theta', \theta]$. By Lemma 3,

[32] As in Proposition 6, interpret τ_i as a tangent to the value function at continuation belief z_i. Intuitively, if we employ a continuation policy at belief z_1 that is optimal for a "more inaccurate" belief $z_3 > z_2$, then the payoff is lower, $\tau_3(z_1) \leq \tau_2(z_1)$.

continuation beliefs rise in public beliefs: $p(a, \pi', \tilde{\theta}) > p(a, \pi, \tilde{\theta})$ for $a = 1, 2$. The two cases in Lemma 6 yield, as desired,

$$\Delta(\tilde{\theta}, \pi') - \Delta(\tilde{\theta}, \pi) = \delta \left\{ \left[\tau_1'(\tilde{\theta}) - \tau_1(\tilde{\theta}) \right] + \left[\tau_2(\tilde{\theta}) - \tau_2'(\tilde{\theta}) \right] \right\} \geq 0. \tag{44}$$

Next, $\alpha(\tilde{\theta}) \equiv g(\tilde{\theta}|\pi')/g(\tilde{\theta}|\pi)$ is a positive and nondecreasing function over $[\theta', \theta]$, since g is log-supermodular by the proof of Lemma 3. Lemma 5 and inequality (44) imply:

$$B_\theta(\tilde{\theta}|\pi') = g(\tilde{\theta}|\pi')\Delta(\tilde{\theta}, \pi') \geq g(\tilde{\theta}|\pi')\Delta(\tilde{\theta}, \pi) = \alpha(\tilde{\theta}) B_\theta(\tilde{\theta}|\pi), \tag{45}$$

By Proposition 2 in **?**, (45) implies that B obeys their interval dominance order. By their Theorem 1, $\Theta(\pi)$ rises in the strong set order — contrarianism.

For Corollary 5, we now argue that the optimizer set *strictly rises*. Suppose that $\theta \geq \theta'$ are respectively optimal at public beliefs $\pi < \pi'$ — contrary to strict contrarianism. By the already proven strong set order, $\theta \in \Theta^*(\pi')$. By Proposition 6, $w(1, \pi, \theta) - w(2, \pi, \theta) = w(1, \pi', \theta) - w(2, \pi', \theta) = w(1, \pi', \theta') - w(2, \pi', \theta') = 0$. The first possibility $\theta > \theta'$ contradicts the fact that $w(2, \pi', \rho) - w(1, \pi', \rho)$ is a strictly increasing function of ρ. This fact follows from equation (37). For the natural action order implies that $\bar{u}(2, \rho) - \bar{u}(1, \rho)$ is strictly increasing, and convexity of v implies that its tangent difference $\tau_2'(\rho) - \tau_1'(\rho)$ is weakly increasing. Consider the other possibility $\theta = \theta'$. Now $\pi < \pi'$ implies $p(a, \pi, \theta) < p(a, \pi', \theta)$, and Claim 4 implies that at least one of $p(1, \pi', \theta), p(2, \pi, \theta)$ is outside the cascade set. Lemma 6 gives the contradiction $w(1, \pi, \theta) - w(2, \pi, \theta) > w(1, \pi', \theta) - w(2, \pi', \theta)$ — the inequality is strict because Lemma 4 provides strict convexity of v outside the cascade set. □

3.10 Conclusion

We have fully characterized the socially-planned herding model with two states in our model, and derived a general novel contrarianism prediction. Much of our analysis can be generalized to more states, but this comparative static exploits specifically scalar analysis. Extending the log-concavity condition and its monotone posterior implication beyond two states is a challenging open problem, as is the many state extension of contrarianism.

3.11 Appendix

3.11.1 Value Function Characterization Proofs

Value Convexity and Monotonicity: Proof of Lemma 11

The Bellman operator T is given by Tv equal to the RHS of (33). Note that for $v \geq v'$ we have $Tv \geq Tv'$. As is standard, T is a contraction, and so has a unique fixed point v_δ. This fixed point lies in the space of bounded, continuous, weakly convex functions. We simply show convexity. Since the operator T is a contraction, it suffices to prove that whenever v is convex, Tv is also convex. Let $\pi_\lambda = \lambda\pi_1 + (1-\lambda)\pi_2$, where $\lambda \in (0,1)$. Fix an optimal rule ξ mapping private beliefs to actions at π_λ. Using Bayes' rule, $p(a,\pi,\xi) = \pi\psi(a,H,\xi)/\psi(a,\pi,\xi)$, we get:

$$p(a,\pi_\lambda,\xi) = \frac{\lambda\psi(a,\pi_1,\xi)}{\psi(a,\pi_\lambda,\xi)}p(a,\pi_1,\xi) + \frac{(1-\lambda)\psi(a,\pi_2,\xi)}{\psi(a,\pi_\lambda,\xi)}p(a,\pi_2,\xi). \tag{46}$$

The first, myopic term in (34) at π_λ is the convex combination of the terms with π_1 and π_2, since $\bar{u}$ is linear in beliefs. The second, future term obeys:

$$\psi(a,\pi_\lambda,\xi)v(p(a,\pi_\lambda,\xi)) \leq \lambda\psi(a,\pi_1,\xi)v(p(a,\pi_1,\xi)) + (1-\lambda)\psi(a,\pi_2,\xi)v(p(a,\pi_2,\xi)).$$

given (46), since v is convex. Summing over actions $a = 1,\ldots,A$ yields $Tv(\pi_\lambda) = T_\xi v(\pi_\lambda) \leq \lambda T_\xi v(\pi_1) + (1-\lambda)T_\xi v(\pi_2) \leq \lambda Tv(\pi_1) + (1-\lambda)Tv(\pi_2)$, as desired.

Let $\tilde{u}(\pi) = \max_a \bar{u}(a,\pi)$ denote the payoff frontier. The bound on tangent slopes follows from the observations that $v(0) = u(1,L)$ and $v(1) = u(A,H)$, that the convex function v exceeds the payoff frontier $\tilde{u}$, and that $\bar{u}(1,\rho)$ and $\bar{u}(A,\rho)$ define the most extreme slopes of $\tilde{u}$ by supermodularity.

Claim 1. *The function sequence $\{T^n\tilde{u}\}$ is pointwise increasing and converges to v_δ. The value v_δ weakly exceeds $\tilde{u}$, and strictly so outside the cascade sets.*

Proof. To maximize $\sum_{a=1}^{A}\psi(a,\pi,\xi)\left[(1-\delta)\bar{u}(a,p(a,\pi,\xi)) + \delta\tilde{u}(p(a,\pi,\xi))\right]$ over Ξ for given π, one rule $\tilde{\xi}$ almost surely chooses the myopically optimal action. Then $p(\tilde{\xi}(\sigma),\pi,\tilde{\xi}) = \pi$ a.s., resulting in value $\tilde{u}(\pi)$. Optimizing over all $\xi \in \Xi$, $T\tilde{u}(\pi) \geq \tilde{u}(\pi)$ for all π. By induction, $T^n\tilde{u} \geq T^{n-1}\tilde{u}$, yielding a pointwise increasing sequence converging to the fixed point $v_\delta \geq \tilde{u}$. Finally, when π is outside the cascade sets, by definition it is *not* optimal to almost surely induce one action. So, $v_\delta(\pi) > \tilde{u}(\pi)$ $\forall\delta \in [0,1)$ and $\forall\pi \notin \cup_{a=1}^{A} C_a(\delta)$. □

Claim 2 (Weak Value Monotonicity)**.** *When $\delta_2 \geq \delta_1$, $v_{\delta_2}(\pi) \geq v_{\delta_1}(\pi)$ for all π.*

Proof. Clearly, $\sum_{a=1}^{A} \psi(a,\pi,\xi)\bar{u}(a,p(a,\pi,\xi)) \leq \sum_{a=1}^{A} \psi(a,\pi,\xi)v(p(a,\pi,\xi))$ for any ξ and any function $v \geq \tilde{u}$. At higher δ, then $T_\xi \tilde{u}$ is pointwise higher, since more weight is placed on the larger component of the RHS of (34). By (33), $T\tilde{u}$ is pointwise higher. Iterating this argument, $T^n\tilde{u}$ is higher. Let $n \to \infty$ and apply Claim 1. □

3.11.2 Cascade Sets: Proof of Lemma 1 and More

A. Proof of Lemma 1. The certain choice of a is optimal iff $v_\delta(\pi) = \bar{u}(a,\pi)$. As $\bar{u}(a,\pi)$ is affine in π, and v_δ is weakly convex, this equality holds on a closed interval $C_a(\delta)$. When δ is greater, v_δ is weakly higher by Claim 2, and hence $C_a(\delta)$ is weakly smaller.

Next, action 1 is myopically strictly optimal when $\pi = 0$. Since it updates to continuation belief $\pi = 0$ for any rule, it is also dynamically optimal for any discount factor $\delta \in [0,1)$. A similar proof holds for $\pi = 1$. Since the private signal is valuable in the selfish problem, $\cup_{a=1}^{A} C_a(0) \neq [0,1]$.

We now prove that for bounded beliefs, namely with $\text{supp}(F) \subseteq [\sigma_1,\sigma_0] \subset (0,1)$, the cascade sets for actions 1 and N are $C_1(\delta) = [0,\underline{\pi}(\delta)]$ and $C_A(\delta) = [\bar{\pi}(\delta),1]$, where $0 < \underline{\pi}(\delta) < \bar{\pi}(\delta) < 1$. For low beliefs, it is optimal to let the rule ξ induce 1; the argument for high beliefs is similar. Action 1 is optimal at belief $\pi = 0$, and there is no tie, so 1 is the optimal selfish choice for beliefs $\pi \leq \pi'$, for some $\pi' > 0$. In particular, $\bar{u}(1,\pi) > \bar{u}(a,\pi) + \eta$ for all $a \neq 1$ for some $\eta > 0$, and for all beliefs π in the interval $[0,\pi'/2]$. No action can reveal a stronger private signal than any $\sigma \in \text{supp}(F) \subseteq [\sigma_1,\sigma_0] \subset (0,1)$. So any initial belief π is updated to at most $\bar{p}(\pi) = \pi\sigma_1/[\pi\sigma_1 + (1-\pi)(1-\sigma_1)]$. For π small enough, $\bar{p}(\pi) \in [0,\pi'/2]$ and $\bar{p}(\pi) - \pi$ is arbitrarily small. By continuity of v_δ, $v_\delta(\bar{p}(\pi)) - v_\delta(\pi)$ is less than $\eta(1-\delta)/\delta$ for small enough π. By the Bellman equation (33), any action $a \neq 1$ is strictly suboptimal for such small beliefs.

Next, assume unbounded beliefs. Smith and Sørensen [2000] prove that $C_a(0) = \varnothing$ for all $a \neq 0,1$, and that $C_1(0) = \{0\}$ and $C_A(0) = \{1\}$. □

B. Cascade Sets as Limit Beliefs. As in Smith and Sørensen [2000], the martingale convergence theorem implies that public beliefs converge, and their limit is not fully wrong:

Claim 3. *The belief process* $\langle \pi_n \rangle$ *is a martingale unconditional on the state, converging a.s. to some limiting random variable* π_∞*. The limit* π_∞ *is concentrated on* $(0, 1]$ *in state* H*.*

Smith and Sørensen [2000] find for $\delta = 0$ that the public belief process converges upon the cascade set. The result extends to the case $\delta > 0$:

Theorem 5 (Convergence of Beliefs)**.** *Consider a solution of the planner's problem. The limit belief* π_∞ *has support in the cascade sets* $C_1(\delta) \cup \cdots \cup C_A(\delta)$*. In particular,* π_∞ *is concentrated on the truth for unbounded private beliefs.*

Proof: At least two actions occur with positive chance for any belief π not in any cascade set. By the interval structure of Corollary 1, the highest such action is more likely in state H, and the lowest in state L. So the next period's belief differs from π with positive probability. Intuitively, or by the characterization result for Markov-martingale processes in Appendix B of Smith and Sørensen [2000], π cannot lie in the support of π_∞.

3.11.3 Differentiable Continuations: Proof of Corollary 2

Abbreviate $p(a, \pi)$ for $p(a, \pi, \xi)$. For simplicity, at $\pi \notin C(\delta)$, consider 2 active actions. By (36) and (37), $v(\pi) = F^\pi(\sigma)w(1, \pi, p(1, \pi)) + (1 - F^\pi(\sigma))w(2, \pi, p(2, \pi))$, where the optimal private belief threshold σ satisfies $w(1, \pi, r(\pi, \sigma)) = w(2, \pi, r(\pi, \sigma))$.

From Proposition 6, $w(a, \pi, \rho)$ is defined through a tangent τ_a of v at $p(a, \pi)$. If v is not differentiable at $p(2, \pi)$, there exists a different tangent $\hat{\tau}_2$ of v at $p(2, \pi)$. Define the new index function $\hat{w}(2, \pi, \rho) = (1 - \delta)\bar{u}(2, \rho) + \delta\hat{\tau}_2(\rho)$ for action 2. At the same threshold σ, we obtain the same value, $v(\pi) = F^\pi(\sigma)w(1, \pi, p(1, \pi)) + (1 - F^\pi(\sigma))\hat{w}(2, \pi, p(2, \pi))$. Since both actions are active, σ lies strictly inside the convex support of F, and $r(\pi, \sigma) < p(2, \pi)$. Thus, $w(1, \pi, r(\pi, \sigma)) \neq \hat{w}(2, \pi, r(\pi, \sigma))$. It follows that this choice of tangents allow for a different threshold obtaining a higher value than $v(\pi)$, in contradiction to optimality of σ.

3.11.4 Implementation by Transfers: Proof of Proposition 7

We start with a general new property of informational herding models with cascades:

Claim 4. *Let $\delta > 0$. For any belief π, continuation beliefs lie in at most one cascade set.*

Proof: Given unbounded private beliefs, continuation beliefs never lie in a cascade set. Assume bounded private beliefs. Let $\underline{\sigma} = \min \operatorname{supp}(F)$ and $\bar{\sigma} = \max \operatorname{supp}(F)$. Suppose that for some π, two continuation beliefs $\pi_1 < \pi_2$ lie in distinct cascade sets, namely, $C_{a'}(\delta)$ below $C_{a''}(\delta)$. Then $\pi_1 \in C_{a'}(0)$ and $\pi_2 \in C_{a''}(0)$ by weak monotonicity of cascade sets in δ (Proposition 8). Let $\pi' = \max C_{a'}(0) \leq \pi'' = \min C_{a''}(0)$. Then $\pi_1 \leq \pi'$. There exist x_1, x_2 in $[\underline{\sigma}, \bar{\sigma}]$ with $r(\pi, x_1) = \pi_1$ and $r(\pi, x_2) = \pi_2$. Since (a) Bayes-updating commutes, (b) $r(\pi, \bar{\sigma}) \geq r(\pi, x_2) = \pi_2$ and $x_1 \geq \underline{\sigma}$, (c) $\pi_2 \geq \pi''$, and (d) $\pi'' \in C_{a''}(0)$ while $\pi' \in C_{a'}(0)$:

$$r(\pi_1, \bar{\sigma}) = r(r(\pi, x_1), \bar{\sigma}) = r(r(\pi, \bar{\sigma}), x_1) \geq r(\pi_2, \underline{\sigma}) \geq r(\pi'', \underline{\sigma}) \geq r(\pi', \bar{\sigma})$$

and so $\pi_1 \geq \pi'$. Thus $\pi_1 = \pi'$, which contradicts Claim 5. □

Claim 5. *The interior endpoints of each cascade set $C_a(0)$ are not in $C_a(\delta)$, for any $\delta > 0$, and any action $a \in A$.*

Proof: Let $\breve{\pi} = \min C_a(0)$ where $a \neq 1$. Denote the minimal private posterior belief by $\breve{\rho} = r(\breve{\pi}, \min \operatorname{supp}(F))$. Then $\bar{u}(a-1, \breve{\rho}) = \bar{u}(a, \breve{\rho})$. Define

$$w_{a-1}(\rho) = (1-\delta)\bar{u}(a-1, \rho) + \delta\tau(\rho) \quad \text{and} \quad w_a(\rho) = (1-\delta)\bar{u}(a, \rho) + \delta\bar{u}(a, \rho) \tag{47}$$

where τ is a tangent of v_δ at $\breve{\rho}$. Next, since $\breve{\rho} < \breve{\pi}$, we have $\breve{\rho} \notin C_a(0)$. So $\breve{\rho} \notin C_a(\delta)$ by Lemma 1-(d); thus, $\bar{u}(\breve{\rho}, a) < v_\delta(\breve{\rho}) = \tau(\breve{\rho})$. Plugging this inequality into (47) gives $w_a(\breve{\rho}) < w_{a-1}(\breve{\rho})$. If $\breve{\pi} \in C_a(\delta)$, then $w_a(\breve{\rho})$ is the value to the altruistic agent with posterior belief $\breve{\rho}$, but our inequality then contradicts the behaviour of Proposition 6. □

3.11.5 Contrarianism Proofs

*The Role of Posterior Monotonicity in Contrarianism We show by an example that the posterior monotonicity property obtained in Lemma 3 is necessary for contrarianism in Proposition 9 when the convex value function v can be chosen freely (see footnote 29). We use a version of the two-period professor-student example with $\delta = 1$ in §3.8 to show the principle. The student has three actions available, while the professor has two actions taken in the natural order. The student gets no private signal. The professor's signal is described by the conditional density $g(\rho|\pi)$. This signal structure violates posterior monotonicity for an interval, say $[\hat{\theta}, 1]$. Thus,

$$p' \equiv \int_{\hat{\theta}}^{1} \rho g(\rho|\pi')d\rho / \int_{\hat{\theta}}^{1} g(\rho|\pi')d\rho > \int_{\hat{\theta}}^{1} \rho g(\rho|\pi'')d\rho / \int_{\hat{\theta}}^{1} g(\rho|\pi'')d\rho \equiv p''.$$

By this reversal, $\hat{\theta}$ must lie strictly inside the posterior belief supports at π', π'', so $p'' > \hat{\theta}$.

Figure 6 illustrates the convex value function that we construct for the example. First choose an arbitrary $\theta_{23} \in (p'', p')$. For any $\varepsilon > 0$, the convex function $\hat{v}(p|\varepsilon)$ consists of three linear segments $\ell_1, \ell_2, \ell_3(\varepsilon)$. Segments ℓ_1, ℓ_2 intersect at $\hat{\theta}$, while $\ell_2, \ell_3(\varepsilon)$ intersect at θ_{23}. ℓ_2 is steeper than ℓ_1, and the slope of ℓ_3 is $\varepsilon > 0$ higher than ℓ_2. The intersection of the extended line segments $\ell_1, \ell_3(\varepsilon)$ is denoted $\theta_{13}(\varepsilon)$.

We will show that when $\varepsilon > 0$ is small enough, $\hat{\theta}$ is the unique optimal threshold at π'', while only the strictly higher $\theta_{13}(\varepsilon)$ and θ_{23} are candidates for optimal thresholds at the lower π'. In either case, contrarianism fails.

Observe that the three kink points $\hat{\theta}, \theta_{12}, \theta_{23}(\varepsilon)$ describe the only candidates for optimal policies. By construction, they are the only ones that solve for index indifference — given discount factor $\delta = 1$, only the tangents to the value function matter. It remains to check suboptimality of a cascade policy, whereby the posterior is the prior. But the interior threshold $\hat{\theta}$ gives strictly more than $\hat{v}(\pi|0)$ at $\pi = \pi', \pi''$, due to the kink at $\hat{\theta}$.

Consider π'. The first order condition fails at $\hat{\theta}$ for any $\varepsilon > 0$, as the tangent at the upper posterior p' is ℓ_3. So optimal posterior cut-offs are among $\theta_{13}(\varepsilon), \theta_{23}$.

Consider π''. First, suppose we use the belief cutoff $\theta_{13}(\varepsilon)$. As $\varepsilon \downarrow 0$, the crossing point θ_{13}

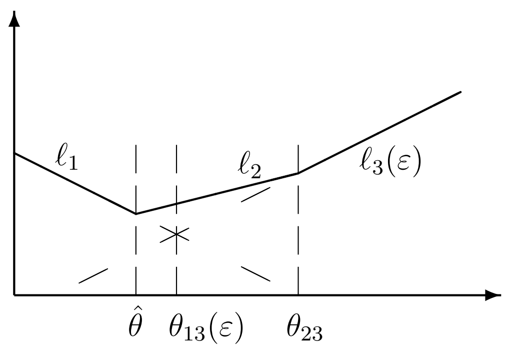

Figure 6: **Necessity Principle.** The student's value function for §3.11.5.

converges to $\hat{\theta}$, and the upper posterior belief converges to p''. In other words, it is eventually below θ_{23}, since $p'' < \theta_{23}$. At that point, the tangents at the continuation posteriors after π'' are ℓ_1 and ℓ_2. These tangents intersect at $\hat{\theta}$, and therefore the first order condition fails at $\theta_{13}(\varepsilon)$. Second, suppose we use the belief cutoff θ_{23}. Since $\theta_{23} \in (p'', p')$, it is strictly inside the posterior belief support. Thus, the upper posterior lies in $(\theta_{23}, 1]$, and the lower one either lies in $[0, \hat{\theta})$ or $[\hat{\theta}, \theta_{23})$. If in $[0, \hat{\theta})$, the tangents at the continuation beliefs are ℓ_1 and $\ell_3(\varepsilon)$. These cross at $\theta_{13}(\varepsilon)$, and so the first order condition fails at θ_{23}. If in $[\hat{\theta}, \theta_{23})$, the first order condition holds. But as $\varepsilon \downarrow 0$, the continuation value approaches $\hat{v}(\pi''|0)$. But as noted before, $\hat{\theta}$ yields a strictly higher continuation value than $\hat{v}(\pi''|0)$. □

3.11.6 Bellman Derivative Formula: Proof of Lemma 5

From (42), the Bellman function is almost everywhere differentiable in θ. For assumption (LC) implies that $p(a, \pi, \theta)$ is strictly monotone and differentiable, and the convex function v must be differentiable almost everywhere. Following the derivation of (36) and using Proposition 6, we can rewrite the Bellman function as

$$B(\theta|\pi) = \int_0^\theta w(1, \pi, \rho) g(\rho|\pi) d\rho + \int_\theta^1 w(2, \pi, \rho) g(\rho|\pi) d\rho. \tag{48}$$

The result follows. Later on, we need a many action generalization.

Claim 6 (Bellman Derivative)**.** *Let $\theta \in \Theta(\pi)$. Assume $\theta_a = \cdots = \theta_{a+j} = x$ for some $a \geq 1$ and $j \geq 0$*

with $a+j \leq A-1$, and suppose that $\theta_{a-1} < x < \theta_{a+j+1}$.[33] *Then the Bellman function B in (50) is absolutely continuous with respect to x, and its derivative in x almost everywhere equals:*

$$B_x(\theta|\pi) \equiv g(x|\pi)\left(w(a,\pi,x) - w(a+j+1,\pi,x)\right). \tag{49}$$

Moreover, for all $\pi'' > \pi'$, there exists a positive and increasing function $\alpha(x)$ such that the Bellman function $B(\theta|\pi)$ almost everywhere obeys $B_x(\theta|\pi'') \geq \alpha(x) B_x(\theta|\pi')$ when $\theta \in \Theta(\pi') \cap \Theta(\pi'')$.

The omitted proof follows closely on Lemma 5, since we take action a for the posteriors $\rho \in [\theta_{a-1}, x]$, and action $a+j+1$ for posteriors $\rho \in [x, \theta_{a+j+1}]$. So the derivative of the Bellman function B in x is similar to (43) which had payoffs and tangents for actions $a=1$ and $a+j+1=2$. Thus, (49) follows, and the inequality follows similarly from (45). □

3.11.7 Subtangents to a Convex Function: Proof of Lemma 6

When v is affine on $[z_1, z_2]$, subtangents τ_1 and τ_2 can coincide, with $\tau_1(z_3) = \tau_2(z_3)$. Otherwise, the subtangent τ_2 is steeper than τ_1. Thus, $\tau_2(z_3) - \tau_2(z_2) > \tau_1(z_3) - \tau_1(z_2)$, whence $\tau_2(z_3) - \tau_1(z_3) > \tau_2(z_2) - \tau_1(z_2)$. Since v is convex, the subtangent τ_1 lies below v at z_2, so that $\tau_2(z_2) = v(z_2) \geq \tau_1(z_2)$. We conclude that $\tau_2(z_3) > \tau_1(z_3)$. The analysis at z_1 is similar. □

*Contrarianism: Proof of Proposition 9 for Multiple Actions

Claim 7. *The threshold space $\Theta(\pi)$ is a lattice, and B is supermodular for $\theta \in \Theta(\pi)$.*

Proof. Assume $\theta, \theta' \in \Theta(\pi)$. Then $\theta \wedge \theta' \in \Theta(\pi)$ since $(\theta \wedge \theta')_a = \theta_a \wedge \theta'_a \leq \theta_{a+1} \wedge \theta'_{a+1} = (\theta \wedge \theta')_{a+1}$ for every a. Similarly, $\theta \vee \theta' \in \Theta(\pi)$. Next, to show that B is supermodular in θ, let $\theta'_a > \theta_a$. If θ_{-a} increases, both continuation beliefs $p(a,\pi,\theta)$ and $p(a+1,\pi,\theta)$ increase. Since $p(a,\pi,\theta) < \theta_a < p(a+1,\pi,\theta)$, Lemma 6 implies that $w(a,\pi,\theta_a)$ increases while $w(a+1,\pi,\theta_a)$ decreases. So the difference $(w(a,\pi,\theta_a) - w(a+1,\pi,\theta_a))$ increases in θ_{-a}. Then by (49), the Bellman difference $B(\theta'_a, \theta_{-a}) - B(\theta_a, \theta_{-a})$ increases in θ_{-a}. Supermodularity can now be decomposed into a summation of differences of this form. □

[33] Notation: $\theta_0 = r(\min \operatorname{supp}(F), \pi)$ and $\theta_A = r(\max \operatorname{supp}(F), \pi)$.

Fixing the action ordering, the Bellman function (34) for a convex continuation value v is:

$$B(\theta|\pi) = \sum_{a=1}^{A} \psi(a,\pi,\theta)[(1-\delta)\bar{u}(a,p(a,\pi,\theta)) + \delta v(p(a,\pi,\theta)). \tag{50}$$

We now prove Proposition 9 for finitely many actions. Pick beliefs $\pi < \pi'$ and assume that $\theta \in \Theta^*(\pi)$ and $\theta' \in \Theta^*(\pi')$. If $\theta \le \theta'$, we are done. Assume next that they are inversely ordered $\theta' < \theta$. We verify $\theta \in \Theta^*(\pi')$ and $\theta' \in \Theta^*(\pi)$. First, both $[\theta_1, \theta_{A-1}]$ and $[\theta'_1, \theta'_{A-1}]$ are subsets of $\Theta(\pi) \cap \Theta(\pi')$, since $[\theta_1, \theta_{A-1}] \subset \Theta(\pi)$ and $[\theta'_1, \theta'_{A-1}] \subset \Theta(\pi')$ and $[\theta_1, \theta_{A-1}]$ lies above $[\theta'_1, \theta'_{A-1}]$ in the strong set order, and yet $\Theta(\pi)$ lies below $\Theta(\pi')$ in the strong set order. Second, let X be the set of all cut-off rules with cut-off points in $\Theta(\pi) \cap \Theta(\pi')$. By **?**, $B(\theta|\pi')$ dominates $B(\theta|\pi)$ in the interval dominance order over X, since by Claim 6, the condition in the Proposition 2 in **?**, is satisfied.

Finally, suppose that θ and θ' are not ordered. We now need a stronger proof ingredient — specifically, we exploit the supermodularity of B (Claim 7). Our result follows if:

$$B(\theta|\pi) - B(\theta \wedge \theta'|\pi) \ge 0\ (>0) \Longrightarrow B(\theta \vee \theta'|\pi') - B(\theta'|\pi') \ge 0\ (>0). \tag{51}$$

Let's see why this suffices. Since θ is optimal at π, the left side is non-negative, and thus $\theta \vee \theta'$ is optimal at π' by the weak inequality in (51). Conversely, if $\theta \wedge \theta'$ is not optimal at π, then θ' is not optimal at π', by the strict inequality in (51).

We split the proof of (51) into two parts, since the choice domain $\Theta(\cdot)$ depends on the public belief. Let $(\theta_a, ..., \theta_{A-1})$ be the components of θ inside $\Theta(\pi')$, for some $a < A$. Choose $z \in \Theta(\pi')$ with $z < \min\{\theta_a, \theta'_1\}$. Let $\hat{\theta} = (z, ..., z, \theta_a, ..., \theta_{A-1})$, where the first $a-1$ components are z. Then $\hat{\theta} \in \Theta(\pi) \cap \Theta(\pi')$, since $\theta_{a-1} < z$ follows from $\theta_{a-1} \notin \Theta(\pi')$.

By supermodularity of $B(\cdot|\pi')$, and because $\hat{\theta} \vee \theta' = \theta \vee \theta'$, we have:

$$B(\hat{\theta}|\pi') - B(\hat{\theta} \wedge \theta'|\pi') \ge (>0) \Longrightarrow B(\theta \vee \theta'|\pi') - B(\theta'|\pi') \ge (>0). \tag{52}$$

Then (51) follows if we also argue:

$$B(\theta|\pi) - B(\theta \wedge \theta'|\pi) \geq (> 0) \Longrightarrow B(\hat{\theta}|\pi') - B(\hat{\theta} \wedge \theta'|\pi') \geq (> 0). \tag{53}$$

We now prove (53). First, for all $\theta'' \in [\hat{\theta} \wedge \theta', \hat{\theta}]$, we have $\hat{\theta} = \theta \vee \theta''$ and so:

$$B(\hat{\theta}|\pi) - B(\theta''|\pi) \geq B(\theta|\pi) - B(\theta \wedge \theta''|\pi) \geq 0, \tag{54}$$

by supermodularity of $B(\cdot|\pi)$ and optimality of θ at π, respectively. When $\theta'' = \hat{\theta} \wedge \theta'$ in (54), we have $B(\hat{\theta}|\pi) - B(\hat{\theta} \wedge \theta'|\pi) \geq B(\theta|\pi) - B(\theta \wedge \theta'|\pi)$, since $\theta \leq \hat{\theta}$. Hence, if $B(\theta|\pi) - B(\theta \wedge \theta'|\pi) > 0$, then $B(\hat{\theta}|\pi) - B(\hat{\theta} \wedge \theta'|\pi) > 0$. Finally, the interval dominance ordering of $B(\cdot|\pi')$ over $B(\cdot|\pi')$ lets us conclude (53). □

3.11.8 Strict Contrarianism: Proof of Corollary 5

Pick $\pi' > \pi$. Let $\theta \in \Theta^*(\pi)$ and $\theta' \in \Theta^*(\pi')$. By Proposition 9, behavior is contrarian. Suppose for a contradiction that it is not strictly so, and thus $\theta'_k \leq \theta_k$ for some k. By Proposition 9, $\theta \vee \theta'$ is optimal under π'. Since $\theta'_k \leq \theta_k$, we have $(\theta \vee \theta')_k = \theta_k$. Suppose that a_j is the highest active action below a_k, and a_m the least active action above a_k. Then $(\theta \vee \theta')_j < (\theta \vee \theta')_{j+1} = \cdots = (\theta \vee \theta')_k = \cdots = (\theta \vee \theta')_{m-1} < (\theta \vee \theta')_m$, since θ and θ' have the same active actions in natural order. Our proof for two actions then carries over to this case, by considering a neighboring pair of active actions. □

References

Wolfgang Adamski. Capacitylike set functions and upper envelopes of measures. *Math. Ann*, 229 (237-244), 1977.

Bernd Anger. Representation of capacities. *Math. Ann*, 229(245-258), 1977.

Susan Athey. Monotone comparative statics under uncertainty. *The Quarterly Journal of Economics*, 117(187-223), 2002.

Abhijit V. Banerjee. A simple model of herd behavior. *Quarterly Journal of Economics*, 107(797–817), 1992.

Dallas Banks. An integral inequality. *Proc. Amer. Math. Soc*, 14(823-828), 1963.

Gary Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(813-846), 1973.

Dirk Bergemann, Ji Shen, Yun Xu, and Edmund M. Yeh. Mechanism design with limited information: the case of nonlinear pricing. *Game Theory for Networks*, 75, 2012.

Dimitri Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice Hall, Englewood Cliffs, N.J., 1987.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as information cascades. *Journal of Political Economy*, 100:992–1026, 1992.

Lluís Bru and Xavier Vives. Informational externalities, herding and incentives. *Journal of Institutional and Theoretical Economics*, 158:91–105, 2002.

Hector Chade and Lones Smith. Simultaneous search. *Econometrica*, 74(5), 2006.

Christophe Chamley and Douglas Gale. Information revelation and strategic delay in a model of investment. *Econometrica*, 62:1065–1085, 1994.

Gustave Choquet. Theory of capacities. *Annales de I'institut Fourier*, 5(131-295), 1954.

Leon Yang Chu and David E.M. Sappington. Simple cost-sharing contracts. *The American Economic Review*, 97(1), 2007.

Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(1431-1451), 1982.

James Dow. Search decisions with limited memory. *Review of Economics Studies*, 58(1), 1991.

Matthew Doyle. Informational externalities, strategic delay, and the search for optimal policy. ISU Economics Working Paper, 2002.

David Easley and Nicholas Kiefer. Controlling a stochastic process with unknown parameters. *Econometrica*, 56:1045–1064, 1988.

Bennett Fox. Discrete optimization via marginal analysis. *Management Science*, 13(3), 1966.

Lars P. Metzger Gerhard Jager and Frank Riedel. Voronoi languages equilibria in cheap-talk games with high-dimensional types and few signals. *Games and Economic Behavior*, 73(517-537), 2011.

J.C̃. Gittins. Bandit processes and dynamical allocation indices. *Journal of the Royal Statistical Society, Series B*, 41:148–177, 1979.

Robert M. Gray and David L. Neuhoff. Quantization. *IEEE Transacctions on Information Theory*, 44(6), 1998.

Faruk Gul and Ennio Stacchetti. Walrasian equilibrium with gross substites. *Journal of Economic Theory*, 87(95-124), 1999.

Hugo A. Hopenhayn and Edward C. Presscott. Stochastic monotonicity and stationary distribution for dynamic economics. *Econometrica*, 60(6), 1992.

E.L. Lehman. Comparing location experiments. *The Annals of Statistics*, 16(2), 1988.

Benny Lehmann, Daniel Lehmann, and Noam Nisan. Combinatorial auctions with decreasing marginal utilities. *Games and Economic Behavior*, 55(153), 2006.

Ehud Lehrer and Roee Teper. The concave integral over large spaces. *Fuzzy Sets and Systems*, 159 (2130-2144), 2008.

Barton L. Lipman. Why is language vague? *working paper*, 2009.

STUART P. Lloyd. Least square quantization in pcm. *IEEE Transactions in Information Theory*, 28 (127-135), 1982.

L. Lovasz. Submodular functions and convexity. *Mathematical Programming: The state of the art*, (235-257), 1982.

Massimo Marinacci and Luigi Montrucchio. Introduction to the mathematics of ambiguity. *memo*, 2004.

Albert W. Marshall and Ingram Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, San Diego, 1979.

Joel Max. Quantizing for minimum distortion. *IEEE Transactions in Information Theory*, 6(7-12), 1960.

R. Preston McAfee. Coarse matching. *Econometrica*, 70(2025-2034), 2002.

Luis A. Medrano and Xavier Vives. Strategic behavior and price discovery. *Rand Journal of Economics*, 32:221–248, 2001.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2), 2002.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(157-180), 1994.

H. Moulin. Axioms of cooperative decision making. *Cambridge University Press, Cambridge, UK*, 1988.

B. M. Oliver, J. Pierce, and C. E. Shannon. The philosophy of pcm. *Proc. IRE*, 36(1324-1331), 1948.

Nicola Persico. Information acquisition in auctions. *Econometrica*, 68(135-148), 2000.

John K.-H. Quah and Bruno Strulovici. Comparative statics with the interval dominance order 2. *mimeo*, 2007.

John K.-H. Quah and Bruno Strulovici. Comparative statics, infomativeness, and the interval dominance order. *Econometrica*, 77(1949-1992), 2009.

Roy Radner. Team decision problems. *Annals of Mathematical Statistics*, 33:857–881, 1962.

William P. Rogerson. Simple menus of contracts in cost-based procurement and regulation. *The American Economic Review*, 93(3), 2003.

Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9: 185–202, 1974.

Walter Rudin. Real and complex analysis (third version). *McGraw-Hill Companies*, 1987.

D. Schmeidler. Subjective probabilities without additivity. *Econometrica*, 57(571-587), 1989.

Daniel Sgroi. Optimizing information in the herd: Guinea pigs, profits, and welfare. *Games and Economic Behavior*, 39:137–166, 2002.

Ran Shao. Generalized coarse matching. *working paper*, 2011.

Lones Smith and Peter Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68: 371–398, 2000.

Lones Smith, Peter Sørensen, and Jianrong Tian. Informational herding, optimal experimentation and contrarianism. *Revised and resubmitted for Review of Economic Studies*, 2012.

Dezso Szalay. Strategic information transmission and stochastic orders. *working paper, University of Bonn*, 2012.

Jianrong Tian. Monotone pragmatics. *working paper*, UW Madison, 2015.

Donald M. Topkis. Minimizing a submodular function on a lattice. *Operations Research*, 26(305-321), 1978.

Donald M. Topkis. supermodularity and complementarity. *Princeton University Press*, 1998.

Xavier Vives. How fast do rational agents learn? *Review of Economic Studies*, 60:329–347, 1993.

Xavier Vives. Learning from others: A welfare analysis. *Games and Economic Behavior*, 20:177–200, 1997.

Robert Wilson. Efficient and competitive rationing. *Econometrica*, 57 (1)(1-40), 1989.

Adam Chi Leung Wong. The choice of the number of varieties: Justifying simple mechanisms. *Journal of Mathematical Economics*, 54(7-21), 2014.